To the Graduate Council:

I am submitting herewith a thesis written by Scott M. Thayer entitled "DESIGN AND ANALYSIS OF HUMAN-MACHINE AND AUTOMATIC VISION SYSTEMS FOR CYLINDER DETECTION AND LOCALIZATION." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Mohan M. Trivedi

Prof. Mohan M. Trivedi, Major Professor

We have read this thesis
and recommend its acceptance:

Reid Kress

D. Vaughn Blalock

Accepted for the Council:

Associate Vice Chancellor
and Dean of the Graduate School

DESIGN AND ANALYSIS OF HUMAN-MACHINE AND AUTOMATIC VISION SYSTEMS FOR CYLINDER DETECTION AND LOCALIZATION

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Scott M. Thayer

August 1993

Dedicated to the idea that sacrifice is the toll of success

to my parents for demonstrating it

to my friends for allowing me to forget

Mirrors and Images

Portal of perception,
Pool of deception
Get from my sight!
Why can't we separate truth from light?

- Scott M. Thayer
September 1991

Acknowledgements

I am deeply indebted to Dr. Mohan Trivedi for taking a chance and giving me an opportunity to prove myself at a point in my life when I did not deserve one. His insight, advice, and example have been invaluable. I would like to thank my committee members Dr. Reid Kress and Dr. T. V. Blalock for their patience, criticism, and guidance. I would also like to thank the Computer Vision and Robotics Research group for promoting a creative and enjoyable work place and for suffering through an endless stream of pranks and bad jokes.

A special thanks goes to Fred Depiero, Hugh Costello, Phil Butler, and Tim Noell at the Robotics and Process Systems Division of the Oak Ridge National Laboratory for their supervision and assistance. From the CVRR Laboratory, I thank Steve Rich, Gopalan Ravichandran, and Chris Gourley for their comments on the manuscript.

This research was sponsored by a grant awarded by the Office of Technology Development, U.S. Department of Energy under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

Abstract

For most robotic tasks, two different philosophies dictate how the applications designer approaches a problem. The first philosophy attempts to produce an application that is fully operational inside its task boundaries without human intervention. Examples of these types of applications are robot welders. These robots have the necessary knowledge base and/or intelligence to accurately perform the welding task. These types of systems are highly desirable for a large number of tasks; however, certain applications may require, due to the lack of technology or the task nature, intervention at critical steps in the process. Man-machine system design philosophies govern such situations. Telerobotic control is an example of man-machine systems design. In these systems, a remote operator interacts with a user interface, usually a master arm, and directs the movement of the actual robot in the work space.

Most of the research in machine vision has been directed toward the development of fully autonomous vision systems. Such systems have been met with varying degrees of success depending on the complexity of the task and the skill of the designer. Logically, vision systems would benefit from the addition of a human operator into the "perceptual loop." Applying these concepts to perception, the research in the second section of this text describes a human-machine cylinder detection system. The design process includes a hierarchical task analysis decomposition and temporal-based "pick-and-place" studies. Stereo algorithms for interactive detection and localization are discussed. Finally, a basic error analysis is presented.

In the third part of the text, the design of an automatic vision system for cylinder detection is presented. Using the Marr or Sobel edge detectors, the linear extremal features of cylinders are isolated via the Hough transform. Morphological filters (grey level) are then applied to enhance "feature clusters" in the Hough space. The Hough space is segmented and additional morphological filtering (binary dilation) provides the

region growing necessary to connect the (possibly) segmented feature clusters. From this point, a recursive labeling algorithm isolates individual clusters. Centroids of these clusters are used to represent the linear extremal features present in a given scene. The set of linear features is segmented using unsupervised graph theoretic techniques into subsets that represent the candidates for stereo analysis. A thorough region-based stereo analysis discriminates between the actual cylinders and false alarms caused by the clustering algorithm. Again, a simple error analysis is outlined.

Naturally, some design issues are common to both systems. The first section of the work involves implementation of the *Two Planes* stereo calibration algorithm using a second order approximation. It is followed by an extensive numerical stability analysis. Also some basic ideas used in the feature extraction algorithms use similar geometric constructs. These assumptions and their repercussions are fully developed. Lastly, a hardware design that was used in the implementation of a pipe-cutting system developed for the Department of Energy's Decontamination and Decommisioning project is outlined. The study is concluded with a comparative analysis of the design methods and a presentation of some empirical results.

TABLE OF CONTENTS

CHAPTER		PAGE
1. Introduction		1
1.1	Motivation for the Development of Cylinder Detection Systems	2
1.2	Literature Review	3
1.3	A Man-Machine System for Right Circular Cylinder Detection and Localization	5
	1.3.1 User-Interface Design	7
	1.3.2 Feature-Enhancement Toolkit	8
	1.3.3 Interactive Three-Dimensional Object Localization	9
1.4	Machine Vision for Right Circular Cylinder Detection and Localization	10
	1.4.1 Feature Selection	11
	1.4.2 Edge Detection and Segmentation	12
	1.4.3 Straight Line Detection	14
	1.4.4 Pipe Candidate Identification	15
	1.4.5 Region-Based Stereo	15
2. Hardware Architecture for Vision-Based Localization Systems		**19**
2.1	Stereo Vision Sensor Package	19
2.2	An ASM-Based Hardware Configuration	20
3. Camera Calibration		**23**
3.1	Calibration Mathematics	25
	3.1.1 Derivation of $\mathbf{A}_{nci}$	26
	3.1.2 Extensions of Two-Planes Model	27

CHAPTER PAGE

3.2 Calibration Error Analysis 29
3.3 Transformation Error 30
3.3.1 Numerical Stability of Two-Planes 35
3.3.2 Geometric Significance of the Error 39
3.4 Conclusion on Calibration 40

4. Geometry Error Analysis 42
4.1 Error: The Infinite ViewPoint Model 42
4.2 Error: The Stereo Projection Model 46
4.3 Spatial Quantization Error 48
4.4 Geometry Error Conclusions 50

5. Design of the Graphical User Interface for Interactive Cylinder Detection 51
5.1 Task Analysis of Interactive Cylinder Localization 51
5.2 An Overview of Human Performance 57
5.2.1 Observation and Perception of Reflectance Maps 58
5.2.2 Cataloging the Relevant Features 60
5.2.3 Conclusions on Feature Selection 66
5.3 Constructing the User Interface 66

6. An Interactive Feature Enhancement Toolkit 69

7. Algorithms for Interactive Cylinder Localization 72
7.1 Variations on the Feature Selection Mechanism in Stereo Matching . . . 74
7.2 Experimental Results 75
7.3 Conclusion of Human-Machine Performance in Estimating Cylinder Pose 76

CHAPTER PAGE

8. Automatic Segmentation **80**

8.1 Edge Detection 80

8.2 Threshold Generation for Sobel 80

8.3 Processing the Edge Maps 84

8.4 Conclusions on Edge-Based Segmentation 86

9. Hough Space Processing **87**

9.1 Grey Level Morphology 89

9.2 Hough Space Segmentation 96

10. Boundary Formation by Edge Line Clustering **110**

10.1 Cluster Analysis 111

10.1.1 Pattern Vector Extraction 111

10.1.2 Graph-Theoretic Cluster Analysis 113

10.2 Simple Rules for Candidate Promotion 116

11. Region-Based Stereo **117**

11.1 Scan Line and Artificial Feature Generation 119

11.2 Experimental Results 122

12. Conclusions **125**

BIBLIOGRAPHY **128**

VITA **133**

LIST OF TABLES

TABLE		PAGE
3.1	Geometric error analysis showing calibration uncertainty (in inches) through three different norm functions.	41
5.1	Results of interactive cylinder localization under contrast variation. Cylinders are translated with respect to the coordinate frame with constant standoff of 22.0 inches.	62
5.2	Statistical analysis of contrast data	64
5.3	Results of interactive cylinder localization under temporal variation at constant standoff.	65
5.4	Statistical analysis of contrast data	65
7.1	Translation error measurement in interactive cylinder localization with "point-to-point" matching: standoff = 16.5in.	77
7.2	Translation error measurement in interactive cylinder localization with "point-to-region" matching:standoff = 16.5in.	78
9.1	Forward Histogram Moment Extension Metric	93
9.2	Reverse Histogram Moment Extension Metric	93
9.3	Class Prototype of Constant Theta Vectors	96
9.4	Relative Segmentation Error: Manual vs. Automatic Segmentation	103
9.5	Integration of Segmented Hough Space(in pixels)	105
9.6	Regional Similarity Metric	105
9.7	Sobel H-Space Lines: Manual vs Automatic Segmentation	106
9.8	Marr H-Space Lines: Manual vs Automatic Segmentation	107
9.9	Intra-Cluster Prototype Similarity Metric	107
11.1	Ground truth information for image I.	122

TABLE PAGE

11.2 Ground truth information for image II. 122
11.3 Ground truth information for image III. 123
11.4 Ground truth information for image IV. 123
11.5 Pose Estimation of image I. 123
11.6 Pose Estimation of Image II. 124
11.7 Pose Estimation of image III. 124
11.8 Pose Estimation of image IV. 124

LIST OF FIGURES

FIGURE		PAGE
1.1	System flow diagram for the interactive detection of right circular cylinders. Solid lines depict interaction channels while dashed line show data flow paths.	7
1.2	Object recognition philosophy for human-machine stereo vision systems.	8
1.3	Object recognition philosophy for automated stereo vision systems.	11
1.4	Image Space Ambiguity in Identification of a Cylinder: showing how cylindrical and rectangular objects may be imaged with similar edge features in the correct imaging pose.	13
1.5	Edge Energy vs. Threshold	14
1.6	(*i*)Clusters in Hough Parameter Space, and (*ii*) imaging ambiguity caused from the perspective projection of the linear extremal features of cylindrical and flat objects.	17
1.7	System flow diagram for the automatic detection of right circular cylinders.	18
2.1	Stereo Vision Platform.	20
2.2	The ASM-based machine vision architecture capable of supporting both human-machine and fully automatic cylinder detection systems.	22
3.1	Side view of calibration geometry.	24
3.2	Monocular Two-Planes Calibration	24
3.3	Stereo Two-Planes Calibration: Depicting imaging vectors from each stereo camera pair, and their corresponding intersections with the calibration planes.	25
3.4	Calibration Target	30
4.1	Monocular Viewpoint Error	43

FIGURE **PAGE**

4.2 Stereo Viewpoint Error 43

4.3 Percent Error vs. Standoff 45

4.4 Stereo Mismatch vs. Standoff: $\rho = 1$ 47

4.5 Stereo Mismatch vs. Standoff: $\rho = 2$ 47

4.6 Spatial quantization as a function of sensing element projection and camera focal length 48

4.7 Pixel projection vs. object standoff distance for three commonly used lens systems. 49

5.1 Hierarchical task decomposition of user interface requirements for cylinder detection 54

5.2 A completely rigid fixed sequence control plan. 55

5.3 A more flexible contingent sequence control plan. 56

5.4 An anarchist control plan that allows maximum freedom. 56

5.5 Test scene shown in "good" contrast conditions. 63

5.6 Test scene shown in "low" contrast conditions. 63

5.7 Graphical user interface for interactive cylinder detection and localization. 67

6.1 A typical input image to before image enhancement. 70

6.2 Histogram equalized version of the input image 71

6.3 Edge overlay of input image after equalization. 71

8.1 Input to Cylinder Detection System 81

8.2 Input to Cylinder Detection System 81

8.3 Sobel-Based Edge Detection 82

8.4 Binarized Sobel Edge Map 85

8.5 Marr-Hildreth Edge Detection: $\sigma = 1.0$ 85

9.1 Hough Transformation of Edge Maps 88

9.2 Histogram of Hough Maps 88

FIGURE PAGE

9.3 Hough Space Morphology 91
9.4 Histogram of Dilated Hough Maps 91
9.5 Manual Segmentation of Hough Space 97
9.6 Binary Region Growth in Manually Segmented Hough Space 97
9.7 Automatic Segmentation of Hough Space 99
9.8 Binary Region Growth in Automatically Segmented Hough Space 99
9.9 Labeled Hough Space 100
9.10 Hough Space Cluster Selection Measure: Manual Segmentation 100
9.11 Hough Space Cluster Selection Measure: Automatic Segmentation . . . 101
9.12 Manual vs. Automatic Segmentation of Hough Space 101
9.13 Synthesized Hough Space Line Images 109
9.14 Original Image with Synthesized Line Overlays 109
11.1 Basic stereo imaging geometry 118
11.2 Solution to point of minimum distance stereo triangulation problem via two planes calibration procedure 120
11.3 Solution of stereo correspondence problem showing scan-line and feature generation for foreshortened cylindrical object. 121

CHAPTER 1

Introduction

In this text, two fundamentally different philosophies for attacking problems in machine vision are explored in the context of developing vision systems for a robotic pipe-cutting and scanning operations at the Oak Ridge National Laboratory. The development of a human-machine vision systems for the detection and localization of arbitrarily oriented pipes, or right circular cylinders, relates to the first school of thought. This school believes that computing systems can be significantly improved by designing into them the capabilities of the user, and likewise adding to the capabilities of the user through an applied knowledge of human-machine interaction[She89a]. Recent advances in cognitive psychology and related sciences tend to support this hypothesis[CMN83]. The application of human cognitive science to goal-oriented engineering problems has and will continue to improve the tri-fold understanding of the human, the computer, and their interaction.

Development of completely autonomous systems motivates the second ideology. Here, the impetus becomes removing the supervision from the system. The designer attempts to instill in the application controller the necessary intelligence, flexibility, and fault tolerance to perform a task without the need for human intervention. A second system exploiting autonomous machine vision attempts to detect and localize right circular cylinders; accordingly, it is an analysis of the techniques and procedures necessary for solving problems via the second school of thought. No attempt is made to prejudge the success of these separate systems based on the prejudice of the author; rather, an attempt will be made to draw conclusions based on the empirical results derived after ex-

perimentation with each of the implementations. Obviously, both will have weaknesses and strengths that will be cataloged and outlined to illustrate the overall effectiveness of each system. It is also recognized that certain critical steps in each system may benefit from the application of its competitor's technology. To illustrate this, a hypothetical "hybrid-system" is constructed that uses, at each step, the algorithms and procedures of the ideology that works best for it. In this way, a composite of techniques from each school could be compiled into a localization system that is unparalleled by either of its ancestors.

An attempt is also made to isolate from these two separatist ideologies the modules that are common to them both. For example, any application that uses vision to affect a three-dimensional localization requires a calibrated camera system. In this case, a pair of calibrated stereo cameras is the fundamental to both cylinder detection systems. Also both application philosophies due to their goal-oriented and task specific natures make certain approximations and assumptions about the environment with respect to that task. By definition, these characteristics often arise from the very nature of the physical problem, not from any particular approach to solving it. Considerations of this type may lead to the establishment of some common ground between the two schools of thought. Naturally, this commonality could provide valuable insight in any future machine vision work.

1.1 Motivation for the Development of Cylinder Detection Systems

One of the most common requirements of any scientist is the justification of his work. Aside from a totally academic pursuit, it is desirable that the effort produce some benefits in an application or working prototype or perhaps it fills some gap in the technology when completed. Decontamination and Decommissioning (D&D) of old nuclear facilities is an important requirement of the U. S. Department of Energy's En-

vironmental Restoration and Waste Management (ER&WM) program[oE89]. Methods need to be devised that minimize radiation exposure to humans involved in the D&D operations. Examples of such operations include remote inspection, radiation scanning, work-cell mapping, and decommissioning of nuclear hot-cells. The research presented in this paper describes (*i*) a human-machine and (*ii*) an automatic system which can be employed in radiation scan and pipe cutting operations in a typical D&D work cell. Using an existing telerobotic manipulator, the Advanced Servo Manipulator (ASM) for the Oak Ridge National Laboratory's Robotics and Process Systems Division, either system, or a combination of both, could be used for these D&D operations. Some initial testing on the manipulation capabilities of each approach is presented.

1.2 Literature Review

Any work in science or engineering must be built upon the infrastructure of existing relevant research in the field, and a competent technical report must include a review of the relevant research issues. Consequently, a review of the appropriate literature is presented. The review is restricted to a few sample cases of the relevant work involving two- or three-dimensional cylinder detection and localization.

Mohtadi, et. al., [MSS92] describe a machine vision system for the detection of tubes in a "noisy industrial" environment. In their system, tube diameters range from 20mm to 200mm, and the imaging environment is characterized by a noisy background. Their methodology employs a sampling of the image space in thin bands followed by the application of a normalized correlation filter derived from a reference template. Image space sampling results in a reduction of the model size and an increase in algorithm efficiency. Template generation, by necessity, must have knowledge of the tube diameter, which in their application is readily available. They define a tube after application of the filter as a peak characterized by the following properties:

- Coordinates of local maxima in the processed image band.
- A reliability factor based on peak height and filter background response.
- A local estimate of tube orientation based on grey level uniformity.

From here, peaks are mapped directly into a Hough accumulator array with an accumulator size of 1 degree. From a Hough space analysis, it becomes easy to determine the relative position of the tubes with respect to the reference frame and the method is unaffected by missing peaks in the accumulator array. At this point, the Hough space is smoothed to join close maxima, and a global threshold is applied. Next, a labeling algorithm is applied and the centroids of the individual labeled regions are used for the representation of the central lines of the tubes. In simulation results with 500 synthetic and 70 real images, they report a 100% detection rate.

Hallset [Hal91] reports a two-dimensional vision-based pipe-tracking system for an autonomous underwater vehicle. Hallset's technique uses a model based recognition strategy where symbolic descriptions of real world objects are formed by solid geometric primitives. A system of pipes, or a pipeline, is described in a reference map by a scalar indicating the diameter of the pipe section and a piecewise linear approximation to the network. The first step in the vision processing involves a histogram equalization to enhance the contrast between the object, a pipe section, and the background. Next, a $5\times$ 5 Sobel edge detection and global thresholding process is applied to the equalized image. After edge detection, a traditional recursive region growing algorithm is applied. The last stage in the vision processing is the fitting of a rectangle to each of the labeled image regions from the recursive segmentation process. From here, the work deviates from classical vision techniques and uses other sensing modalities to obtain an approximate scale factor for the world-to-image transformations and the approximate heading of the camera system. Using information integrated from these modalities, the reference map

is updated. No numerical results pertaining to the accuracy of the localization process were reported; however, initial results were described as promising.

Shiu, et. al., [SZ92] address the determination of the pose of spheres, right-circular cylinders, and right-circular cones for perspective projections. Again a model-based framework is introduced. Using elliptical projections from the circular surfaces of a cylinder, the three-dimensional pose is calculated. Also, a method is described to compute the pose of a right circular cylinder from its linear extremal contours using techniques from analytic geometry. Unique solutions are found only if the terminations of linear features are known. In the linear case, it is stated that a knowledge of termination points will produce a six degree-of-freedom geometric representation of a cylinder; however, if no termination points are located a degree-of-freedom remains along the central axis, or axis of least inertia, of the cylinder. Using the linear extremal features, two planes that pass through these lines and the focal point are constructed. The cylinder axis can be computed by taking the cross product of the normals of these two planes. The diameter of the cylinder can be expressed by taking the distance between a reference point of the cylinder axis and the point of minimum distance in the tangent plane. No implementation or results were reported.

1.3 A Man-Machine System for Right Circular Cylinder Detection and Localization

A significant portion of the research in telerobotics has been directed to a study of remote maintenance and inspection of nuclear hot cells, particularly in the use of telerobotics to reduce the potential hazards to human beings and the environment in the handling of radioactive materials[Her91, She89b, oE89]. Man-machine systems offer the inherent flexibility and intelligence of a human worker while simultaneously removing that worker from the hazardous work environment. Such an application is characterized

by a robot in an unstructured environment in which the precise locations of objects are not known in advance [HST93]. Development of generic sensory and software tools for these tasks creates tremendous difficulties for the man-machine system designer. Variations in manipulator geometry, available sensory hardware, and task goals prohibit a generalized applications package for all but the most basic endeavors.

The primary advantage of man-machine systems is solving problems that traditional autonomous implementations experience. Most of these problems concern the application of artificial intelligence to derived sensory data. Many times this approach has encountered technical difficulties as the needed artificial intelligence proved more difficult to realize than originally expected[Utt89]. Unfortunately, solutions to these problems usually cannot wait until "tried-and-true" technology for a particular application exists. In such situations, exploitation of inherent human understanding skills is merited. Our task now becomes the application of human interpretive capabilities to derived sensory data, in effect by-passing specific autonomous bottlenecks. This application, although general-purpose in nature, can best be served by tailoring the choice of sensory devices to the task at hand. By designing this into a system, the remote operators can be offered the optimal set of sensory data, while minimizing the costs and complexity required for a general purpose applications package.

A natural extension to this philosophy is the enhancement of the raw sensory data. Advantage can be taken of existing computer graphics and machine vision technology to offer the remote operator an interactive sensory data enhancement toolkit. From this toolkit, the operator can implement well-defined feature enhancement tools in an interactive fashion. Using algorithms for interactive object localization, a remote supervisor can quickly and reliably extract information necessary for three-dimensional object localization. Once the necessary information has been extracted from the enhanced sensory data pools, this model-based three-dimensional data can be applied to

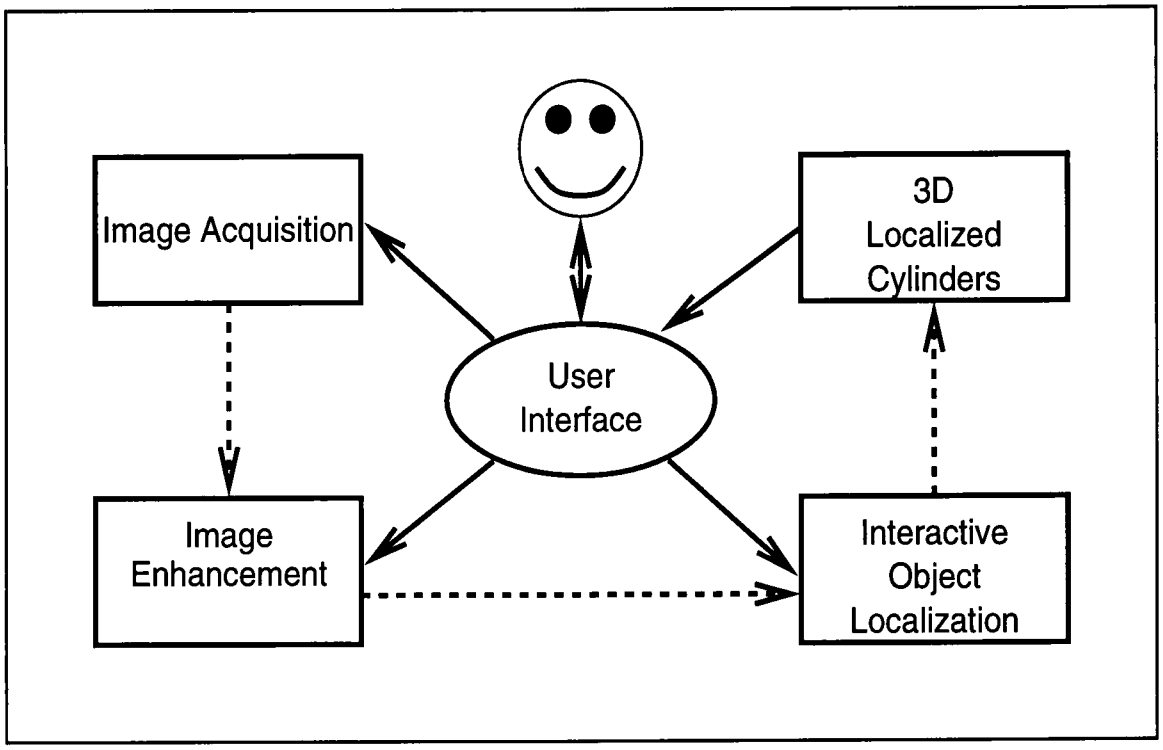


Figure 1.1: System flow diagram for the interactive detection of right circular cylinders. Solid lines depict interaction channels while dashed line show data flow paths.

motivate a typical D&D task. Figure 1.1 conceptually displays the interaction of a human operator with each of the system components via the user-interface. Figure 1.2 represents the design philosophy of the interactive vision system where high-level tasks are left to the human operator and low-level, number crunching jobs are given to the machine.

1.3.1 User-Interface Design

A first step in the design of a vision-based man-machine system for three-dimensional object localization is the design and implementation of an adequate user interface. The term "user interface" refers to the combined environment of data presentation using output devices and different interaction methodologies using input devices [KY91]. The

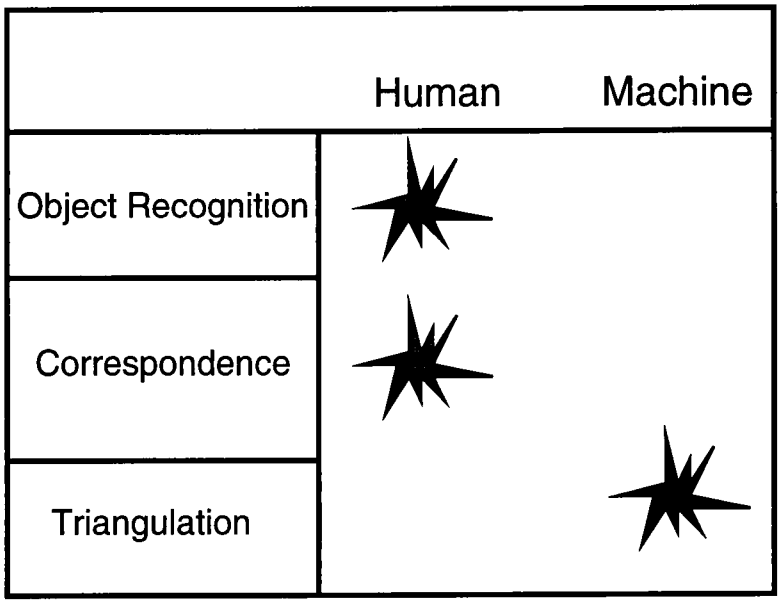


Figure 1.2: Object recognition philosophy for human-machine stereo vision systems.

primary goal of the user interface should be to enhance productivity in the application task through a simplification of the interaction between the data and the user. Therefore, a necessary step in user interface development is the task analysis. *Task analysis* is the process of trying to make sense of what people should do or what they actually are doing [She89a]. Once the appropriate task analysis has been performed, the user interface can be tailored to fit the application. Next, some decisions must be made relating to the form in which the sensory data is to be presented to the user. From this analysis the information displays may be added to the user interface.

1.3.2 Feature-Enhancement Toolkit

If the images presented to the human operator are degraded by some environmental conditions, a set of image enhancing tools becomes useful. Improving the quality of an image for human interpretation is a highly subjective process [GW87]. The operator should have complete freedom over the form and fashion that these tools are to be

applied and should be able to localize the effects of a particular operation to a desired image region. The effects of these operators on these regions should be cumulative. The operator must also be allowed to undo the effects of the previous operations on an image area. The basic software package consists of

- Variable mask median filtering
- Variable mask neighborhood filtering
- Pseudo-color enhancement
- Edge detection toolkit
- Thresholding & segmentation

1.3.3 Interactive Three-Dimensional Object Localization

A common approach for passively acquiring range data, *binocular stereo*, uses the difference between corresponding image points obtained from different views of the same scene. The key to success in such a stereo system requires a robust method for determining these corresponding points. Thus, the *correspondence* problem is that of identifying features in two images that are projections of the same entity in the three-dimensional world [Hor89]. Current automated techniques for solving this problem include cross-correlation, gray-level matching, sum-of-squared difference (SSD), and edge- or region-based methods [FGL87, Hor89, MT93, MT89, OK93]. These methods demand large amounts of computation. Even with current computing technology, the calculations may take minutes or even hours of CPU time depending on the correspondence strategy. Computational problems can be aggravated by the changes in perspective necessary for the stereo system [Nof85]. Simplifications of the stereo correspondence problem have been proposed. Usual techniques for effecting this simplification involve the projection

of a structured pattern of light onto the scene. This pattern can then be used as a basis for selecting candidate points [Nof85]. Although the computation time for selected points drops dramatically, the amount of range data recovered is severely limited. While interpolation can approximate the missing data, the quality of the interpolated data may not be appropriate for certain machine vision applications.

A human-machine system for cylinder detection requires algorithms to determine quickly and reliably the location of the target cylinder in a cluttered work space. Assuming that we could derive range maps from an automated stereo vision system at rates suitable for a human operator, the problem of identifying cylinders from an unstructured scene introduces a new set of difficulties. While automated systems for solving this type of machine vision problem certainly merit study, a human-machine stereo vision system offers a highly robust, efficient and inexpensive alternative. Drawing on years of experience, a human operator can use highly developed pattern recognition skills to match features in corresponding images. Exploiting the model-based nature of the application, an efficient computational framework can be introduced. In this way, the stereo correspondence, object identification, and object location problems are solved simultaneously. The human operator becomes responsible for high-level feature matching assisted by the computer in the number-crunching tasks.

1.4 Machine Vision for Right Circular Cylinder Detection and Localization

Humans are able to view an image that may contain a virtually limitless variety of objects that can be intuitively recognized. Generally human perception of an object in a scene is unaffected by rotational, scale, or translational variations. However, when similar operations are performed on a digital computer the presence of these variations coupled with additive noise require levels of complexity that may prohibit the develop-

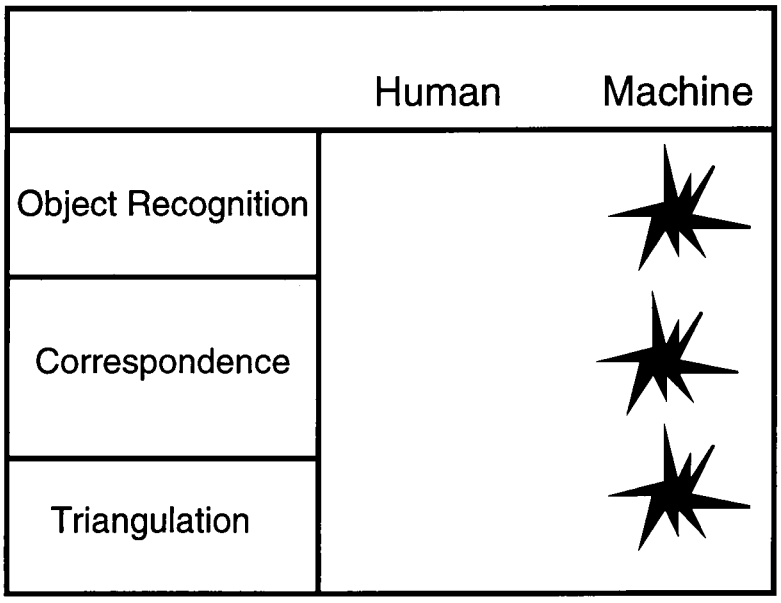


Figure 1.3: Object recognition philosophy for automated stereo vision systems.

ment these types of systems. Certain simplifications or assumptions about the scene and the task goals reduce this complexity. Our study is limited to the identification of cylindrical objects, in a three-dimensional scene. Now that the task specification has been sufficiently limited, attention is turned to extracting the best possible features for locating pipes in an image. Figure 1.3 shows the design philosophy of an automatic vision system with respect to three fundamental vision problems. Note that in this case all the problems are solved by the computer, whereas in the human-machine case shown if Figure 1.2 only stereo computation is handled by the computer.

1.4.1 Feature Selection

As in any pattern recognition problem, the most difficult task involves the identification of a complete set of discriminatory features from the observed data[TG74, DH73]. Ideally, a set of features that uniquely defines a cylinder needs to be isolated. No assumptions concerning the texture, orientation, diameter, etc. are made. From ele-

mentary calculus, a right circular cylinder is defined by the following: *Let* **C** *be a curve in a plane and let* **l** *be a line that is not in a parallel plane. The set of points on all lines that are not parallel to* **l** *and intersect* **C** *is a* **cylinder**. *A right circular cylinder, or a pipe, is such a set of points where* **l** *is perpendicular to the plane, and the curve* **C** *is a circle*[Swo84]. Using this basic geometrical description, a simple set of features that describes (not uniquely) the projection of an arbitrarily oriented pipe in image space was derived. In Hough or image space, converging sets of roughly parallel lines are tagged as pipe candidates. A simple visual example shown in Figure 1.4 reveals the inherent flaw in this logic. From the figure, the projection of the cylinder and the ruler in image space becomes a set of convergent lines. Stereo processing, however, will reveal the curved geometric form of the actual cylinder and discriminate between the two.

1.4.2 Edge Detection and Segmentation

Assuming the set of features that will best describe an object as a possible right circular cylinder will be a set of parallel or convergent lines, methods must be devised to produce from the reflectance map a binary map of edges in the scene. The *Sobel* and *Marr* operators were chosen as the primary edge detection algorithms. The Sobel was attractive because of its inherent noise tolerance, a by-product of a smoothing effect, and it does not require a thresholding step[GW87]. From the gradient operation, a raw grey level edge map is created. The raw edge map (Sobel) was binarized using a energy-based segmentation scheme. In this segmentation algorithm, a thresholding of the raw edge map over each of the possible intensity levels occurs. After this thresholding, a binary integration over the image space is performed. The resulting sequence of *edge energy* values is plotted on a graph as a function of the interim threshold value. A sample plot is show in Figure 1.5. The *knee* of the plotted curve is chosen as the global threshold for the raw edge map. A complete analysis of this segmentation process is discussed

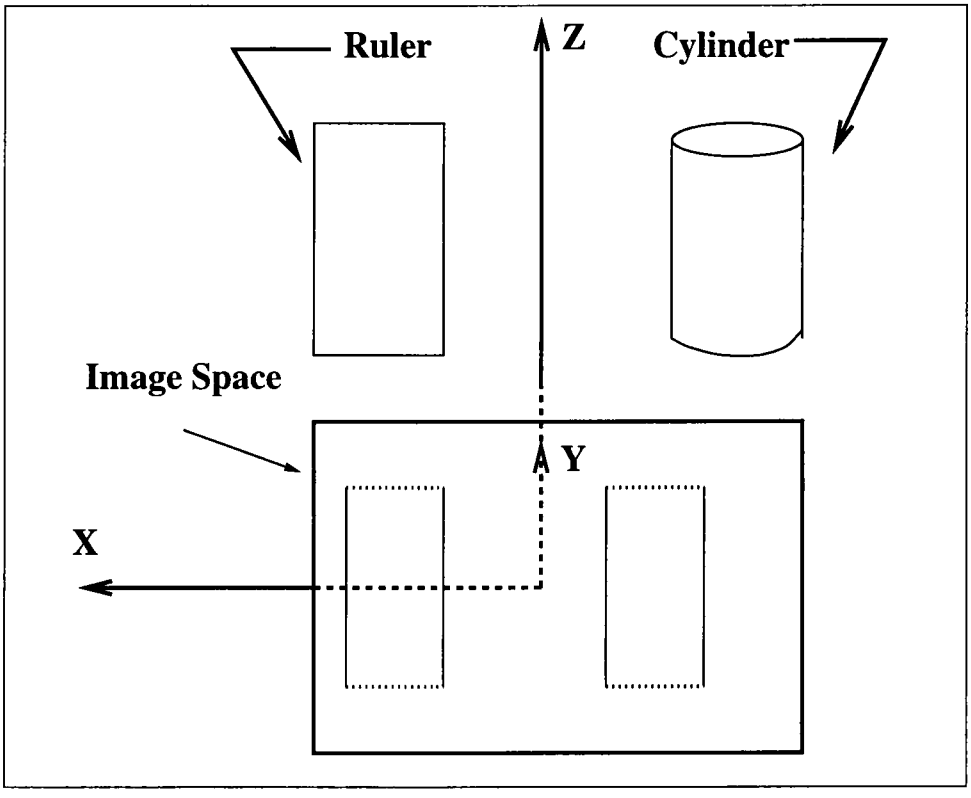


Figure 1.4: Image Space Ambiguity in Identification of a Cylinder: showing how cylindrical and rectangular objects may be imaged with similar edge features in the correct imaging pose.

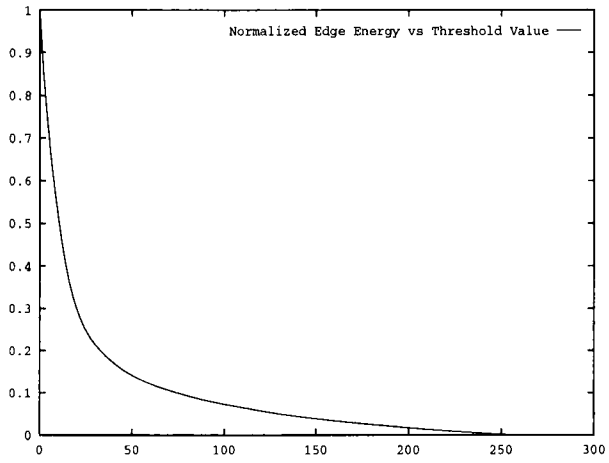


Figure 1.5: Edge Energy vs. Threshold

in Chapter 8. At this point in the algorithm, a binarized edge map of the image space exists. The next step is the identification of lines from this edge data.

1.4.3 Straight Line Detection

The Hough transform is a well-known method for the detection of parametric curves, in this case lines, in binary images[PLJ92, GW87]. Advantages of the Hough transform are its robustness to noise and discontinuities in the curve; however, storage requirements and computational costs can be prohibitive[Ati92, PLJ92]. Edge pixels, based on their spatial locations, are indexed into a quantized polar parameter space. In this space, compact clusters may be thought of a sets of collinear points[VC92]. Figure 1.6 exhibits several of these clusters in a 512×480 (pixels) parameter space. Examining the figure reveals four or five primary clusters located horizontally along the center line of the image. Also, four secondary clusters are located near the corners of the image. Some process must localize these clusters and separate them from the background. This

Hough space representation of edge lines is inadequate for selecting the feature set. Venketaswar and Chellapa suggest using the Hough Transform as an initial pass for generating groupings of collinear, but not necessarily connected edge pixels[VC92]. For this application however, the tendency of the transform to fit lines without regard to spatial contiguity is desired. In this way, occlusions are handled automatically by the standard transform and require no further attention. What is required, however, is modification of the raw Hough space to produce sharp peaks near the inertial center of significant clusters and to remove spurious background noise. From this transformed space, morphological filtering sharpens line clusters into distinctive peaks.

1.4.4 Pipe Candidate Identification

The final task in monocular pipe candidate detection is an exhaustive search of the modified Hough space for pairs of convergent lines. Simply identifying pairs of lines as pipe candidates is not adequate as show in Figure 1.6. Any of the regions **A**, **B**, and **C** satisfy the initial constraint of being bounded by convergent set of lines. Obviously, however, only the regions **A** and **C** can physically be pipes. Application of these physical constraints reduces the number of false matches in most cases that were studied. Stereo processing can also reveal false matching in high clutter images where the simple assumptions breakdown.

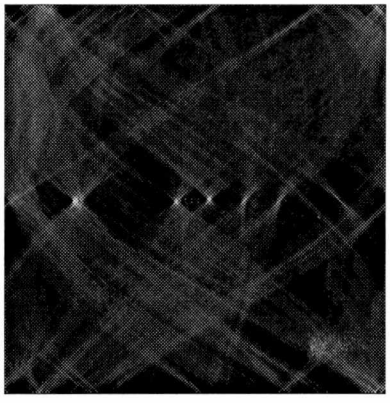
1.4.5 Region-Based Stereo

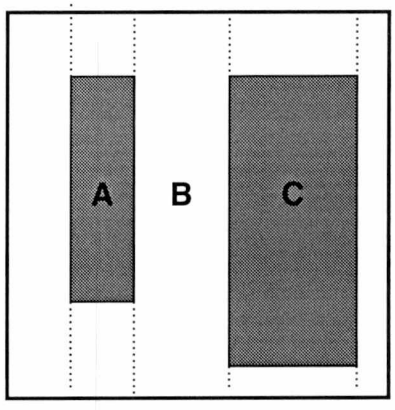
Input to the Region-Based Stereo consists of images obtained from the geometrical arrangement of cameras depicted in Figure 2.1. The third camera, not shown in the figure, is mounted in the lateral center of the platform with a transverse displacement of 4.25 inches. Matching is not performed using Okutomi and Kanade's method of computing the sum of squared-difference values; rather, a normalized cross-correlation

was substituted[OK93].

Ambiguity between the pipe candidates is resolved using the definition of a right circular cylinder from elementary calculus. If it is assumed that a particular pair of lines isolated by the segmentation and transformation processes previously described was actually derived from a pipe present in the scene, and it is further assumed that a line segment defined by two points, one from each line, passes through the center of the pipe. (This approximation, valid within certain constraints that are primarily functions of baseline and object distance, is distance outlined in Chapter 4), Then within some boundary, a general class of disparity functions is implied. The second assumption really states that from any arbitrary viewpoint exactly one half of the cylinder is imaged. This idea is similar to what Storjohann proposed in a paper on the inverse perspective mapping[Sea90]. The structure of the class of disparity functions, or conversely distance functions, has no meaning at the pixel level. However, if a structuring element is imposed on the region, over some line segment perpendicular to both edge lines, then it can be seen that the disparity function will take on an identifiable pattern if the object in view is a cylinder. Using this idea, the ambiguity between the set of possible pipe candidates is resolved and the true pipes are exposed. Figure 1.7 contains a block diagram representation of the complete automatic pipe detection system.

In the next chapters, the design and calibration of a camera system intended for a area-based stereo system is explored. A mathematical representation of the calibration procedure and an error analysis are outlined. Also, some new derivations of the *Two-Planes* calibration method that are readily applicable to the Storjohann's inverse perspective mapping procedure are presented[Sea90, IPS87]. The condition of the resultant transformation matrices is also examined.





(a) Hough Parameter Space (a) Ambiguity in Region Matching

Figure 1.6: (*i*)Clusters in Hough Parameter Space, and (*ii*) imaging ambiguity caused from the perspective projection of the linear extremal features of cylindrical and flat objects.

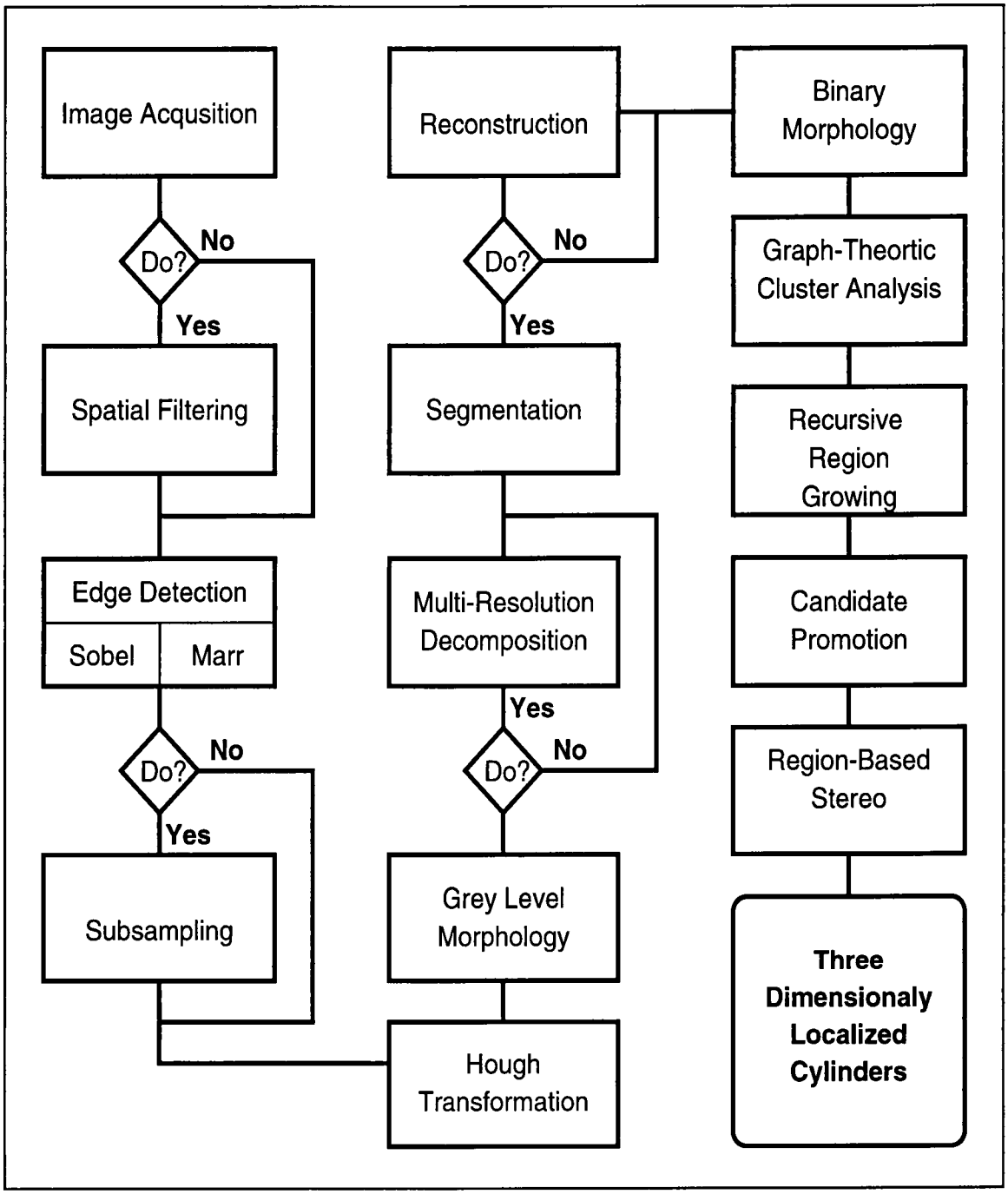


Figure 1.7: System flow diagram for the automatic detection of right circular cylinders.

CHAPTER 2

Hardware Architecture for Vision-Based Localization Systems

Implementation of three-dimensional vision systems, by necessity, must begin with the design of the necessary sensing and computing hardware system. In general, it would be appropriate to design into the hardware system a degree of independence from the task and the robotic manipulator intended to affect that task. Although the ASM is intended to be the primary user of this vision systems, only a small portion of the hardware system is devoted to it. In fact, the sensor package, the vision software, and the communications system are completely manipulator independent. The vision system is completely portable to other architectures with minimal effort; however, if some control or manipulation is desired upon porting to a new robotic architecture, a new control interface module must be designed for each new manipulator. The following sections briefly describe the sensor package layout and the hardware architecture.

2.1 Stereo Vision Sensor Package

Two Panasonic GP-MF502 black-and-white CCD cameras where used in the stereo vision system. A Lasaris variable-diameter, infra-red, circle-projecting laser complements the stereo unit. Simplification of the sensor geometry was achieved by forcing the individual units to lie on the same lateral axis while maintaining nearly parallel transverse axes. The laser occupies the geometric center of the lateral axis with the cameras centered around it. In general, this configuration supports the stereo-scopic view of the laser light required to offer adequate backup support and structured visual queues in future work.

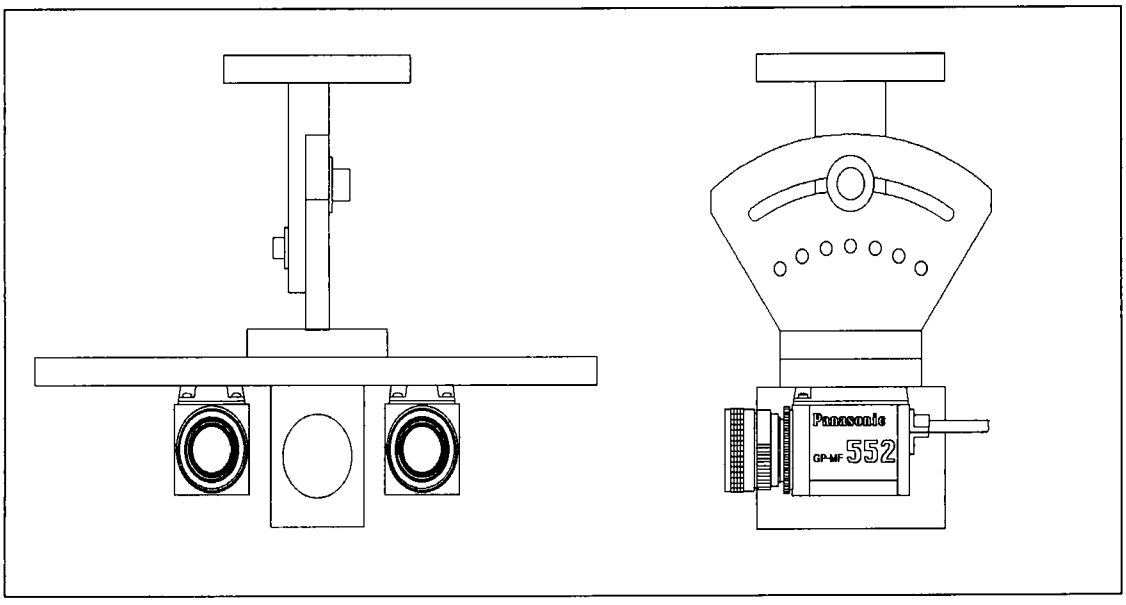


Figure 2.1: Stereo Vision Platform.

Mounting bracket design includes two degrees of freedom along the lateral baseline and pitch on the transverse axes. Figure 2.1 shows each of the sensory units in one of the possible arrangements on the mounting bracket. Both degrees of freedom can be manually changed to rigid discrete positions. By restricting these movements to well-known locations, accurate calibration of the sensing package becomes a simple matter of describing a transformation between a point in three-dimensional space and a static reference frame on the mounting bracket. Transformations between this reference frame and the individual sensory units become simple translations.

2.2 An ASM-Based Hardware Configuration

The Advanced Servo Manipulator is a six degree-of-freedom, remotely maintainable, force-reflecting master-slave teleoperator developed for remote maintenance and handling studies[BG89]. The communications and control of the existing ASM system

were developed on a Multibus system. The transparency of the cylinder localization system was realized by tapping directly into the existing ASM arm controller using a VME-to-Multibus repeater. Hence, when the system is not in use or has been ported to another manipulator, arm control sequences will naturally be derived from the ASM master manipulator. Aside from the existing ASM hardware, the localization system includes VME- and Silicon Graphics based computing environments. The Silicon Graphics workstation and the VME system are connected using a VME-VME bus repeater to allow high-speed communications. A two megabyte dual-ported memory board offers shared memory between the three computing architectures. Lower level communication and control functions reside on a VME bus rack. The VME sub-system consists of two 68040 processors, Datacube image capture and memory boards, and multiple bus repeater links. These are controlled using the VxWorks real time operating system. The master 68040 board is used to handle communications, bus arbitration, shared memory access, and image acquisition. The remaining 68040 board is solely responsible for synchronizing arm control with the Multibus systems for the ASM test case. An RS-232 serial port provides an interface to a radiation sensor. A cut tool control linkage is accomplished using a 68040 parallel port.

High-level human-machine interaction occurs on the Silicon Graphics 310VGX workstation. The human-machine interface, interactive object localization, and graphical task modeler systems exploit the capabilities of the Unix-based graphical workstation. The autonomous vision system is not restricted to reside on the Silicon Graphics machines and can be operated from a Sparc or similar unix system. Each of these uses the X-Windows based Motif widget package for the application interface. All code was written using a C++ Translator. Figure 2.2 displays a block diagram of the system hardware configuration.

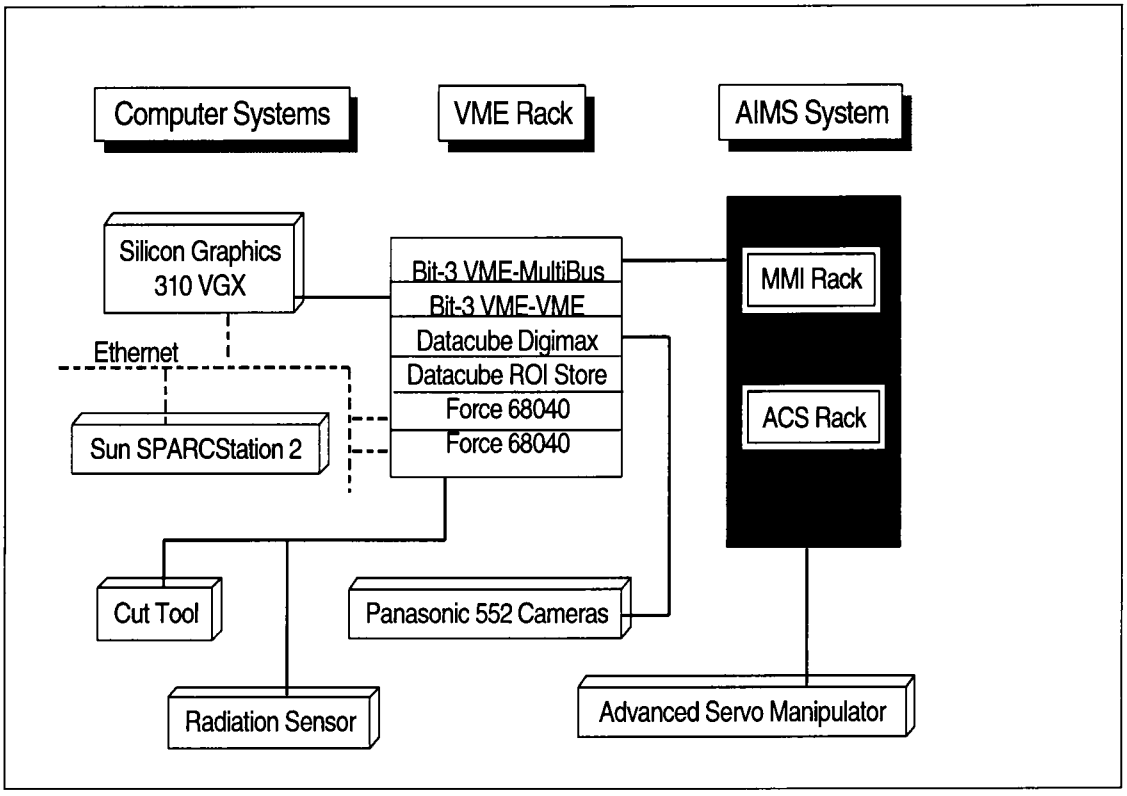


Figure 2.2: The ASM-based machine vision architecture capable of supporting both human-machine and fully automatic cylinder detection systems.

CHAPTER 3

Camera Calibration

Accurate calibration of camera systems is an essential first step in a computer vision system requiring three-dimensional measurements[WCH92, IPS87, HN91, Nom92]. There are two general philosophies for attacking this problem. The first methodology concentrates on experimentally determining internal camera parameters[WCH92]. In this scheme, knowledge of first-order camera parameters and possibly distortion models work through the perspective transformation to uncover a relationship between image space and three-dimensional space. The second methodology, however, requires specific knowledge of external parameters (i.e. a calibration scene geometry) and uses this to provide a derived transformation between image and world space. Here, the principle technique uses a nonlinear least-squares algorithm to extract this transformation empirically from the environment[IPS87, HN91]. The attractiveness of this method stems from its ability to produce accurate calibration without knowledge of internal camera parameters. Calibration is achieved by fixing the scene geometry to a known accuracy, as in Figure 3.1, and deriving these transformations from it. The Two-Planes method as described by Izaguirre[IPS87] was chosen for these reasons.

Conceptually, Two-Planes calibration originates from the idea that a line from a reference point, W_{ref}, to a point in a scene will intersect a plane at a unique point. When an additional plane is added to the scene, as in Figure 3.2, a vector emanating from the image plane in W_{ref} space is described, assuming that W_{ref} is the origin of a known three-dimensional coordinate system. The power of this concept becomes clear. If the transformation between two physically known planes in calibration space and a

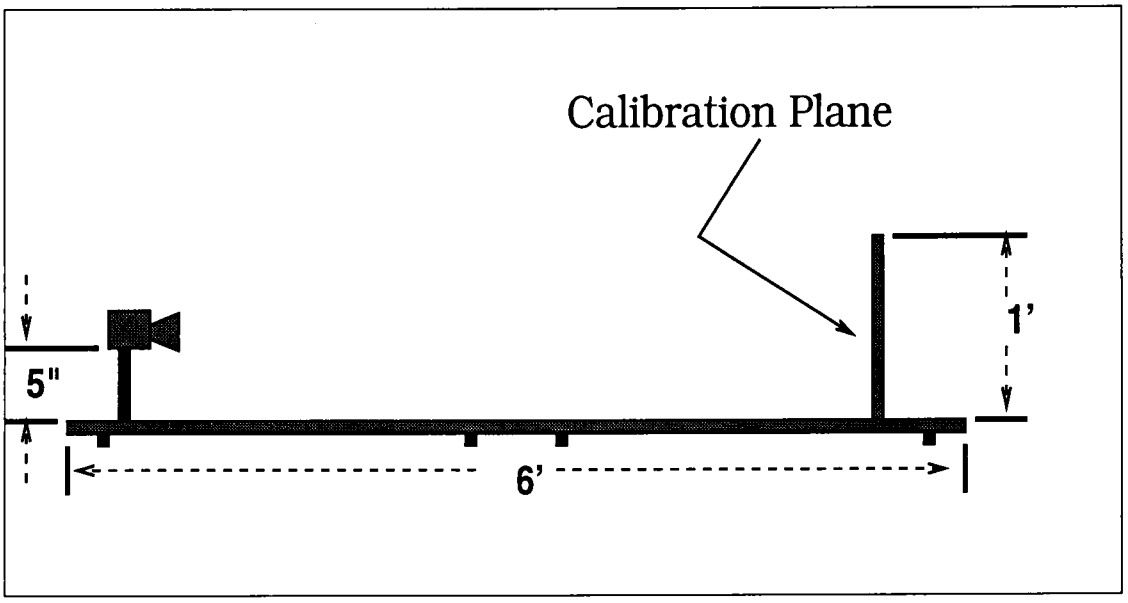


Figure 3.1: Side view of calibration geometry.

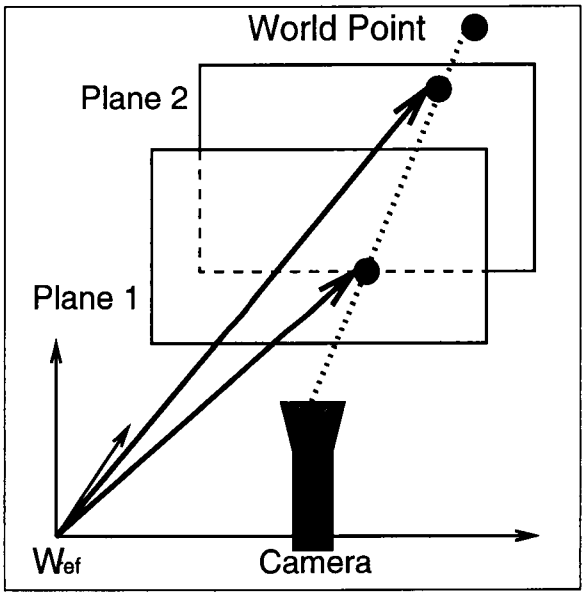


Figure 3.2: Monocular Two-Planes Calibration

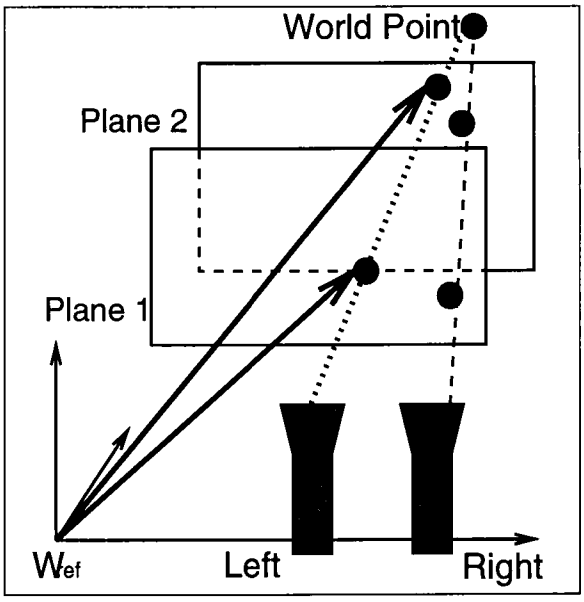


Figure 3.3: Stereo Two-Planes Calibration: Depicting imaging vectors from each stereo camera pair, and their corresponding intersections with the calibration planes.

world reference frame is known, a point in three dimensional space can be described by a vector in W_{ref}. This description is expected from this type of transformation. In essence, it is a projection of a three-dimensional point onto a two-dimensional image space. Necessarily, some information will be lost. This method lends itself readily to stereo vision problems. Adding another camera to the scene will produce two vectors in the reference space for each three-dimensional point. Solving for the intersection of these vectors describes uniquely that point in space. Figure 3.3 conceptually displays the application of the Two-Planes method to stereo vision.

3.1 Calibration Mathematics

It was stated that the exact location of two planes in the world coordinate frame must be known. Therefore for each world plane, a transformation that describes the geometrical relationship between points in the image plane and points in the each cali-

bration plane will be derived. Using standard linear interpolation a general form of this transformative process is

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{\text{plane } i} = \mathbf{A}_{nci}(x^n, x^{n-1} \times y, \ldots, y^n, \ldots, x, y, 1)^T, \tag{3.1}$$

where i identifies the particular world pane; n is the order of the polynomial fit; $(XYZ)_i^T$ are the coordinates of a world point from plane i; $(x, y)_c$ are the coordinates of the same world point in the image plane of camera, c; and A_{nci} is the interpolation matrix for the plane i and the camera c. Using matrix notation this expression may be written as

$$\mathbf{P} = \mathbf{L}\mathbf{A}_{nci}^T. \tag{3.2}$$

Here, $\mathbf{P}$ is a $3 \times N$ matrix of world points, and $\mathbf{L}$ is a $N \times n$ polynomial interpolation matrix. In calibration $\mathbf{P}$ and $\mathbf{L}$ are known. Using matrix techniques the interpolation matrix $\mathbf{A}_{nci}$ is derived in the next section.

3.1.1 Derivation of $\mathbf{A}_{nci}$

Given a matrix of world coordinate points from plane i with respect to W_{ref} and their corresponding projections into camera c, Equation 3.2 is left with one unknown. Theoretically, the linear solution should exist. In most cases, the matrix $\mathbf{L}$ is not square; therefore, a simple premultiplication by its inverse is prohibited. It is well-know that the multiplication of a matrix by its transpose produces a square matrix. Making use of this and premultiplying Equation 3.2 by the transpose of $\mathbf{L}$ produces

$$\mathbf{L}^T\mathbf{P} = (\mathbf{L}^T\mathbf{L})\mathbf{A}_{nci}^T. \tag{3.3}$$

The placement of the parenthesis in Equation 3.3 immediately suggests the next operation. If the matrix $\mathbf{L^T L}$ is well-conditioned, its inverse, $\mathbf{L^T L}^{-1}$, will exist. Premulti-

plication by this the this inverse, called the *Moore-Penrose Pseudo Inverse* gives

$$(\mathbf{L^T L})^{-1}\mathbf{L^T P} = (\mathbf{L^T L})^{-1}(\mathbf{L^T L})\mathbf{A}_{nci}, \tag{3.4}$$

which reduces to

$$(\mathbf{L^T L})^{-1}\mathbf{L^T P} = \mathbf{A}_{nci}. \tag{3.5}$$

This process is repeated for the second calibration plane and again for each additional camera that is to be calibrated. Additional planes may be added to demark regions of operation where only a local transformation is valid. An analysis of the error inherent in this least-squares approximation for experimental data highlights Section 2.2. At this point, the camera model contains a pair of transformation matrices that when presented with a vector of image space coordinate will define a vector in the reference space.

3.1.2 Extensions of Two-Planes Model

Thinking about the model involved in two planes calibration, then from Equation 3.2 it is logical that a relationship between the two transformation matrices exist with respect to a single calibration plane. Assume that two cameras with convergent optical axes are focused on a point in the calibration plane. Further, assume that a complete and accurate calibration of the system has been performed. The Equations that describe the system are

$$\mathbf{P} = \mathbf{L}_{11}\mathbf{A}_{n11}^{T}, \tag{3.6}$$

and

$$\mathbf{P} = \mathbf{L}_{12}\mathbf{A}_{n12}^{T}. \tag{3.7}$$

$\mathbf{L}_{11}$ and $\mathbf{L}_{12}$ represent the interpolation matrices of Plane 1 for the first and second cameras, respectively. Since the vector $\mathbf{P}$ in Equations 3.6 and 3.7 represents the same physical point, the expressions can be equated, resulting in

$$\mathbf{L}_{11}\mathbf{A}_{n11}^{T} = \mathbf{L}_{12}\mathbf{A}_{n12}^{T}. \tag{3.8}$$

Premultiplying this by $\mathbf{L}_{11}^T$ produces

$$\mathbf{L}_{11}^T\mathbf{L}_{11}\mathbf{A}_{n11}^T = \mathbf{L}_{11}^T\mathbf{L}_{12}\mathbf{A}_{n12}^T \tag{3.9}$$

and applying the pseudo-inverse techniques from Equation 3.4 offers

$$(\mathbf{L}_{11}^T\mathbf{L}_{11})^{-1}(\mathbf{L}_{11}^T\mathbf{L}_{11})\mathbf{A}_{n11}^T = (\mathbf{L}_{11}^T\mathbf{L}_{11}^{-1})(\mathbf{L}_{11}^T\mathbf{L}_{12})\mathbf{A}_{n12}^T \tag{3.10}$$

this reduces to

$$\mathbf{A}_{n11}^T = (\mathbf{L}_{11}^T\mathbf{L}_{11}^{-1})(\mathbf{L}_{11}^T\mathbf{L}_{12})\mathbf{A}_{n12}^T \tag{3.11}$$

From the definition of the transformation in Equation 3.6, it is obvious that Equation 3.11 can be premultiplied by the interpolation matrix $\mathbf{L}_{11}$ to obtain $\mathbf{P}$ in the form,

$$\mathbf{P} = \mathbf{L}_{11}(\mathbf{L}_{11}^T\mathbf{L}_{11}^{-1})(\mathbf{L}_{11}^T\mathbf{L}_{12})\mathbf{A}_{n12}^T. \tag{3.12}$$

At this point most readers will probably conclude that we have taken the elegant Equation 3.7, expanded it with terms that will most likely introduce numerical instability into the transformation, and arrived at the same point. In fact, it can be shown that Equation 3.12 does simplify to Equation 3.7. Now the question becomes, after a somewhat lengthy analysis what has been accomplished?

The answer to this perplexing question comes directly for Equation 3.11. It suggests that given a knowledge of set of image coordinate vectors for each camera and *one* of the transformation matrices the other can be derived. This is not an exciting discovery when considering the fact that the cameras have converging optical axes. By definition, this constraint implies the same matrix, $\mathbf{P}$, for each of transformations, $\mathbf{A}_{nic}$. From Equations 3.6 and 3.7, the information contained in $\mathbf{P}$ is implicit in $\mathbf{A}_{nic}$. What is exciting, however, is the possibility that one of the image space coordinate vectors can be derived from the knowledge of, or an assumption of, the matrix $\mathbf{P}$ and its corresponding image space projection. If this were possible, vectors of one image space coordinates could be projected into the image space of another camera. This procedure is exactly

what Storjohann does in[Sea90]. Back solving Equation 3.1 presents some interesting difficulties. Solving for the image space coordinate vector given the matrix $\mathbf{P}$ and the transformation $\mathbf{A}_{nic}$ involves an over constrained problem. Using knowledge of the perspective transformation, the Z term can be neglected resulting in two equations in two unknowns. The nonlinear nature of the equations demands an iterative numerical solution. Therefore, in Equation 3.11, with a prior knowledge of a set of transformation matrices and a vector of coordinates from a single image space, the vector of coordinates in the other image space can be derived. This type of convergent camera system is used in location of *isodisparity* surfaces, or constant depth surfaces, as described by Storjohann, which are particularly useful in robot navigation on a planar surface. This derivation shows the applicability of the Two-Planes method to a common stereo vision problem. Solutions to the general stereo problem are offered in Chapter 9. An extensive error analysis of the intrinsic and extrinsic uncertainty associated with Two-Planes follows.

3.2 Calibration Error Analysis

The accuracy of the calibration hinges on the ability of the system to recover the precise location of the individual targets in the calibration plane. One instance of the physical realization of this calibration plane is shown in Figure 3.4. After morphological processing, the solid disks in the figure will be isolated and their geometric centers will represent calibration targets. Another important aspect of the calibration error is the error intrinsic in the least-squares fitting process. We will analyze the intrinsic error of the transformation and observe what happens to it as some of the geometric parameters of the scene are perturbed.

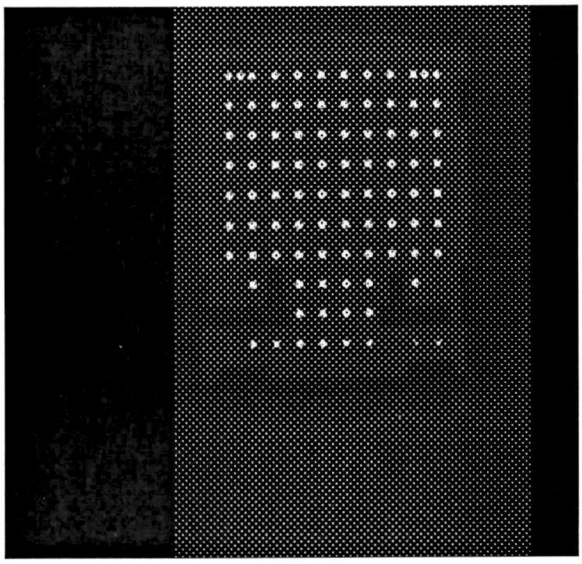


Figure 3.4: Calibration Target

3.3 Transformation Error

If it is assumed that the matrix of image space points that are derived from Figure 3.4 is ideal and the geometry of the calibration plane is known to some degree of precision an error function can be defined as

$$\mathbf{E} = \mathbf{P}_{geom} - \mathbf{P}_{trans} \tag{3.13}$$

where $\mathbf{P}_{geom}$ is the actual geometry of the scene and $\mathbf{P}_{trans}$ is computed geometry from Equation 3.2. This, however, is not a true measure of the error, since some assumptions of the geometry were made. Extending Equation 3.13 to account for this additional error gives

$$\mathbf{E} = [\mathbf{P}_{ideal} - \mathbf{Q}] - \mathbf{P}_{trans}. \tag{3.14}$$

This relates to Equation 3.13 by

$$\mathbf{P}_{geom} = [\mathbf{P}_{ideal} - \mathbf{Q}]. \tag{3.15}$$

$\mathbf{Q}$ is an abstract measure of the uncertainty associated in the geometric description of the scene. It can take on an arbitrary form; however, for the extent of this paper, it is defined as

$$\mathbf{Q} = \epsilon \mathbf{U}. \tag{3.16}$$

$\mathbf{U}$ is the *unity* matrix, where each element is a one, and ϵ is 0.005. A study of one of the particular transform cases is merited. The calibration plane geometry, or $\mathbf{P}$ is

$$\mathbf{P} = \begin{bmatrix} 3 & 5.5 & 18.0 \\ 4 & 3.5 & 18.0 \\ 4 & 4.5 & 18.0 \\ 4 & 5.5 & 18.0 \\ 5 & 3.5 & 18.0 \\ 5 & 4.5 & 18.0 \\ 5 & 5.5 & 18.0 \\ 6 & 3.5 & 18.0 \\ 6 & 4.5 & 18.0 \\ 6 & 5.5 & 18.0 \\ 7 & 3.5 & 18.0 \\ 7 & 4.5 & 18.0 \\ 7 & 5.5 & 18.0 \end{bmatrix}. \tag{3.17}$$

The second order interpolation matrix derived from Figure 3.4 is

$$\mathbf{L}^T = \begin{bmatrix} 3975.63 & 14004. & 749333.1 & 63.0526 & 222.1111 \\ 14363.7 & 9403.7 & 6156.47 & 119.848 & 78.46321 \\ 14318.4 & 17926.8 & 22444.6 & 119.66 & 149.8151 \\ 14454 & 26645.8 & 49121.2 & 120.225 & 221.6331 \\ 31321.4 & 13809.7 & 6088.71 & 176.978 & 78.03021 \\ 31336.7 & 26485.2 & 22384.9 & 177.022 & 149.6161 \\ 31568.1 & 39341.6 & 49029.2 & 177.674 & 221.4251 \\ 54827.4 & 18192.5 & 6036.52 & 234.152 & 77.69511 \\ 54926.1 & 34929.6 & 22213 & 234.363 & 149.041 \\ 55178.2 & 51893 & 48803.3 & 234.9 & 220.9151 \\ 84725.1 & 22647.4 & 6053.72 & 291.076 & 77.80571 \\ 85015.1 & 43369.8 & 22124.8 & 291.573 & 148.7441 \\ 85393.1 & 64474.7 & 48680.6 & 292.221 & 220.6371 \end{bmatrix}. \tag{3.18}$$

Evaluating Equation 3.5 produces

$$\mathbf{A}_{ni}^T = \begin{bmatrix} 1.61E-08 & -7.32E-08 & 7.15E-18 \\ -4.16E-07 & 1.64E-07 & -1.08E-19 \\ -8.17E-07 & -7.31E-07 & -1.84E-18 \\ 0.017 & 8.21-05 & -2.66E-15 \\ 0.00023 & 0.0141 & 3.88E-16 \\ 1.889 & 2.384 & 18 \end{bmatrix}. \tag{3.19}$$

Applying Equation 3.4 generates

$$\mathbf{E} = \begin{bmatrix} 3.00156 & 5.49953 & 18 \\ 4.00054 & 3.50074 & 18 \\ 3.99728 & 4.5003 & 18 \\ 3.99877 & 5.49894 & 18 \\ 5.00044 & 3.49884 & 18 \\ 4.99957 & 4.5024 & 18 \\ 5.0009 & 5.50162 & 18 \\ 6.00124 & 3.49783 & 18 \\ 6.00165 & 4.49875 & 18 \\ 5.99931 & 5.4996 & 18 \end{bmatrix}. \tag{3.20}$$

Equation 3.15 with ϵ set to 0.005 synthesizes a lower error bound of

$$\mathbf{E}_{\text{low}} = \begin{bmatrix} -0.00343766 & -0.00546969 & -0.005 \\ -0.00446217 & -0.00425504 & -0.005 \\ -0.00772256 & -0.00469588 & -0.005 \\ -0.00622575 & -0.00606341 & -0.005 \\ -0.00456184 & -0.00615808 & -0.005 \\ -0.00543156 & -0.00259955 & -0.005 \\ -0.00410243 & -0.0033797 & -0.005 \\ -0.00376345 & -0.00716618 & -0.005 \\ -0.0033487 & -0.00624526 & -0.005 \\ -0.00568884 & -0.00540486 & -0.005 \\ -0.00723707 & -0.00239221 & -0.005 \\ -0.0034608 & -0.00648903 & -0.005 \\ -0.00555716 & -0.00468111 & -0.005 \end{bmatrix}. \tag{3.21}$$

Equation 3.15 with ϵ set to -0.005 synthesizes an upper bound on the expected error.

$$\mathbf{E}_{\text{up}} = \begin{bmatrix} 0.00656234 & 0.00453031 & 0.005 \\ 0.00553783 & 0.00574496 & 0.005 \\ 0.00227744 & 0.00530412 & 0.005 \\ 0.00377425 & 0.00393659 & 0.005 \\ 0.00543816 & 0.00384192 & 0.005 \\ 0.00456844 & 0.00740045 & 0.005 \\ 0.00589757 & 0.0066203 & 0.005 \\ 0.00623655 & 0.00283382 & 0.005 \\ 0.0066513 & 0.00375474 & 0.005 \\ 0.00431116 & 0.00459514 & 0.005 \\ 0.00276293 & 0.00760779 & 0.005 \\ 0.0065392 & 0.00351097 & 0.005 \\ 0.00444284 & 0.00531889 & 0.005 \end{bmatrix}. \tag{3.22}$$

An analysis of the geometric significance of the upper and lower error matrices will be informative. First, however, the numerical stability of the transformation must be proved. Equation 3.5 reveals the weakness of the algorithm. In general, the transformation will be stable if the $(\mathbf{L}^T\mathbf{L})^1$ matrix is nonsingular. Since this matrix is square, the singular value decomposition is employed to study the condition number of the matrix. If the matrix is *ill-conditioned*, it will be found unacceptable.

3.3.1 Numerical Stability of Two-Planes

Singular Value Decomposition theory states that every $p \times q$ matrix $\mathbf{A}$ can be decomposed as

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H, \tag{3.23}$$

where $\mathbf{U}$ is $p \times p$ and unitary, $\mathbf{V}$ is $q \times q$ and unitary, and $\mathbf{\Sigma}$ is $p \times q$ and diagonal in that $< \mathbf{\Sigma}_{ij} >= 0$ unless $i = j$, in which case we write $< \mathbf{\Sigma}_{ii} = \sigma_i >$, real and nonnegative[ND88].The values, σ_i, are called the *singular values* of $\mathbf{A}$, and the *rank* of $\mathbf{A}$ is the number of nonzero singular values of $\mathbf{A}$. It must be shown that the inverse of the matrix $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ is stable (i.e. $\mathbf{A}$ does not lose rank). If the matrix $\mathbf{A}$ is square, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ are likewise square matrices of the same size. Through the decomposition of $\mathbf{A}$, any instability in $\mathbf{A}^{-1}$ may be exposed. The matrices $\mathbf{U}$ and $\mathbf{V}$ are square and unitary; therefore, the inverses are their hermitian transposes. The matrix $\mathbf{\Sigma}$ is diagonal and its inverse may be obtained by taking the reciprocal of the individual σ_i. Press[Pre88] formalizes this as

$$\mathbf{A} = \mathbf{V}[\mathbf{diag}(1/\sigma_i)]\mathbf{U}^H, \tag{3.24}$$

and states that the instability will occur only if one or more of the σ's becomes zero, or if $\mathbf{A}$ losses rank. Because this theory supposes that the numerical tools have infinite resolution, modifications must be made to support the finite precision of a digital computer. Therefore, any inverse will not be stable if the reciprocal of one of the σ_i's is so small that is value is dominated by roundoff error and therefore unknowable[Pre88]. This is expressed formally as the ratio of the largest eigenvalue, or σ_i, to the smallest, and is called the *condition number.* If the condition number is infinite, then $\mathbf{A}$ is singular. If the number is large with respect to the floating point precision of the machine, the $\mathbf{A}$

is *ill-conditioned.* The vector of eigenvalues from $\mathbf{\Sigma}$ is

$$\mathbf{Trace}(\mathbf{\Sigma}) = \begin{bmatrix} 4.51553 \\ 0.0322346 \\ 0.00623286 \\ 5.429548 \times 10^{-5} \\ 4.23481 \times 10^{-6} \\ 1.19951 \times 10^{-5} \end{bmatrix}. \tag{3.25}$$

and, the condition number γ for this system becomes

$$\gamma = \frac{4.51}{1.19 \times 10^{-6}} \approx 3.79 \times 10^{-6} \tag{3.26}$$

The IEEE 754 standard for floating point arithmetic outlines the smallest increment that is representable by a given floating point number as a *long double* without loss of precision is 1×10^{-19}[Pla92]. Upon comparison of this with the number generated in Equation 3.26, the stability of the transformation becomes clear. Thirteen orders of magnitude shield the condition number from the floating point limit of the machine. We could conclude that the matrix $\mathbf{A}$ is *well-conditioned* and the numerical analysis unto this point is valid; however, one question remains. How is the stability of the transformation process affected by uncertainty in the calibration geometry? The same type of singular value analysis is performed where perturbations are imposed on the initial geometry matrix $\mathbf{P}$. If these matrices are well-conditioned then the transformation can be judged as stable. The image space locations that form matrix $\mathbf{L}$ were disturbed by $\pm\delta$ equal

to 1 pixel. The matrix of eigenvalues for $+\delta$ is

$$\mathbf{Trace}(\Sigma_{\text{low}}) = \begin{bmatrix} 5.68023 \times 10^{10} \\ 7.03755 \times 10^{09} \\ 3.38984 \times 10^{08} \\ 25869.6 \\ 951.445 \\ 0.0478875 \end{bmatrix}, \tag{3.27}$$

while the matrix of eigenvalues for $-\delta$ is

$$\mathbf{Trace}(\Sigma_{\text{up}}) = \begin{bmatrix} 5.47354E+10 \\ 6.86323E+09 \\ 3.39446E+08 \\ 25611.4 \\ 973.545 \\ 0.0502348 \end{bmatrix}. \tag{3.28}$$

These matrices yield the following condition numbers:

$$\gamma U = \frac{5.68E+10}{0.0478} \approx 1.18 \times 10^{12} \tag{3.29}$$

and

$$\gamma L = \frac{5.47 \times 10^{10}}{0.0502} \approx 1.08 \times 10^{12} \tag{3.30}$$

for upper and lower bounds, respectively. The condition number with these perturbations imposed still resides within the limits of long doubles; however, double precision and single precision would be inadequate. Appendix A contains five more cases of this analysis.

3.3.2 Geometric Significance of the Error

It is desirable to extract some type of measure from the upper and lower error matrices to describe in a meaningful geometric sense the errors inherent in the least-squares approximation. We turn to models from linear algebra of the various norms of a matrix and try to interpret some meaning from them. The most common norms on a $p \times q$ matrix $\mathbf{A}$ are the *One*, *Two*, and *Infinity* norms which are defined as

$$\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^{p} |a_{ij}|, \tag{3.31}$$

$$\|\mathbf{A}\|_2 = \max \sigma_{ii}, \tag{3.32}$$

and

$$\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^{1} |a_{ij}|. \tag{3.33}$$

respectively. Once these standard tools are applied to the upper and lower error bound matrices, some useful geometric information is derived, as shown in Table 3.1. The first column in the table represents the maximum absolute value of the elements in the matrix. Geometrically, this represents the limiting uncertainty that a single point can have in one dimension. If we assume that this error exists in each element of a three-dimensional coordinate vector obtained by the transformation, the distance formula will produce the radius of a ball centered at that particular coordinate vector. This measure is the maximum error, per recovered point, that the transformation coupled with quantitative assumptions concerning the calibration geometry will acquire. Column three of Table 3.1, the infinity norm, provides a less severe estimation of this uncertainty sphere by maximizing the sum of the absolute values of the rows. The One norm, from column two, describes a cumulative error effect in each one-dimension, such as the sum of errors in the x direction. This is clearly a function of the number of calibration points and not descriptive of error in the recovered geometric structure.

3.4 Conclusion on Calibration

An in-depth mathematical analysis of the calibration mechanism was performed. The transformative process was found to be numerically stable over the operating region via a singular value decomposition and perturbation analysis. Upper and lower error bounds were computed within a band of geometrical uncertainty. Error boundary analysis claims points recovered from the Two-Planes calibration engine are known within 0.023 inches (23 *mils*). While error is certainly contained within this relatively wide band, actual point recovery error taken from the infinity norm will maintain a average value of 17 *mils*. Finally, it is naive to claim this error is not a function of calibration distance. This data was taken from a calibration plane with a standoff of 18 inches. Some degradation as a function of distance is observed from the data contained in Appendix A. A standoff of 48 inches the infinity norm reveals error in the 30-40 *mils* range. The stand-offs define a rough operating region. All points taken from the calibration engine within the region will be assumed to have a maximum error of 30 *mils*. This should more that adequately compensate for any calibration error. It will be shown that an error of this magnitude within the operating region is insignificant when compared to spatial quantization effects. In other words, the maximum error in the system is not a function of uncertainty stemming from the Two-Planes method and its calibration geometry; rather, it is primarily a function of focal length and pixel density. In the application, the Two-Planes method performs adequately.

Table 3.1: Geometric error analysis showing calibration uncertainty (in inches) through three different norm functions.

	Element	One	Infinity
Upper	0.0076	0.650	0.01741
Actual	0.0027	0.016	0.00484
Lower	0.0077	0.650	0.01741

CHAPTER 4

Geometry Error Analysis

Aside from error intrinsic to the least-squares fit of the Two-Planes model, additional error will accumulate due to some fundamental assumptions that were made corresponding to the features the system synthesizes from an input image. As mentioned, the primary features, sets of parallel lines, represent the linear extremal boundary of the pipe in question. The following assumption was made: *a line segment connecting two points, one from each of the boundary lines, will intersect the axis of least-inertia of a cylinder in exactly one point.* Less formally, from any arbitrary viewpoint exactly one-half of the cylinder is observed. Figure 4.1 demonstrates the error in this assumption. Another critical assumption was made. *If a cylinder is observed from two different locations, the same half-cylinder is viewed.* In most cases this is not valid and Figure 4.2 graphically proves this. Readers familiar with perspective geometry will conclude that we have assumed an infinite viewpoint in the model. While it is true that both assumptions are valid from an infinite viewpoint, in reality, cameras must be reasonably close to the scene. However, we will prove that within certain limits these approximations will not affect the over-all accuracy of the system. In fact, the error will be insignificant with respect to the pixel density problem described in Section 2.3.

4.1 Error: The Infinite ViewPoint Model

A mathematical analysis of the first assumption is warranted. The origin of the system in Figure 4.1 is assumed to be the center of the circle which represents top view of a cylinder in two dimensions. The viewpoint resides along the x axis at standoff of

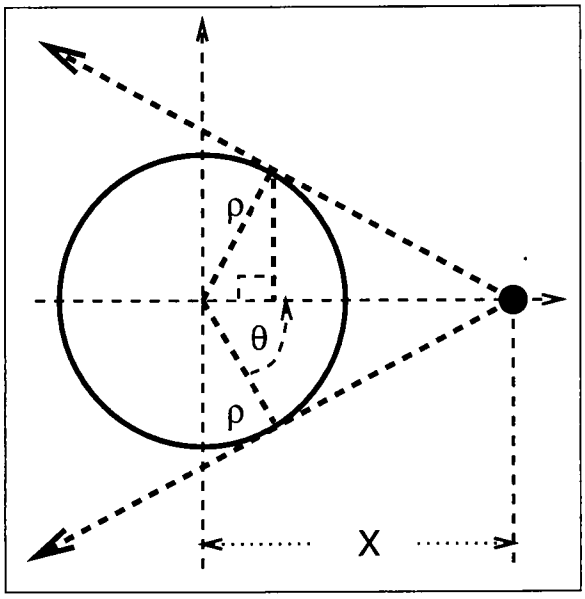


Figure 4.1: Monocular Viewpoint Error

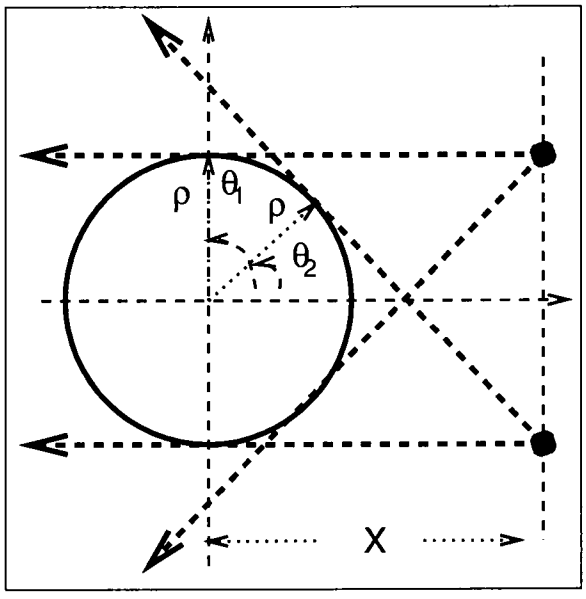


Figure 4.2: Stereo Viewpoint Error

X. Tangent vectors from this point form the portion of the circle that is perceivable. A derivation of the percentage of the surface that is actually imaged as a function of standoff will be obtained. From elementary geometry, any line formed from a point lying on the center of a circle and a tangent point T on that circle is perpendicular to the tangent vector at T. This is described mathematically as

$$\vec{a} \bullet \vec{b} = \vec{0} \tag{4.1}$$

where $\vec{a}$ is a vector from the origin to the tangent point T and $\vec{b}$ is the tangent vector from the observation point. The ($\bullet$) operator represents the *dot-product.* From trigonometry

$$T = (\rho \cos\theta, \rho \sin\theta), \tag{4.2}$$

also by definition, the viewpoint V is defined as

$$V = (X, 0). \tag{4.3}$$

Using this information,

$$\vec{a} = < \rho \cos\theta, \rho \sin\theta > \tag{4.4}$$

and,

$$\vec{b} = < \rho \cos\theta - X, \rho \sin\theta - Y > . \tag{4.5}$$

Substitution into Equation 4.1 produces

$$< \rho \cos\theta, \rho \sin\theta > \bullet < \rho \cos\theta - X, \rho \sin\theta - Y >, \tag{4.6}$$

which reduces to

$$X \cos\theta + Y \sin\theta - \rho = 0. \tag{4.7}$$

Examination of Equation 4.7 algebraically describes the relationship between the viewpoint and the angle θ. Also, from the geometry in Figure 4.1, the Y coordinate of the viewpoint is zero. Solving for θ yields

$$\theta = \cos^{-1}(\frac{\rho}{X}). \tag{4.8}$$

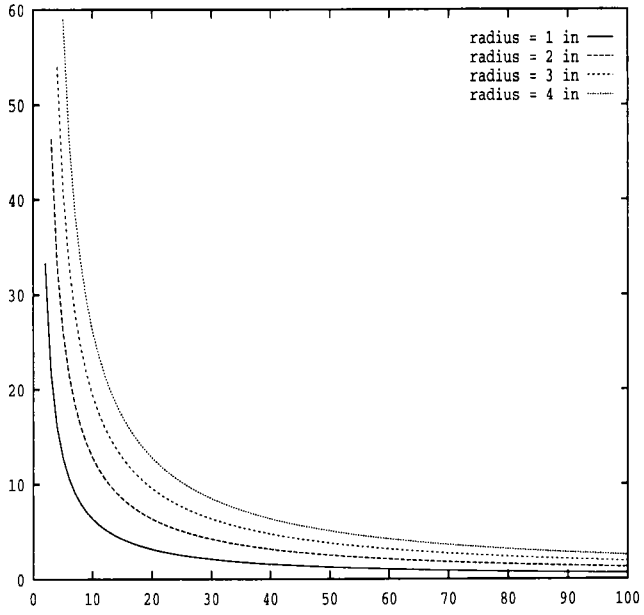


Figure 4.3: Percent Error vs. Standoff

A quantitative description of the error can be derived from a comparison of the arc length encompassed in a particular field of view with respect to the arc length of a half-circle of radius ρ. The equation for arc length viewed for a particular θ is

$$s = 2\rho\theta. \tag{4.9}$$

Figure 4.3 shows the quantitative error as a function of radial standoff for several values of cylinder radii. A careful balance between the allowable range of cylinder diameters and the minimum standoff distance must be preserved. In this system, the standoff defined by the stereo imaging geometry is fixed at 15.0 inches; therefore, from the curves in Figure 4.3 unacceptable inaccuracies will begin to accumulate at cylinder diameters of about 2 inches and greater. Additional analysis will be required before an adequate operating range can be defined. The accuracy of stereo matching with respect to standoff distance must be quantified.

4.2 Error: The Stereo Projection Model

Error introduced from this approximation originates from the mismatch in views from different observation points as depicted in Figure 4.2. The stereo matching assumes that the two views are the same. A quantification of the error can be found by studying the difference in the angle θ subtended by each of the viewpoints assuming lateral displacements along the y-axis from viewpoints set at

$$\left\{ \begin{array}{l} V_1 = (X, +D) \\ V_2 = (X, -D) \end{array} \right\} \tag{4.10}$$

From Equation 4.7 the following system is introduced:

$$\left\{ \begin{array}{l} X\cos\theta_1 + D\sin\theta_1 = \rho \\ X\cos\theta_2 + D\sin\theta_2 = \rho \end{array} \right\}. \tag{4.11}$$

This system must be solved numerically; however, applying the radial formula for θ from Equation 4.8 and translating to a common coordinate frame, the error (in radians) is

$$\psi = \theta_1 - \theta_2 + 2\alpha, \tag{4.12}$$

where

$$\alpha = \tan^{-1}(\frac{D}{X}), \tag{4.13}$$

and the radial distance X from Equation 4.8 becomes

$$\chi = \sqrt{X^2 + A^2}. \tag{4.14}$$

For the geometry displayed in Figure 4.2, θ_1 and θ_2 are equal, further simplifying the expression. Figures 4.4 and 4.5 plot the relative error as a function of X and D for ρ equal to 1 inch and 2 inches, respectively. Based on the empirical data outlined in this and the previous sections, the following *rule-of-thumb* is proposed. If the ratio of pipe

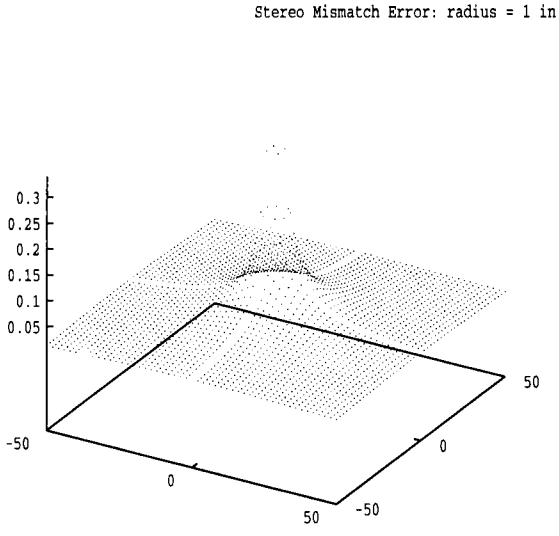


Figure 4.4: Stereo Mismatch vs. Standoff: $\rho = 1$

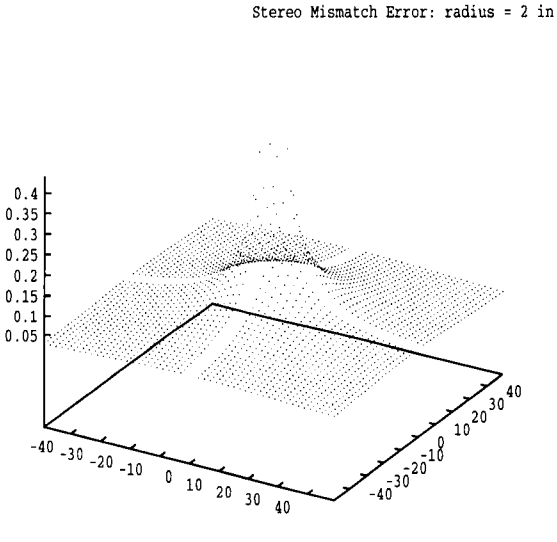


Figure 4.5: Stereo Mismatch vs. Standoff: $\rho = 2$

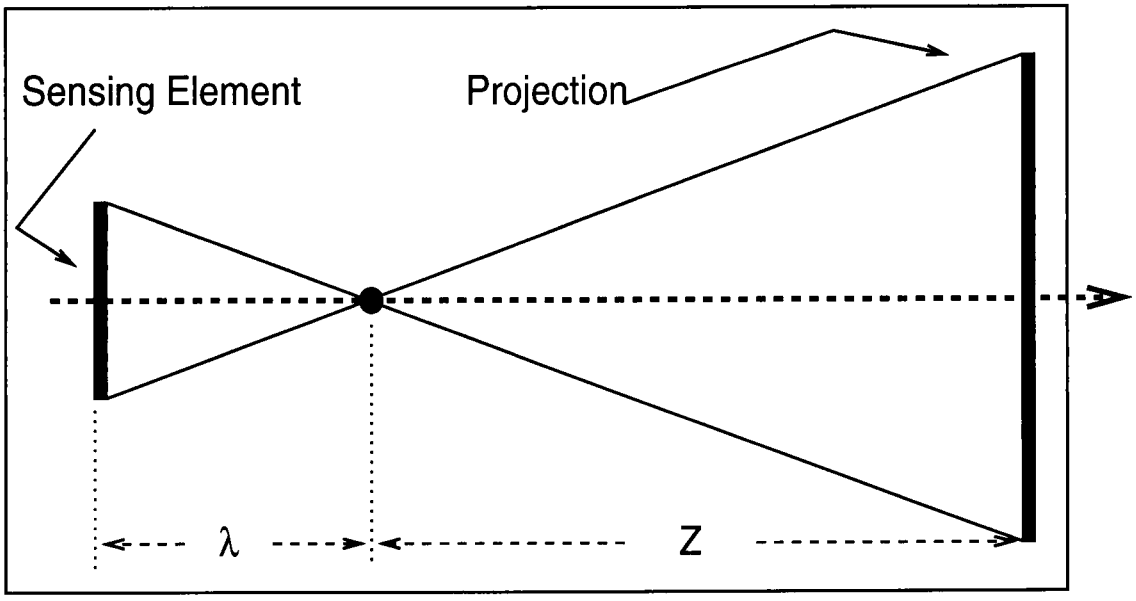


Figure 4.6: Spatial quantization as a function of sensing element projection and camera focal length

diameter to the radial distance to the imaging surface is greater that ten, the error will be less than four percent. Stated numerically,

$$\frac{\rho}{\chi} \geq 10 \Longrightarrow \textbf{error} \leq 4\%. \tag{4.15}$$

4.3 Spatial Quantization Error

Spatial Quantization, or *sampling*, is the formation of a two dimensional spatial grid associated with an image[OW83, GW87, Jai89]. This process is inherent in image formation. Conceptually the imaging surface consists of an array of sensing elements arranged in a rectangular grid. If we assume that these sensing elements are square, Figure 4.6 depicts the sensing element into three space. Using the lens law, the size of this projection as a function of standoff distance can be derived. Obviously, the size

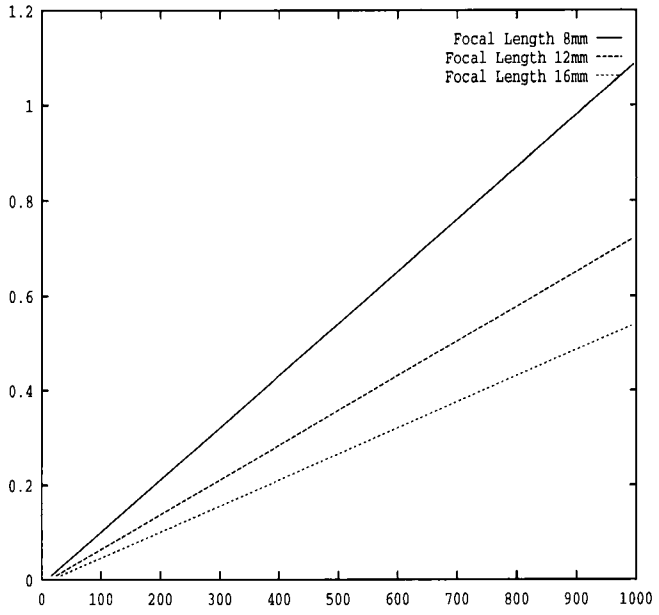


Figure 4.7: Pixel projection vs. object standoff distance for three commonly used lens systems.

increases linearly as the distance to the imaging surface increases, or, more formally,

$$s = d\frac{z - \lambda}{\lambda}. \tag{4.16}$$

The variable s is the length of the projected square, d is the sensing element size, z is standoff distance, and λ is the focal length of camera system. Figure 4.7 shows plots of length of the projected pixel square (in millimeters) as a function of standoff distance for three commonly used focal lengths. The calibration of the camera system was implemented using the 16 mm lenses. From the figure, the projected size is roughly $\frac{1}{2500}$ of the standoff distance for a focal length of 16 mm. At 18 inches, the projected square is approximately 7 mils, and at 50 inches it is about 20 mils. Within this boundary, a maximum uncertainty in point location of 20 mils is assumed. Magnitude of this error is on the order of the error derived from the analysis is Sections 2 and 3. Consequently, we assume maximum uncertainty of 60 mils; 20 mils for calibration error, viewpoint mismatch, and pixel projection. Beyond 50 inches spatial quantization error

will begin to dominate the total error approximation term. A perturbation analysis of how this cumulative error term affects stereo matching is merited; however, this study is more appropriately presented in the stereo vision section section. We sum up the previous analysis with the following assumption: **Total assumed uncertainty in feature location within the defined region of operation will not exceed** 60 **mils.** A discussion of the software algorithms for locating pipe candidates follows.

4.4 Geometry Error Conclusions

A quantitative measure of the error inherent in certain approximations made to simplify the pipe detection process was derived. Error accumulation was found to be inversely proportional to radial standoff from the central axis of the cylinder. In the stereo case, the baseline distance also plays a similar role. Figures 4.3, 4.4 and 4.5 clearly display that these assumptions will be numerically insignificant after some standoff distance; however, the inherent error in stereo vision depth recovery increases in direct proportion with radial standoff. This error is examined in Section 2.4 Constraints derived from this section must be applied to Equation 4.15 before a final operating region will be defined.

CHAPTER 5

Design of the Graphical User Interface for Interactive Cylinder Detection

Although the interaction between a human and a computer is a complex process that is only vaguely understood, a critical component of any man-machine system is the user-interface [CSA89]. Its presence in the system represents a major rift in the two different design philosophies that are being analyzed. Strictly automatic vision algorithms require no interface beyond simple initialization procedures, while the interface constitutes the heart of an interactive vision system. With this in mind, the description of a user interface for interactive cylinder localization is described. The design techniques to a certain degree follow Annet and Duncan's Hierarchical Task Analysis procedure (HTA) as outlined by Shepard [She89a].

5.1 Task Analysis of Interactive Cylinder Localization

Before an analysis of the task requirements in organizing an effective user interface can occur, some basic concepts and definitions from HTA must be outlined. *Goals* are defined as the desired states of the system under observation or control. They may remain constant over the life of an operation or they may change dynamically. The goal of HTA in the development of the user interface is to help the operator fulfill the task requirements in a more efficient and productive manner. An important distinction must be drawn between the goals a system seeks to attain and the methods used to attain them [She89a]. The definition of a *task* permits this distinction. A task is a method

used to achieve the system goals. Any particular goal may require a sequence of tasks to fulfill it. The choice of tasks may have a profound influence on the manner in which a goal is achieved. An *operation* defines the behavior associated with accomplishing a task. The levels of the hierarchy can now be seen. Goals are decomposed into tasks and the operations required to complete them.

The design requirements of a complex user interface may not be constrained to a single goal just as individual goals may not be achieved by single tasks. Goals may also be defined at various levels of precision, and these may in turn be decomposed into a related set of *subgoals.* This relates to the development of the user interface in a highly relevant way. By definition, the major goal of this human-machine system is the recovery of the three-dimensional pose information from cylinders in an unstructured environment. From this statement, only a general guideline can be formalized. Now this general goal must be segmented into a hierarchy of subgoals that impose a more rigid structure on the interface design. The major subgoals of the interactive localization system are (*i*) data acquisition, (*ii*) data enhancement or reduction, (*ii*) object identification, (*iii*) object localization, and (*iv*) communication or storage. Subgoals must be linked in the framework of some overall process or *plan* that directs the execution and synchronization of subgoals at each individual step in the process. A plan, therefore, is an algorithm that controls how subgoals interact to achieve the primary goal. Now that the general terminology of HTA has been related, the framework of the task analysis must be fleshed out. This framework, called the hierarchical task description, is derived from the following rule set:

- Arrange the description in a hierarchy of plans, subgoals, tasks and, operations required to achieve the primary goal.

- Discover the required operation that needs to be implemented before the primary goals can be achieved.

- Decompose this operation in the set of subgoals and plans necessary to motivate it.

- Refine the subgoals, based on the current operation and task, to the required granularity for efficient operation.

If this framework is adequately prepared, a users model can be defined from it. Figure 5.1 graphically depicts this decomposition. The primary goal, cylinder detection and localization, is shown at the root of the tree structure. A level of branches directly below the root describes the subgoals derived from the primary goal. Any particular subgoal is accomplished by a task that is composed of a sequence of operations that form the leaves of the HTA tree.

Generation of a set of plans for achieving the primary goal should consider interaction of the user at multiple levels. In general, the categorization of plans in HTA analysis is not regarded highly [She89a, CSA89]. This results from the highly dynamic environment of a human-machine system. These systems, by definition, use the human operator to generate the planning or control sequence; consequently, deriving a fixed set of plans is more tuned toward automatic vision systems. Interactive systems, however, lend themselves to a certain class of plans where at required stages in the planning sequence the operator is giving the freedom to liberally apply operations at his discretion. A plan, in this particular case, is a dynamic sequence of tasks that realizes each individual subgoal of the HTA tree. The overall operation is the sequence of plans that when executed in proper order complete the major system goals.

Three types of plans are generated for interface strategies. First, a *fixed sequence* procedure is employed. This procedure is outlined as a general guideline for the user to follow in trying to achieve the primary goal. It is intended for operators that are new to the interface and are trying to develop an intuitive feel for the software. As expected, a standardized set of subgoals and its associated tasks and operations are suggested.

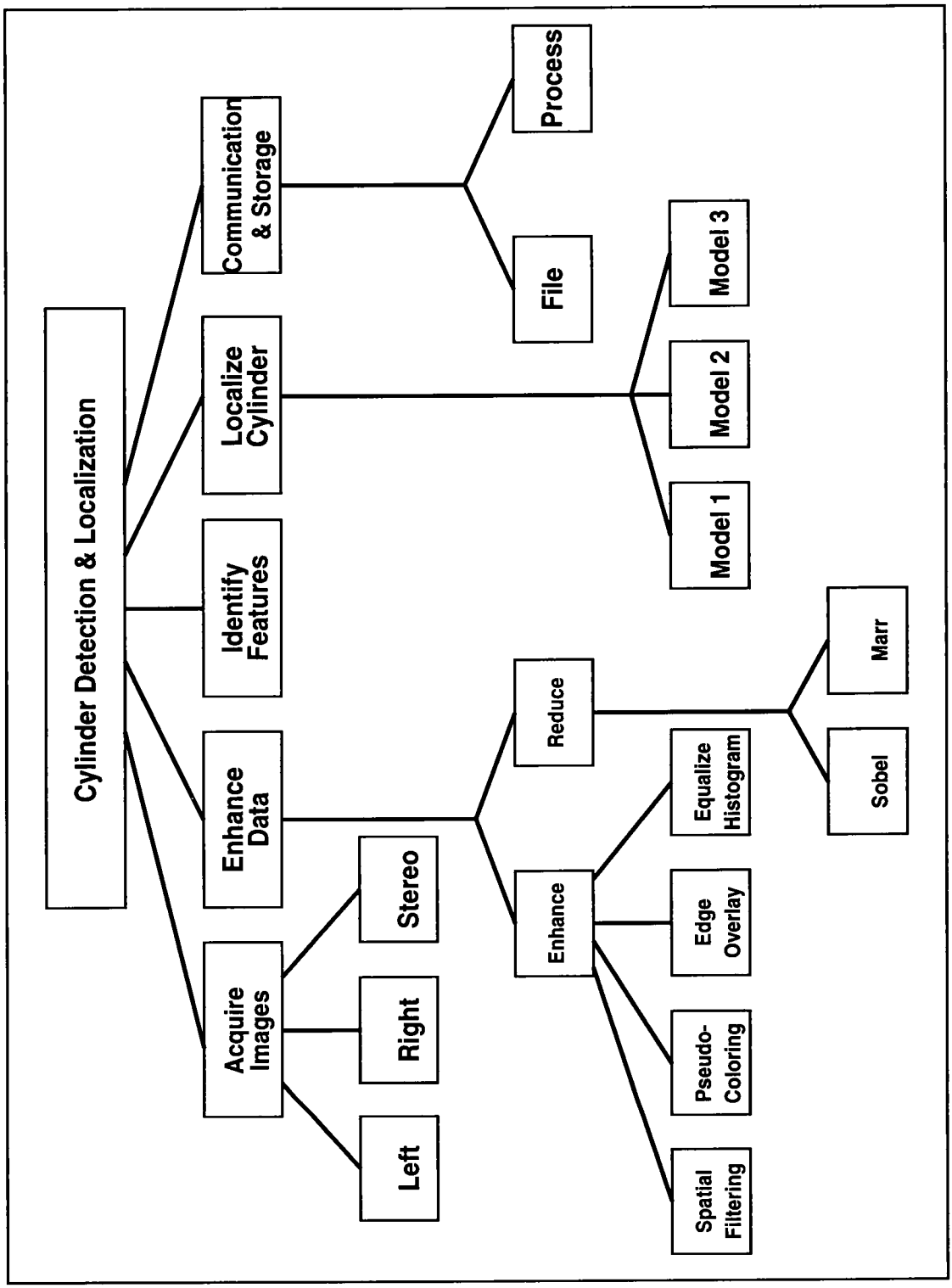


Figure 5.1: Hierarchical task decomposition of user interface requirements for cylinder detection

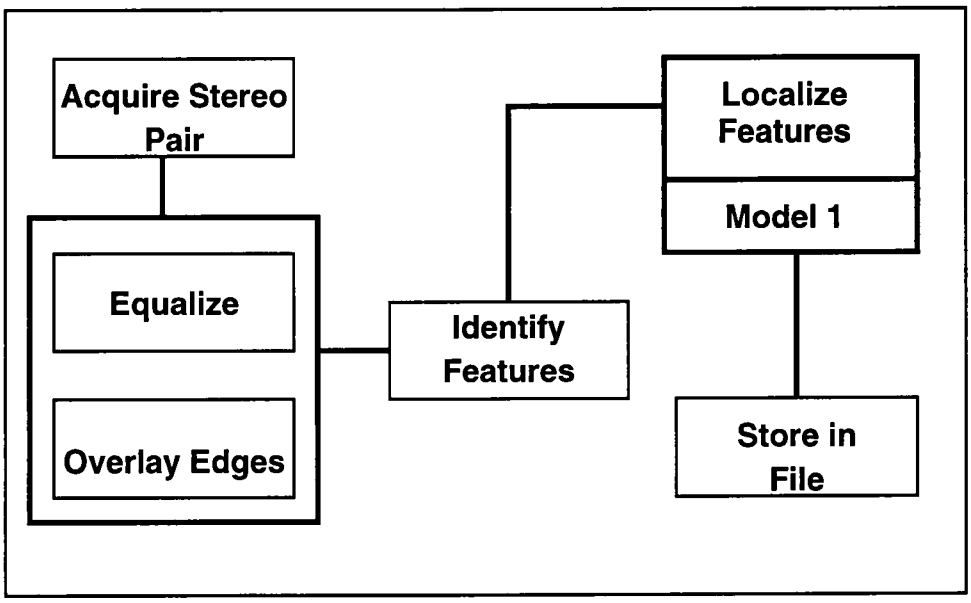


Figure 5.2: A completely rigid fixed sequence control plan.

Figure 5.2 shows a fixed sequence plan in relation to the cylinder detection user interface. Next, a *contingent fixed sequence* is offered for the user who has mastered the initial use of the software and is in an exploration mode. It is characterized by implementation of plans from the fixed sequence control strategy followed by a confirmation that the operation succeeded. In the case of operation failure, a contingency plan helps the operator proceed toward completion of the primary goal. Contingent fixed sequence would be the normal mode of operation for the average operator, and a typical plan is devolved in Figure 5.3 . *Anarchy* is the last stage of plan development. It is represented by a complete breakdown of the hierarchy. In this mode users apply operators without regard to any of the outlined subgoals or their tasks. Primarily, it is intended for the experienced operator who feels constrained by the restrictive nature of the fixed or contingent fixed control plans. An anarchist representation is outlined in Figure 5.4 The structure and control plans for the user interface have been outlined. What remains is a study of human performance capabilities before the software can be realized.

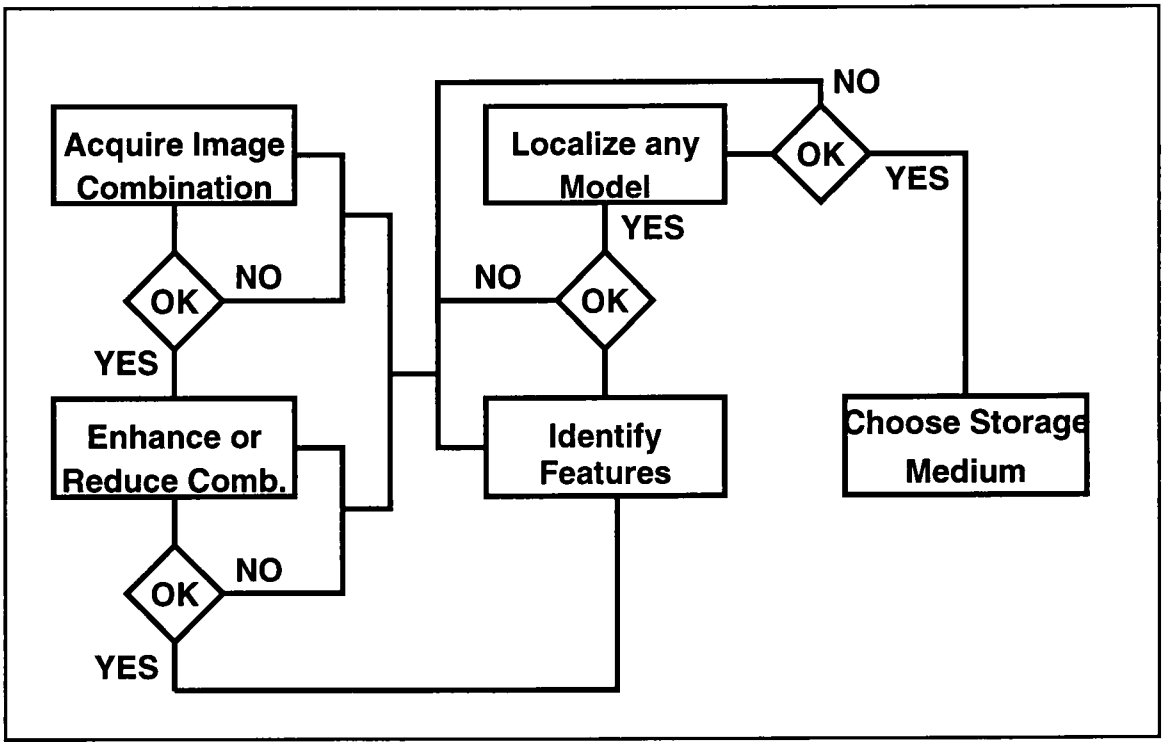


Figure 5.3: A more flexible contingent sequence control plan.

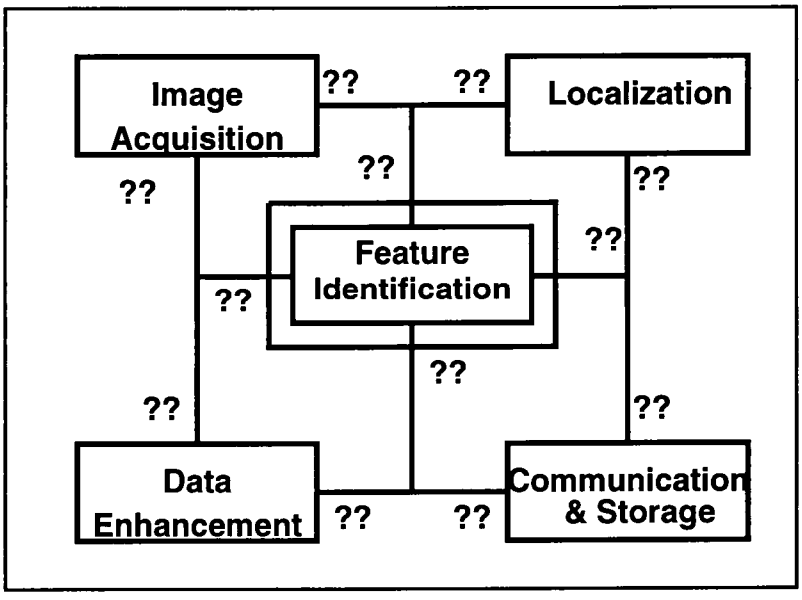


Figure 5.4: An anarchist control plan that allows maximum freedom.

5.2 An Overview of Human Performance

Sheridan [She74] relates that "it is clear, on biological grounds alone, that people have limited resources to devote to information processing, that these resources do no function instantaneously, and, as a result, that people have a limited capacity to handle information". Acceptance of this leads the designer of a user interface to ask the following: (*i*) *How can the maximum amount of sensory data be conveyed to the operator without overloading his information processing capabilities*, and (*ii*) *What algorithms will simplify his interaction with the user interface system?* Only after answering these questions adequately can the completed interface be built upon the HTA framework.

As stated, the task of the user-interface is to provide a means for the operator to interact with vision hardware and software in the task of three-dimensional cylinder localization. Sensing modalities of the vision system were arbitrarily limited to a stereo pair of reflectance maps. From this imagery, the operator is expected to provide cues to the computer that will uniquely identify the same cylinder in both images. The cues, pairs of points that lie on the linear extremal features of the cylinder, will then be used by the computer to synthesize range data. Since the visual data must be directly observed and possibly enhanced by the operator using the methods of Chapter 6, a stereo display must be used. This display unit will show in the proper order, i.e. left-to-right, the pair of images acquired by the camera system. After arbitrary observation time, the operator may choose, depending on the HTA plan, to interactively tune the data. Usually, this process will involve a cyclic processing until adequate (in the subjective sense) enhancement has been obtained. From here, further interaction will be required; consequently, the operator must visually (for mutual benefit to both man and machine) mark the appropriate features. Finally, the software will be signalled that the interaction with the operator is complete and to proceed with the current plan. Two important

interactive areas have been identified:

- Observation and perception of the reflectance maps.
- Cataloging, in the proper order, a list of features required for stereo processing.

Analyzing these two phenomena relates directly to answering the questions we have outlined above.

5.2.1 Observation and Perception of Reflectance Maps

The basic perceptual task requires the operator to observe the stereo image pair and decide the following:

- Does the left image contain the projection of a cylinder?
- Does the right image contain the corresponding projection?

These statements are both rich in connotation and structure. Each requires the presence of significant refractory power in the observer's perceptual process and adequate sensory data representation on the part of the interface. The existence of interactive algorithms for effectively extracting the desired data is also implied. Based on this qualitative outline, what can be expected of the human operator in a given situation? Sheridan points out the following basic predictions of human performance in information processing tasks:

- A human operator has an absolute limit on the amount of information that he can process in a given time frame.
- Performance is not a function of information coding.
- Switching between concurrent processes is possible.

- In operations proceeding at capacity, increased speed is accompanied by a decrease in accuracy.

The constraints imposed by these predictions are particularly relevant to visual processing. The first constraint has an immediate effect on user interface design, i. e., no time limits. This simply says that the user must be provided with all the time he deems necessary to solve the particular task. Sheridan's second observation relates that the performance of the visual task will be relatively uniform across a large range of data presentation styles. While it is acknowledged that human visual processing is a complex integration of multiple sensory responses, the second principle offers the assurance that stereo imagery should be adequate input for the perception process [Ron91]. In this case study, the third principle verifies that the operator will be able to correlate between the set of images without losing any of the perceptual power with respect to an individual image. From this, a serial matching strategy is suggested. An operator can identify a feature in one image, and then correlate it with its associated feature in the complementary image without perceptual loss. The last principle may seem to be a restatement of the first; however, it is concerned with capacity of the individual user to perform a specific task at a desired rate. This capacity is a complex function of a number of variables, and is not very well understood. It is well-known that the efficiency of the operator will improve with practice and increased performance times with invariant accuracy will follow; however, this development will proceed at a much slower rate. Invariantly, the last principle seems to impose a minimum perception time on the operator depending on his aptitude for the task and his experience with the system. A general human performance evaluation has helped to outline some additional attributes that the user interface should contain. What is still missing, however, is an analysis of the primary interfacing task, that of cataloging the linear extremal features in an effective and efficient manner.

5.2.2 Cataloging the Relevant Features

A classical problem in human-machine systems research is that of *human vigilance.* Holland describes it has the problem of "sustained efficiency in monotonous perceptual tasks" where "human beings are required to monitor some display in search of critical signals" [Hol71]. When Holland refers to vigilance, the context is based on changing displays that require constant attention, such as a radar screen; however, the fundamental concept of improving an operators ability to detect some signal whether it is a radar "blip" or an image feature is equally relevant. Howell, et. al., extend the concept of vigilance to "the study of how humans function as a monitor", and conclude that man is not a suitable to function in this capacity [HJG71]. This statement is not necessarily an indictment of human-machine systems; rather, it is motivation for the study of the systems and how to improve human performance in them.

Typical studies of human vigilance concern themselves with monitoring variables that affect perception in time-varying displays. Lack of temporal variation in the display creates some difficulties for traditional studies in vigilance. These studies usually are concerned with time-based variables such as signal frequency and latency. Measures that could be applied to static displays, such as signal density and clutter, require measurement of easily observed phenomena (in the time-varying case) like the number of actual "blips" fed into a radar display or the signal-to-noise ratio. These metrics lose some of their intuitive feel when applied to perception at higher levels of abstraction. For instance, will it become more difficult for humans to discriminate among individual cylinders in an image as the their number increases? In the limit, the answer becomes a definite yes. However, in practical situations where the number of cylinders present is reasonable, to affirm this statement may be too bold. If the concepts of vigilance are to be applied to static displays that require higher levels of reasoning and perception, the measures of it must change also.

Applying some common sense to the problem, it would seem that several factors would have a significant effect on the performance of a human in the feature selection task. Reflectance properties such as *contrast*, and geometric boundaries such as the ratio of cylinder diameter (in pixels) to the image diagonal can have a profound effect on the accuracy of the interactive feature selection mechanism. Two case studies are presented. First, the sensitivity of the feature selection to the contrast of the image is tested. A black cylinder on a white background composes the experimental scene. The scene (constant lighting) is exposed to a decreasing amount of light by reducing the aperture size. The final experimental measure used a simple image to recover the accuracy of the matching as the time constraint applied to the operator becomes increasingly more severe.

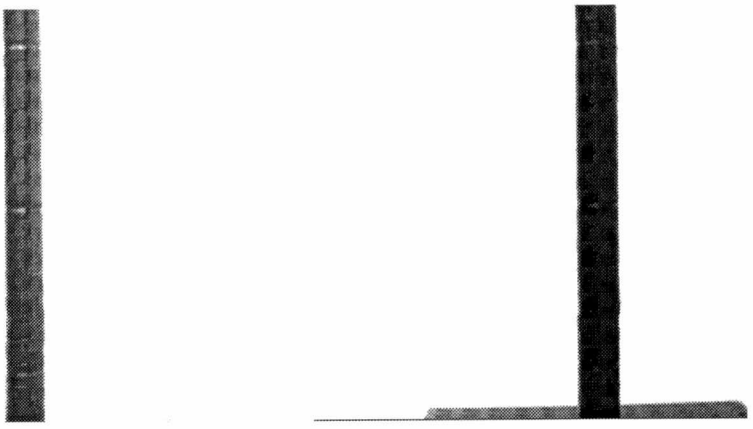
Sensitivity of the Feature Selection to Image Contrast Levels

It is known that the contrast sensitivity of the human eye has been measured by the *Weber* ratio [GW87]. This ratio, calculated by incrementally adjusting the grey level ΔB of a target on a uniform background of grey level B, remains constant at about 0.02 over a large range of brightness levels. It would seem that this is enough support to indicate that a human operator will display tolerance to a wide range of contrast in the input scene. This ratio, however, is based only on visual recognition, and it may not have any implication in an interactive feature selection mechanism. An experiment must be designed that measures the sensitivity of the feature selection, not just the recognition, to the contrast present in the image. A cylinder was presented to the interactive localization system under constant lighting at a standoff of 22.0 inches. Changes in contrast were implemented by varying the aperture setting incrementally from completely open to completely closed. Figures 5.5 and 5.6 display a sampling of this image sequence. Interactive localization was performed and the results recorded

Table 5.1: Results of interactive cylinder localization under contrast variation. Cylinders are translated with respect to the coordinate frame with constant standoff of 22.0 inches.

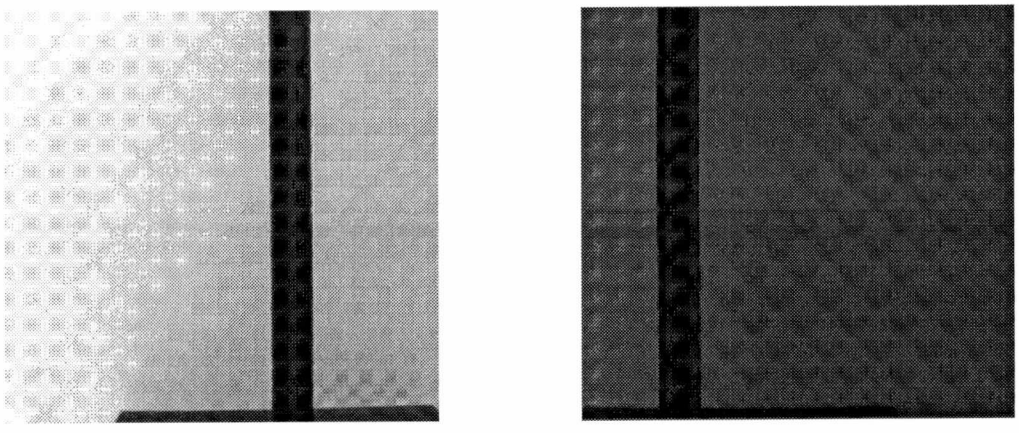
	Aperture Setting						
	1.4	2.0	4	6	8	12	16
Measure 1	22.12	22.1	22.2	22.25	22.05	22.20	FAIL
Measure 2	22.05	22.1	21.95	22.0	21.85	21.90	FAIL
Measure 3	21.85	22.1	22.1	22.25	22.15	22.15	FAIL
Measure 4	21.97	22.0	21.75	22.0	22.05	21.80	FAIL
Measure 5	22.05	22.2	22.25	22.05	22.3	22.28	FAIL
Measure 6	22.24	22.16	21.85	22.05	22.0	22.20	FAIL
Measure 7	21.82	22.2	22.25	22.15	22.2	22.05	FAIL
Measure 8	22.22	22.1	22.0	22.0	21.9	22.09	FAIL
Measure 9	22.18	22.1	22.2	22.31	22.3	21.90	FAIL
Measure 10	22.12	21.85	21.95	22.0	22.05	21.95	FAIL

for each image set. Table 5.2 records the performance of the selection mechanism as a function of aperture setting. Statistical analysis of the positional data in Table 5.1 yields the mean and variance data found in Table 5.2. In all cases but one, the mean value seems to cluster around 22.0in. This seems to indicate that the feature selection mechanism responds over a wide range of lighting conditions in the input scene. Also, it confirms Sheridan's prediction about data presentation and extends the concepts behind Weber's ratio to "pick-and-place" tasks in a graphical user interface. Examination of the variance data from the table reveals a tight clustering of the positional values around the mean. The presence of the small variance tends to solidify the conclusion that feature selection will be resilient in poor lighting conditions. A single failing case is



(a) Aperture Ring = 2 (b) Aperture Ring = 2

Figure 5.5: Test scene shown in "good" contrast conditions.



(a) Aperture Ring = 6 (b) Aperture Ring = 8

Figure 5.6: Test scene shown in "low" contrast conditions.

Table 5.2: Statistical analysis of contrast data

	Aperture Setting						
	1.4	2.0	4	6	8	12	16
Mean	22.06	22.09	22.05	22.10	22.08	22.05	NA
Variance	0.189	0.0968	0.83	0.136	0.216	0.182	NA

reported (Ap = 16) in the experimentation. At this level, no features could be identified and the cylinder was indistinguishable from the background.

Sensitivity of the Feature Selection to Matching Time

It is observed that in the limit improved speed will come at the expense of accuracy [She74]. This seems to make sense; however, it would be useful to have some quantitative outline for the accuracy of the feature selection algorithm as a function of the matching time. The experiment to measure this dependency involved the quantization of time into four categories (*i*) As Fast As Possible (AFAP), (*ii*) Short, (*iii*) Intermediate, and (*iv*) Infinite. AFAP time required matching to be completed in less than two seconds. Intermediate time must be completed in under ten seconds but greater than five. Short time fills the gap between AFAP and Intermediate, while Infinite time is anything greater that ten seconds. Feature selection was performed, and the values were indexed into the time table shown in Table 5.3. Observing the values from the table clearly confirms Sheridan's prediction that speed degrades accuracy. Again, a statistical analysis of the position data reveals the truth of the hypothesis. Mean positional values from Table 5.4 do not indicate a degradation of accuracy; however, the variance becomes significantly larger as the matching time decreases.

Table 5.3: Results of interactive cylinder localization under temporal variation at constant standoff.

	Matching Time			
	AFAP	Short	Intermediate	Infinite
Measure 1	14.89	14.53	14.6	14.62
Measure 2	14.15	14.53	14.6	14.56
Measure 3	14.32	14.71	14.6	14.56
Measure 4	14.95	14.61	14.56	14.61
Measure 5	14.02	14.56	14.56	14.6
Measure 6	14.53	14.65	14.5	14.57
Measure 7	14.62	14.68	14.56	14.63
Measure 8	14.22	14.69	14.56	14.6
Measure 9	15.10	14.71	14.61	14.62
Measure 10	14.34	14.58	14.49	14.58

Table 5.4: Statistical analysis of contrast data

	Matching Time			
	AFAP	Short	Intermediate	Infinite
Mean	14.51	14.62	14.56	14.59
Variance	1.20	0.4825	0.265	0.0063

5.2.3 Conclusions on Feature Selection

Sheridan says that most of the attributes of vision display a nonlinear relationship between the physical quantity and the psychological perception [She74]. Studies of feature selection as a function of image contrast level seem to support this. In the contrast study, image energy is significantly altered; however, the feature selection seems to be unaffected by these changes. Matching time variations, on the other hand, have a dominant affect on feature selection accuracy. As the matching time interval is decreased, the operator can expect the error induced to increase in an inverse proportion. These empirical studies coupled with Sheridan's predictions on sensory perception in man-machine systems indicate that the an implementation of the user interface at this point in the design phase would be successful.

5.3 Constructing the User Interface

Figure 5.7 shows an implementation of a graphical user interface for interactive cylinder detection and localization. The user interface was built on the principles of hierarchical task analysis. A menu-bar provides the various operations necessary for implementation of the task plans. At compile time, a choice of the three planning styles can be selected to govern the human operator's level of freedom in applying the task operations. Currently, the interface is running in anarchy mode.

Based on Sheridan's predictions of performance in concurrent processing, the addition of the stereo display is justified. Using the mouse as the "point-and-click" tool, the operator must select points that represent the linear extremal features of cylinders in one image (a set of rules for this "point-and-click" operation will be outlined later). Concurrency arises from the requirement of the operator to locate a feature in one image and find its match in the corresponding stereo image. Simple matching studies confirm

Figure 5.7: Graphical user interface for interactive cylinder detection and localization.

the success of the feature selection mechanism and indicate that concurrency is not a problem. Another one of Sheridan's predictions that performance is not a function of information coding offers great flexibility in the design of the data presentation format. Data format is not altered by the computer; however, the operator has a wide variety of feature enhancement tools at his disposal.

Studies in human vigilance concerned with image contrast levels and temporal constraints on the feature selection mechanism yield interesting results. It was shown that the perceptive process is capable of operating over a wide dynamic range of lighting levels. In reality, the operator may have no control over the lighting conditions in the environment and the lack of contrast control must be justified. A failing case was reported; however, it represents a lack of incident light on the imaging sensor and is not significant. Constraining the matching time on the operator will reduce the accuracy of the feature selection mechanism; consequently, no time limits where designed into the system. An overall performance evaluation of the stereo matching algorithm is outlined in a later chapter.

CHAPTER 6

An Interactive Feature Enhancement Toolkit

Image enhancement is defined as the processing of images such that the result is better suited to the application than the original images [GW87]. Enhancement techniques by their very nature are highly tuned to specific applications. As a consequence, the enhancement algorithms should be chosen such that they improve the operators ability to perform the required task. Since the primary task is the selection of feature points that geometrically describe the linear extremal features of a cylinder in the imaging space, it is logical that the selected enhancement algorithms primarily concern themselves with the extraction of these features and the suppression of other information.

Some global enhancement techniques should also be available. For example, if the scene is poorly illuminated severe problems might be posed for the operator unless some adequate contrast enhancement algorithm is provided. Another concern to the operator would be the segmentation and thresholding of processed images. Segmentation is a complex process that is usually not very robust. By giving the operator complete control over the segmentation, images can be adequately enhanced for the feature extraction. Enhancement in a human-machine system is a highly subjective phenomenon. We do not pretend to formalize any measure of its success. Rather, a simple presentation and demonstration of the available tools in the user interface package is outlined while the user is left to decide whether the processing was adequate or not.

Figure 6.1 represents a typical input to the cylinder localization system. Histogram equalization of the input yields Figure 6.2. All enhancement operations are pipelined. This means that the resultant image after an operation becomes the input image for

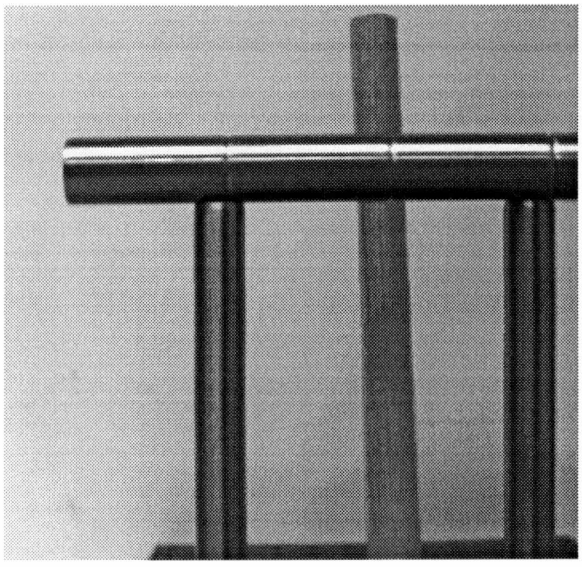


Figure 6.1: A typical input image to before image enhancement.

the next operation. This pipelining can be suspended; however, normally it is desirable. Figure 6.3 displays the edge enhanced version of the equalized image from Figure 6.2. Here, the effects of pipelining can be noticed from the presence of background boundary lines in the image. These are artifacts of applying the edge detection windows to the processed image. A host of other enhancement operators are available for experimentation. These, however, outline the basic functionality of the user interface. In just a few seconds, the operator is able to refine (subjectively) the quality of an input scene to an acceptable grade. After this step, only feature selection and cylinder localization remain.

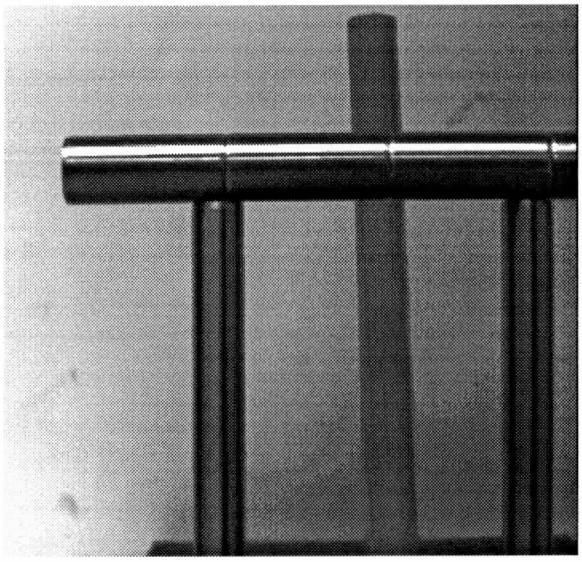


Figure 6.2: Histogram equalized version of the input image .

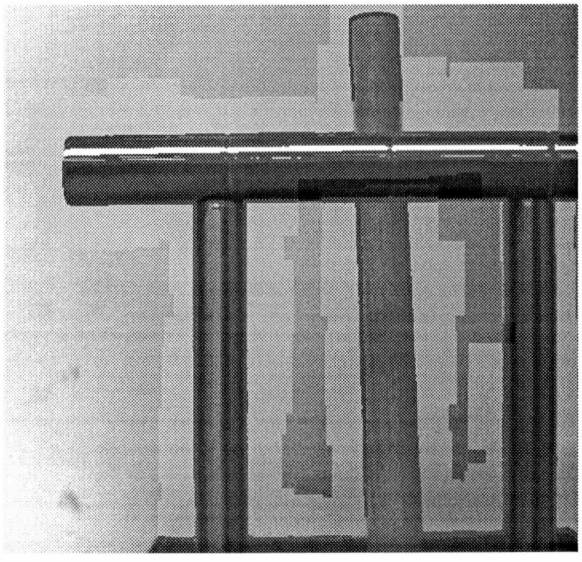


Figure 6.3: Edge overlay of input image after equalization.

CHAPTER 7

Algorithms for Interactive Cylinder Localization

A new objective assuming that the operator is presented with a pair of stereo images becomes the development of suitable operator-assisted algorithms for locating a target cylinder displayed in each of the images. A section of a cylinder in three-dimensional space can be partially described by two points and a scalar representing the diameter. By definition, these two points must lie somewhere on the axis of least inertia [Swo84]. Imposing the structure of the matching derived in the Chapter 5 implies some basic rules that, if followed, help to minimize the effort in locating the cylinder. The following are suggested:

- Any pair of corresponding points entered for pose determination must be approximately perpendicular to the axis of least inertia.
- A minimum of two sets of such points is required.

Figure 5.7 shows a typical situation with two pairs of points that satisfy the first constraint. Assuming that an operator has entered N pairs of points, the diameter can be computed from

$$\mathrm{D} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Dis} < P_{i1}, P_{i2} >, \tag{7.1}$$

where **Dis** is the Euclidean distance operation, P_{i1} is a point returned from the stereo processing satisfying the first condition, and P_{i2} is the corresponding point from image set. An approximation to a point on the axis of least inertia (valid within the error range derived in Chapter 3) can be obtained for each of the N pairs of points from the

midpoint operation.

$$\mathrm{A}_i = \mathbf{mid} < P_{i1}, P_{i2} > . \tag{7.2}$$

Here, A_i is the ith point of the axis of least inertia. A standard least squares fit can be used on this set of N approximate points to refine the estimate of cylinder location, or a piecewise linear approximation could be used. The only remaining task is to define some sort of length for the partially defined cylinder.

Ideally, some benefit could be derived from the tremendous flexibility the human operator introduces by allowing several alternatives in defining the length of the cylinder. Three models are introduced:

- Assume the cylinder length, L, is infinitely long.
- Let the operator interactively set the length.
- Use a standard mathematical approximation.

The first model is useful in an interactive mapping and world model construction software that is not described here. However, the basic method involves an iterative refinement of cylinder length based on the precise movement of the vision system. Alternatively from anchored visual queues or intuition, an operator could interactively tweak the length of the cylinder to any desired value. This model is of use when the operator is forced to select features on the surface of the cylinder as matching points that do not adequately represent the full extent of the cylinder in the image. Exploiting the flexibility of the human-machine system, the operator can compensate for these errors. Finally a simple mathematical description of the length can be derived. An adequate descriptor based on the maximum distance among the set of axis points $\{A\}$ is

$$\mathrm{L} = \mathbf{max}\,\{\mathbf{Dis} < A_i, A_j >\} \ni A_i, A_j \in \{A\}; i \neq j. \tag{7.3}$$

Some basic problems exist with this algorithm. It imposes too much restriction on the operators ability to locate feature matches. The operator is forced to find feature matches then approximate the perpendicular to the two-dimensional axis of least inertia. This practice could induce unnecessary error into the computation. Furthermore, the geometry constraints outlined in Chapter 4 incorrectly state that the perspective projection of a randomly oriented cylinder specifies exactly one half section of the cylinder. Actually, this is only true for a viewpoint that in an infinite distance away; consequently, only a chord of the circular section of the cylinder is being used as a true diameter. In practice, these effects are limited due to the ratio of the standoff of the cameras in the scene to the average cylinder diameter. More error is introduced by assuming that both perspective views specify the exact same portions of the scene. This effect can also be controlled by the same ratio; however, it is important to remember that these occlusions are a by-product of any stereo system and cannot be completely removed.

7.1 Variations on the Feature Selection Mechanism in Stereo Matching

So far, the matching technique described in the feature selection algorithm is a simple point-to-point strategy. The operator must recognize a feature, not necessarily a point, and then approximate its image location at the pixel level. The solution to the stereo problem also demands that the same process be repeated in the corresponding image. While this sort of processing lends itself to the construction of a simple mathematical framework to the solution for the cylinder localization problem, it also demands a great deal of attention from the user. As described, a marked decrease in performance will be experienced if the user is forced to overly concentrate on the task. It would be useful if other types of algorithms were introduced to alleviate some of this burden from the operator. A point-to-region matching strategy could considerably reduce the chance of operator fatigue.

A *point-to-region* matching strategy involves the same basic principle as the previous matching strategy. Again the operator is required to select a point from one of the stereo image pair; however, in the corresponding image a region is demarked. This region becomes the search area for the best possible match of the selected feature point. Searching is accomplished via a normalized cross-correlation with window size, W, centered on the selected feature point, P. At each image point in the search area, R, a corresponding window of size W is likewise selected and the correlation is performed. A match is defined as the point in the search space that gives the numerically largest response to the correlation filter.

7.2 Experimental Results

Until now, the analysis of the human-machine system has been quite subjective. Using heuristics and common sense, some constraints and approximations have been applied to simplify the interface and algorithm design. Final proof of the analysis can only be verified after extensive experimental evaluation. What is the best method to experimentally verify the system? Since its objective is cylinder localization, the experimental verification must measure how well the system can localize cylinders whose geometric placement with respect to the world reference frame is known. An final evaluation metric must account for variants in position. The first measure uses cylinders that are translated with respect to the coordinate frame as in Figure 5.7. The data was taken in 5 separate experiments with ten measurements taken each time. In this case, the geometric "knowns" are the cylinder distance, Z, and the cylinder diameter, D,. We combine them to form the following metric, called the translation error metric:

$$\epsilon_{\text{trans}} = (Z_{\text{ref}} - Z_{ij}) + (D_{\text{ref}} - D_{ij}), \tag{7.4}$$

where Z_{ref} is the standoff distance, D_{ref} is the actual cylinder diameter, and Z_{ij} and D_{ij} are the recovered distance and diameter of the ith measurement in the jth experiment respectively. Since the diameter measure is a differential measurement, (if the match is good) much of the error is cleaved; consequently, the error terms in Tables 7.1 and 7.2 are dominated by the depth error.

7.3 Conclusion of Human-Machine Performance in Estimating Cylinder Pose

A human-machine vision system for the three-dimensional localization of right circular cylinders was designed, implemented, and experimentally verified. System design included a hierarchical task analysis of the requirements of the user interface that produced three general purpose operating plans. These plans, integrated into the software, allow multiple levels of interaction with the sensory data. The first level, called fixed sequence planning, is intended for the novice user. Fixed sequence plans minimize the interaction with the software and greatly reduce the its utility. They are used as a tool for developing familiarity with unfamiliar software systems. A second level, contingent sequenced planning, removes most of the restrictions form the user. Its purpose is to provide a hierarchical level of abstraction to the localization process that greatly reduces the computation burden on the human operator. Contingency is built into this plan to compensate for cases when the freedom given to the operator results in erroneous results. In other words, the operator can undo what has been done. The level of planning, anarchy, is characterized by a complete breakdown of the hierarchy. Its purpose is to provide unconstrained interaction with the data for the experienced user who feels too constrained by the fixed or contingent sequenced plans.

Next a simple study of the human vigilance problems with respect to static displays used in point-and-click operations provided guidance for implementation of the user in-

Table 7.1: Translation error measurement in interactive cylinder localization with "point-to-point" matching: standoff = 16.5in.

	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6
Measure 1	0.10	0.15	0.11	0.08	0.03	0.0
Measure 2	0.10	0.06	0.09	-0.12	0.30	-0.03
Measure 3	0.12	-0.15	0.03	0. 07	-0.09	0.08
Measure 5	-0.07	0.24	-0.01	0.13	0.13	-0.01
Measure 6	0.0	-0.01	0.12	-0.10	0.13	0.0
Measure 7	-0.13	-0.13	0.0	0.01	0.05	-0.16
Measure 8	-0.14	0.16	0.0	0.24	0.17	0.02
Measure 9	0.0	0.03	0.14	-0.07	-0.02	0.07
Measure 10	0.04	-0.05	0.02	-0.04	-0.08	0.02
Measure 11	0.10	0.04	0.13	0.01	0.01	0.11
Measure 12	0.14	-0.09	0.01	-0.12	0.11	-0.09
Measure 13	-0.05	0.13	0.04	-0.03	-0.03	0.09
Measure 14	0.15	0.13	0.14	-0.09	0.05	0.13
Measure 15	0.0	0.02	-0.15	0.14	0.12	-0.04
Measure 16	0.09	0.19	0.10	0.13	-0.03	-0.12
Measure 17	0.04	0.08	0.05	0.06	0.07	0.05
Measure 18	0.23	-0.01	0.03	-0.01	0.04	-0.01
Measure 19	-0.0	-0.32	0.17	0.0	-0.06	0.09
Measure 20	0.12	0.16	-0.11	0.10	0.00	0.14

Table 7.2: Translation error measurement in interactive cylinder localization with "point-to-region" matching:standoff = 16.5in.

	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6
Measure 1	0.10	0.09	0.03	0.09	0.13	0.13
Measure 2	0.09	FAIL	0.02	0.11	0.03	0.13
Measure 3	0.10	0.05	0.18	0. 13	0.11	0.13
Measure 5	0.08	0.14	-0.05	0.20	0.07	0.13
Measure 6	FAIL	0.07	0.09	0.10	0.09	0.13
Measure 7	0.5	0.05	0.03	0.01	0.03	0.13
Measure 8	-0.01	0.11	0.10	0.24	0.10	0.13
Measure 9	0.05	0.09	FAIL	-0.03	FAIL	0.13
Measure 10	FAIL	0.07	0.12	0.11	0.13	0.13
Measure 11	0.10	0.09	0.03	0.09	0.13	0.13
Measure 12	0.10	0.09	0.03	0.09	0.13	0.13
Measure 13	0.10	0.09	0.03	0.09	0.13	0.13
Measure 14	0.10	0.09	0.03	0.09	0.13	0.13
Measure 15	0.10	0.09	0.03	0.09	0.13	0.13
Measure 16	0.10	0.09	0.03	0.09	0.13	0.13
Measure 17	0.10	0.09	0.03	0.09	0.13	0.13
Measure 18	0.10	0.09	0.03	0.09	0.13	0.13
Measure 19	0.10	0.09	0.03	0.09	0.13	0.13
Measure 20	0.10	0.09	0.03	0.09	0.13	0.13

terface. This study produced several interesting results. It was deduced that interaction with the software should not be constrained with respect to time. Also, some justification was found for using serial matching strategies in the concurrent process of stereo feature extraction. From here, some experimental measures of accuracy in a point-and-click operation as a function of the time allowed for each operation was presented. This lead to the justification of the point-to-point matching strategy in the interactive stereo. Point-to-point matching was extended to operate on regions due to possible fatigue problems caused by the perceptual complexity imposed on the user during point-to-point mode. Finally, comparisons of the accuracy of translational localization error between point-to-point and point-to-region matching modes were presented.

The overall accuracy of the human-machine vision was found to adequate for basic robotic tasks within a region of operation. Inside this region, starting at about 12in.(a standoff caused by the baseline) and extending to 36in., the accuracy of the system remained a constant 3%. Operating ranges can be easily extended using a lens systems with a smaller field of view.

CHAPTER 8

Automatic Segmentation

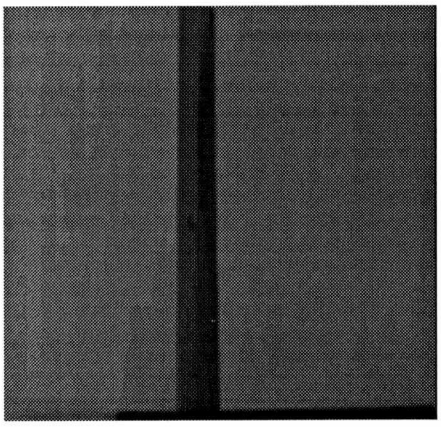
We define segmentation as the process of extracting geometrically significant regions or features from input images. The set of features that needs to be extracted in the segmentation process, as defined in Chapter 1, is parallel clusters of edge lines. Therefore, the primary responsibility of the segmentation algorithm is the synthesis of edge lines from a grey-level map.

8.1 Edge Detection

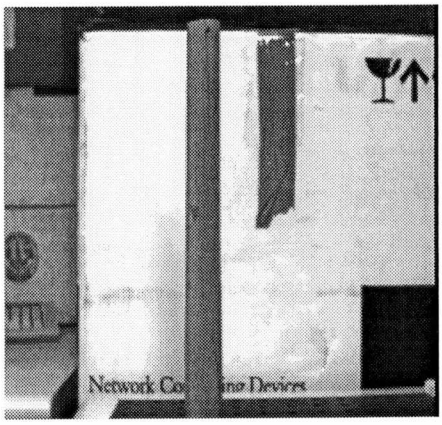
Given a raw grey-level image as in Figure 8.1 the initial task s to derive an edge map from the input scene. Edge detection is the most common approach for detecting meaningful discontinuities in gray level [GW87]. Several types of edge operators are available for selection: *Sobel*, *Marr*, *Canny*, etc. A comparative analysis of the Sobel [GW87] and the Marr-Hildreth [MH80] edge detectors as a function of performance in this particular system is described.

8.2 Threshold Generation for Sobel

The Sobel edge detection algorithm uses the superposition of two 3×3 gradient operators to produce an grey-level edge map as shown in Figure 8.3. Before the map is suitable for processing by the Hough transform, it must be binarized. Traditional means for binarizing the edge map involve the assignment of a threshold derived from repeated experimentation. This sort of approach works well when confronted with the

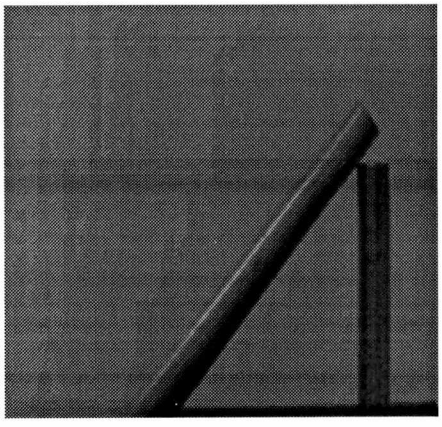


(a) Case 1

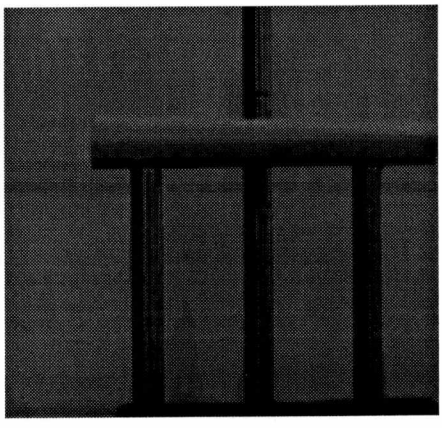


(b) Case 3

Figure 8.1: Input to Cylinder Detection System



(a) Case 2



(b) Case 4

Figure 8.2: Input to Cylinder Detection System

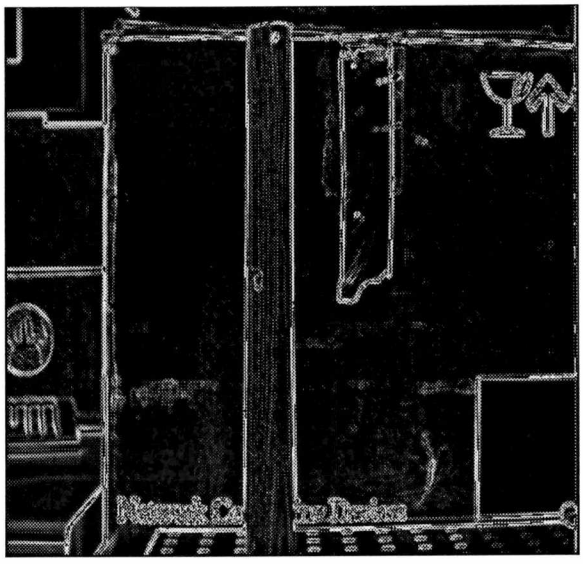


Figure 8.3: Sobel-Based Edge Detection

same type of scene; however, it will tend to fail when differing scenes are input. Ideally, we would like to process the edge map using some algorithm and then binarize it based upon that analysis. The proposed algorithm uses an iterative analysis of *edge-energy spectral density* to compute the appropriate threshold. It is conceptually similar to the process of experimentally determining the optimal number of clusters for an input data set via the *With Group Sum Squared Error* technique [DH73].

Suppose an input scene has some meaningful geometric edge structure that needs to be extracted. Application of the Sobel gradient masks derives an edge map similar to the one shown in Figure 8.3. Because the Sobel is a gradient operation, any sizable localized grey-level variation will trigger a response; consequently, large amounts of background noise and other forms of unwanted variations contaminate the image. Simply viewing the figure, however, reveals the relevant structure in the scene. Processing that will compute a threshold that suppresses the unwanted noise and reveals the true edge structure is needed.

We define a binarized edge map generated from an image as the sum of two responses from the edge detection step as

$$I = I_{ideal} + I_{noise}, \tag{8.1}$$

where I_{ideal} and I_{noise} are the response of the ideal image and the superimposed noise to the gradient operation respectively. Our binarization should trim as much of the noise as possible with degrading the structure of the "ideal". Logically, the magnitude of the noise response to the Sobel should be smaller than that of most of the significant edges in the scene. If we can define a measure across the global image space the represents the presence of edge information in the scene and compute this measure as a function of thresholding values, the measure should achieve a level of stability at some threshold value T that represents the place where most of the noise response has been removed. Suppose the input image has a set of L grey levels, $\{0, 1, \ldots, L-1\}$ associated with it. The following measure is defined

$$E(T) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |B(x,y)|^2 dx dy, \tag{8.2}$$

where $E(T)$ is a measure of the spectral density of the binarized image as a function of $T \in L$. $B(x,y)$ is defined by

$$B(x,y) = \left\{ \begin{array}{ll} 1 & \forall I(x,y) \geq T \\ 0 & \text{otherwise} \end{array} \right\}. \tag{8.3}$$

In the discrete domain Equation 8.2 becomes

$$E(T) = \sum_{i=0}^{m} \sum_{j=0}^{p} B(j,i). \tag{8.4}$$

Simply stated, it is a count of the number of edge pixels present in the $m \times p$ edge map after application of threshold T. If this process is repeated for all possible values of T from the set L the curve shown in Figure 1.5 results. It was stated that the energy

measure should experience a region of stability around the threshold values that cuts most of the smaller magnitude noise responses. Indeed, a step drop in edge density is experienced in the lower thresholding interval. This is immediately followed by a relative stability in the energy curve. We hypothesize that the region of stability in the curve is caused by the reduction of the random responses to the gradient operation. A simple analysis of the resultant curve will reveal this to be true. If gradient responses to noise consist of random, local, small-magnitude variations present in all regions of the image plane, for small threshold values the binarized edge image would contain a large amount of incorrect edge information. As the threshold value grows linearly, a corresponding reduction in local variation occurs *globally* in the image space. This would result in large reduction in edge density for small displacements of threshold. Once most of the global variations are removed, the threshold T begins to act locally, i.e. on the information, in the image. In this region the spectral density of the edge information should gradually tend toward zero as the threshold increases to its maximum value of $L-1$. The energy spectral density measure displays these properties. A threshold is chosen at the boundary of these regions from the following condition

$$T \ni \sum_{i=0}^{m}\sum_{j=0}^{p} |B_T(i,j) - B_{T-1}(i,j)|^2 \leq \epsilon, \tag{8.5}$$

where ϵ is terminal condition on the iterative search. Figure 8.4 displays the results of automatic segmentation of the raw Sobel map using $\epsilon \leq 2.5$. Figure 8.5 depicts the same scene after operation by the Marr edge detector.

8.3 Processing the Edge Maps

Now some of the possible operations on the edge maps that may produce significant improvements in the pipe detection process if they are successful will be considered . *Morphological* operators and *sub-sampling* will be considered. Binary edge maps from

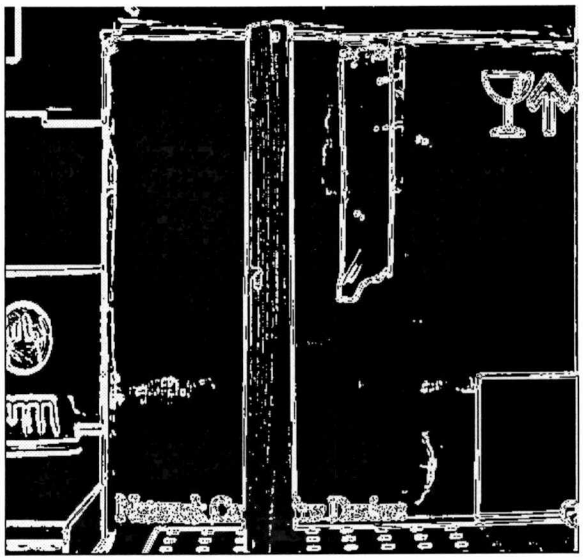


Figure 8.4: Binarized Sobel Edge Map

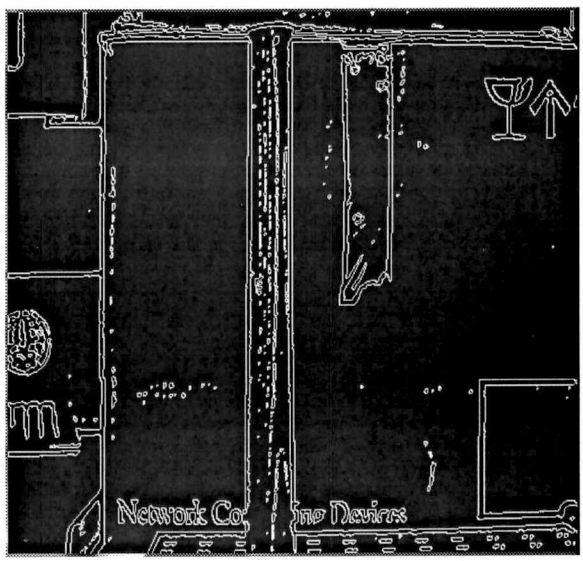


Figure 8.5: Marr-Hildreth Edge Detection: $\sigma = 1.0$

Figures 8.4 and 8.5 will be input to a preprocessing stage that involves a sequence of these operations. The resultant images will be feed into the Hough transform algorithms for line detection. Conclusions about the success or failure of these operations will be reserved until the results have been examined.

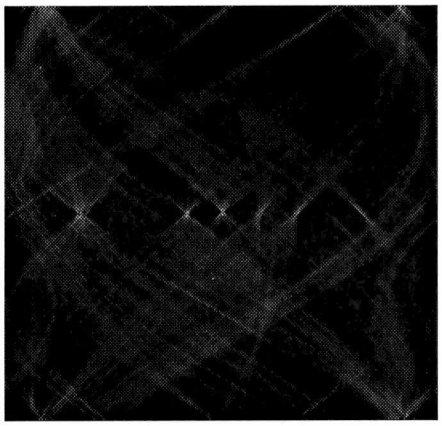
8.4 Conclusions on Edge-Based Segmentation

At this point, some quantitative measures of the effectiveness of the edge detection process are called for. Examination of the Figures 8.4 and 8.5 will reveal both algorithms have the same problem: edge localization. The Marr process produces the cleanest; however, the Sobel does require a thresholding step. The final conclusions will be derived from the comparison of the pipe detection results for each of the Sobel and Marr edge maps.

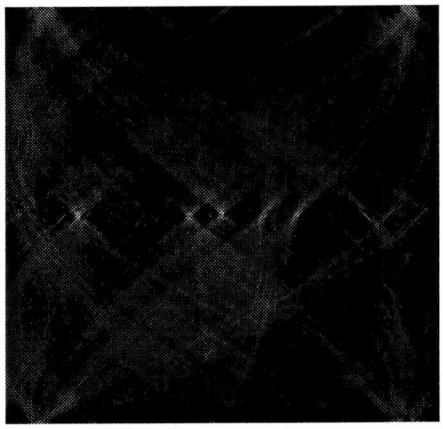
CHAPTER 9

Hough Space Processing

From the binary edge maps generated in Section 5, the Hough transform is implemented to produce all of the possible subsets of lines that can occur in the image space[GW87, PLJ92, VC92, Ati92]. Figure 9.1 demonstrates the Hough space representation of all such subsets while Figure 9.2 displays the Hough space histogram for each case. Our original definition of a cluster in Hough space as the symbolic representation of a mathematical edge line in image space creates some new difficulties. Significant "regions" in Hough space do not form congealed clusters at all; rather, they seem to have an annular or disk-like quality about them. A hasty evaluation would suggest this tendency to form rings in Hough space is a residue from the inability of the edge detection process to effectively localize edges in image space. However, artifacts of this type are present in images that were processed with both the Sobel and the Marr edge detectors. Since the Marr edge detector does not exhibit a spatial location problem, the cause of these effects must originate elsewhere. Answers are located from an examination of the original image in Figure 8.1. Near the edges of objects that occlude regions of the background or other features, false edges are found due to lighting effects. In general, these additional features are unwanted and need to be removed. This type of solution would require a geometrical knowledge of the scene lighting and stringent processing to remove it, and is, therefore, inadequate. If the false edges are allowed to remain some additional uncertainty is added to the recovery of line features from the image space and will contribute to accumulated error. Since we are forced to deal with this error, we must find a way to handle it with accumulation of the least error.

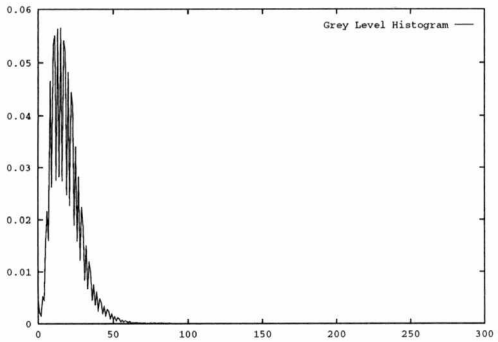


(a) Hough Transformation: Sobel

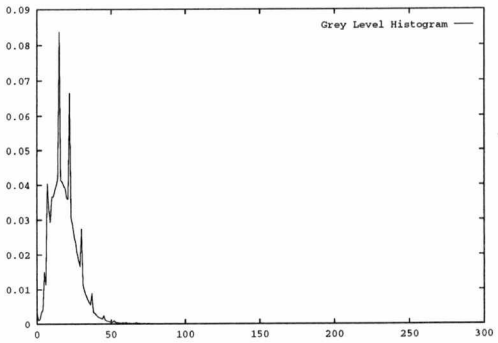


(b) Hough Transformation: Marr

Figure 9.1: Hough Transformation of Edge Maps



(a) Hough Histogram: Sobel



(b) Hough Histogram: Marr

Figure 9.2: Histogram of Hough Maps

Synthesis of clusters in a modified Hough space should preserve the inertial centers of the existing "regions" while simultaneously effecting an increase in the condition of each individual region. These regions also exhibit some tentacle-like projections that are a natural occurring phenomena in the Hough transformation process. By necessity they must also be removed. The task now becomes

- Improve the geometric quality of Hough space clusters.
- Remove unwanted "tentacles" from these regions.
- Suppress residual background information.

Conceptually at least, it would seem that these goals could be reached by some rotationally-invariant low-pass filtering operation. Problems arise in this approach due to the suppression of weak line features in the hough space. This can be overcome by the application of non-linear morphological filters that tend to grow only heavily clustered regions in the processing space. In the interest of completeness we will analysis the effects of three different methods for processing the Hough space. The first concerns the application of grey-level morphology to the raw Hough space.

9.1 Grey Level Morphology

Serra [Ser82] defines binary morphology as the application of a set of structuring elements taken one at a time to the *Hit or Miss* transformation. If the conditions on the image set by a particular structuring element are met, the Hit or Miss transformation responds with a binary true. Otherwise, the transformation replies with a binary negative. Two of the most basic morphological operations that can be applied to any image are erosion and dilation, which are based on Minkowski set addition and subtraction[PZLB93]. Binary erosion simply removes object pixels if they have neigh-

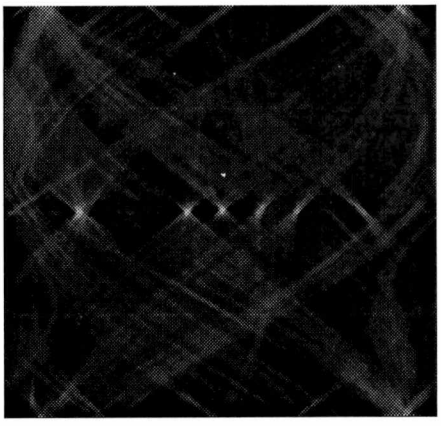
bors that are part of the background, while binary dilation, by duality, is an erosion of the background. Pairings of erosion and dilation operations result in two other fundamental operations termed *opening* and *closing* [PZLB93]. Opening a binary image consists of an erosion of the image followed by a dilation; conversely, closing is an erosion preceded by a dilation of the image. Both operators have a very strong intuitive meaning that is the basis of their use as size and shape descriptors[Ser82]. Opening removes fine features from the image, such as tentacles and small, spurious regions caused by noise. Closing tends to fill small gaps and channels in the image regions while thickening fine details. Both operators can be thought of as forms of image smoothing; however, they affect the geometric structure of image regions in significantly different ways. Since we wish to apply the basic set of morphological tools to process Hough space, binary morphology becomes inadequate due to the grey-level distribution of the Hough space. Here, one of two approaches can be used. First, the Hough space could be segmented using standard automatic thresholding algorithms; then, straight binary morphology could be applied. Or, more appropriately, grey-level morphology could be applied to the raw Hough space with the hope of isolating these high level cluster centers that will represent lines from the test scene.

An objective is to use simple grey-scale morphological algorithms to operate on Hough space images to extract components that adequately represent these clusters. Figure 9.3 displays the original Hough space for both Sobel and Marr edge detectors after the application of a $3x3$ grey-level dilation while Figure 9.4 contains the histograms. Grey-level dilation is defined as

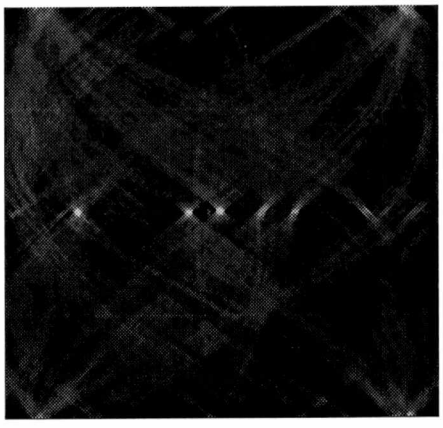
$$f \oplus b(s,t) = \max\{f(s-x, t-y)\}, \tag{9.1}$$

and grey-level erosion becomes

$$f \ominus b(s,t) = \min\{f(s-x, t-y)\}, \tag{9.2}$$

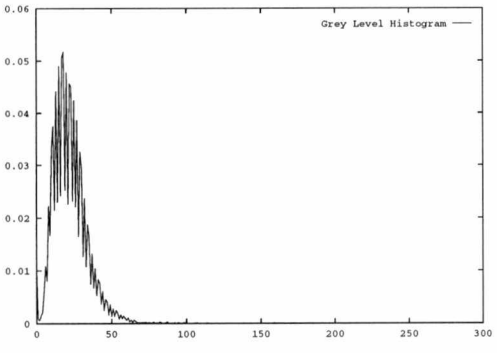


(a) Dilation of Hough Space: Sobel

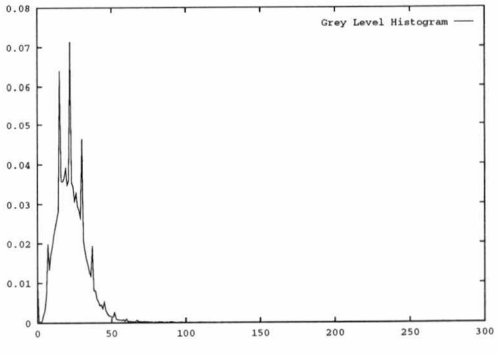


(b) Dilation of Hough Space: Marr

Figure 9.3: Hough Space Morphology



(a) Dilated Histogram: Sobel



(b) Dilated Histogram: Marr

Figure 9.4: Histogram of Dilated Hough Maps

where

$$(s-x),(t-y) \in D_f; (x,y) \in D_b. \tag{9.3}$$

Here $(s-x)$ and $(t-y)$ represent the sweeping of the mask over the image plane.

At this point it becomes important to have a quantitative measure of the success of the operation with respect to system goals. From the figures, the quality of the cluster centers undergoes a drastic improvement; however, some empirical proof that the simple morphological procedure has helped these efforts in some way must be found. A *Histogram Moment Extension* (HME) metric to extract this quantification is defined. Histogram Moment measures the relative change in histogram moment beyond some specified threshold. If the processing has succeeded, the HME should return a large positive percentage value. Histogram Moment Extension is defined as

$$\text{HME} = \frac{\sum_{i=T}^{L-1}(i-T)p_i^{\Delta} - \sum_{i=T}^{L-1}(i-T)p_i}{\sum_{i=T}^{L-1}(i-T)p_i}, \tag{9.4}$$

where T is the starting threshold, L is the number of grey levels, and p_i^{Δ} is the pixel count at the ith position in the histogram of the processed image. Grey-level dilations should create a marked increase in histogram moments beyond the threshold T, which in turn invokes a positive response from HME. Erosions would effect a decrease in the histogram moments and a negative response from the HME metric. Computing the HME for the Sobel operator followed by grey-level dilation provides

$$\text{HME}_{\text{Sobel}} = \frac{0.412 - 0.189}{0.189} = 1.179, \tag{9.5}$$

and Marr operator generates

$$\text{HME}_{\text{Marr}} = \frac{0.459 - 0.173}{0.173} = 1.653, \tag{9.6}$$

Table 9.1: Forward Histogram Moment Extension Metric

	Sobel	Marr
Case 1	0.78	2.34
Case 2	2.08	0.70
Case 3	3.52	0
Case 4	1.179	1.653

Table 9.2: Reverse Histogram Moment Extension Metric

	Sobel	Marr
Case 1	-0.030	-0.021
Case 2	-0.011	-0.010
Case 3	-0.028	0.0
Case 4	-0.108	-0.128

Each of the measures displays a increase in histogram moments beyond a specified threshold of 50 of in excess of 100%. Before concluding that the simple operation was a smashing success, we must verify that the change in opposing moments about the threshold point was not as severe. Again, we turn to the HME metric which responds with

$$\text{HME}_{\text{Sobel}} = \frac{27.678 - 31.062}{31.062} = -0.108 \tag{9.7}$$

for the Sobel, and

$$\text{HME}_{\text{Marr}} = \frac{27.289 - 31.305}{31.305} = -0.128 \tag{9.8}$$

in the Marr case. The counter-clockwise moments around the threshold experienced a slight shift toward that point; however, as the metric displays the relative change is not significant when compared to the shift in the clustered peaks. Examination of the histograms for each of the dilated Hough spaces in Figure 9.4 confirms this quantification. Comparing the histogram of the dilated Hough space with the histogram of the raw Hough space reveals some interesting characteristics of the processing. The first quality noticed is the tendency of the operator to not produce a noticeable shift in the distribution of the "background" content of the histogram. This becomes important since the quality of the segmentation will be directly linked to the separation of the background from the clustered peaks in the Hough space. Careful examination of the figures also reveals the most desired characteristic of the operator. Its application has produced a two-fold benefit. First, a noticeable upward shift in the grey level of the cluster peaks has been realized. Secondly, the annular clusters have been transformed into disks. Surprisingly, the HME metric confirms these observations. The symmetric nature of the operator has also preserved the inertial centers of the peaks. Localization of the peaks will constitute a segmentation of the Hough space followed by a "blob" analysis to recover the various geometric and shape properties of the peaks. Had we applied any operators that destroyed inertial measurements the accuracy of the process

would have been severely degraded. Inherent nonlinearity in the dilation process has allowed grey-level distribution of Hough space to be altered such that the peaks have been significantly enhanced while background information remains mostly unchanged; consequently, segmentation will require less stringent processing to achieve adequate results. Having postponed conclusions on the relative performance of the Sobel- and Marr-based edge detection process until some quantitative results could be derived, we can use the HME to draw some elementary conclusions. Equations 9.5 and 9.6 rate the responsiveness of each of the edge detectors to morphological massaging. Analytically, they are a measure of the edge detectors' ability to localize clusters in Hough space, or accordingly, lines in image space. When image space lines are appropriately localized, the Hough transformation generates compact clusters; however, when they are not well localized, segmented clusters may occur. Severely-disconnected clusters in Hough space may not be joined by the simple region growing filter; consequently, collinear line segments would be severed causing an increase in false detection rates. In each of the HME metrics, the ability of the Marr edge detector to effectively localize edges is observed. The metrics report greater clockwise and counterclockwise moment increases after the filtering process has been applied. In an ideal case, a decrease in the counterclockwise moment would be observed. Since the simple filter is not capable of such a complex response, a comparison with the counterclockwise moments from the Sobel operator confirm that the moment increase is approximately equivalent with that of the Marr. Amazingly, the clockwise moment of the Marr 50% above that of its Sobelian counterpart. Based on these observations, we can now draw the implied conclusion that the Marr edge detector is superior to Sobel in this application.

Table 9.3: Class Prototype of Constant Theta Vectors

	X	Y	Z
Class 1	0.0	7.3	0.0
Class 2	0.0	240.2	0.0
Class 3	0.0	472.03	0.0

9.2 Hough Space Segmentation

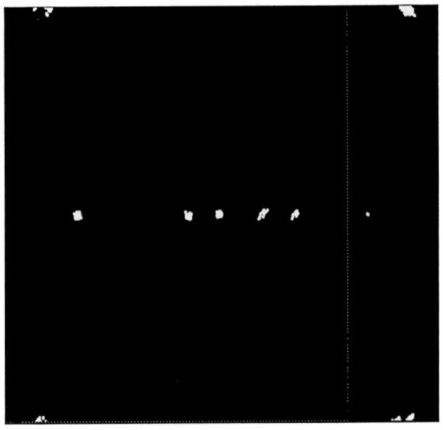
At this stage in the processing, the condition of the Hough space is critical. Cluster centers should be compact and maintain simple geometric shapes, and background intensity must be well separated from the intensity at the cluster peaks. Before a geometric analysis of cluster centers can be initiated, binary peaks must be separated from the background. Success in this segmentation step is crucial to the effectiveness of the stereo system to recover the real cylinders from the set of cylinder candidates presented to it. We consider two options:

- Hard thresholding of the processed hough space
- Segmentation using iterative segmentation procedure

Figure 9.5 reports the result of manual segmentation with a threshold $T = 90$ for both Sobel and Marr Hough spaces. From the figures, the observed regions satisfy the implied constraints of "compact" cluster centers and non-complex geometric shapes. However, some of the intra-cluster regions have been segmented. Again, this is a residue of the annular phenomenon discussed in previous chapters. If these segmentations are not too severe, simple morphological processing can be used to reconnect the related regions. Figure 9.6 results from the application of background erosion to the images of Figure 9.5. Naturally, the resolution of the system is affected by the additional

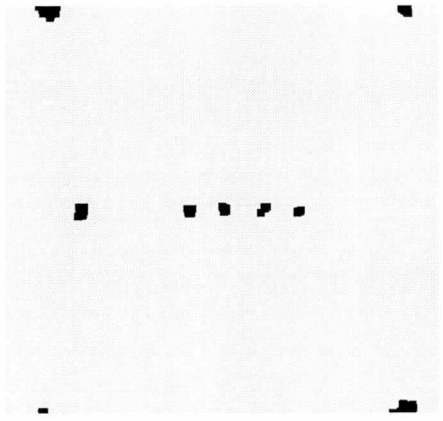


(a) Segmentation of Hough Space: Sobel

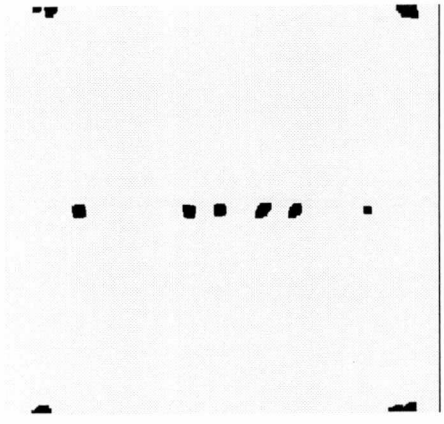


(b) Segmentation of Hough Space: Marr

Figure 9.5: Manual Segmentation of Hough Space



(a) Background Erosion : Sobel



(b) Background Erosion: Marr

Figure 9.6: Binary Region Growth in Manually Segmented Hough Space

processing. Care must be taken to avoid shifting the inertial centers of each cluster by a significant amount. Limiting these side effects implies the reconnection of the related regions without connecting any unrelated regions. Justifying the use of morphology at this point really boils down to a simple assumption. It is assumed that the distance between related regions in clustered hough space is much smaller the distance between unrelated regions. In general, this assumption may not be true, and is a function of a large number of variables. However, for most cases it is logical that the assumption holds. Given this, connecting these segmented clusters can be reduced to a simple region growth algorithm (i.e. binary dilation). Since the distance was assumed to be relatively small, dilation will grow each individual region and will connect regions that are geometrically related. Figure 9.6 confirms the hypothesis and reaffirms belief in the assumption. From the figure, it becomes clear that no unrelated cluster centers have been connected, while all related cluster centers have established spatial connections. Figures 9.7 and 9.8 return similar results for the automatic segmentation of Hough space. Recursive region growing produces the set of images contained in Figure lhspace. These images are input into the cluster selection mechanism to produce the cluster quality graphs shown in Figures 9.10 and 9.11.

At each step in the development of this system, we have tried to develop some sort of numerical "feel" for the error generated by certain critical assumptions or steps in the process. In segmentation, however, generating actual numbers depicting a measure of confidence in the success of the operation becomes difficult due to highly abstract nature of the application. No ground truth can be found with which to compare results with, and the non-linear nature of the processing preempts back-solving to a more stable point in the algorithm. In reality, the measurement of error, even with ground truth information, is relative to parameters or variables that are out of the experimenters' control. With this in mind, it would seem that much could be learned about the

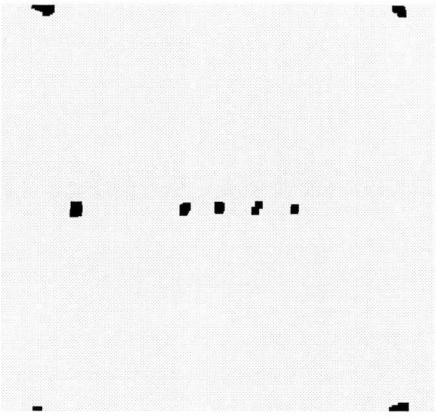


(a) Segmentation of Hough Space: Sobel

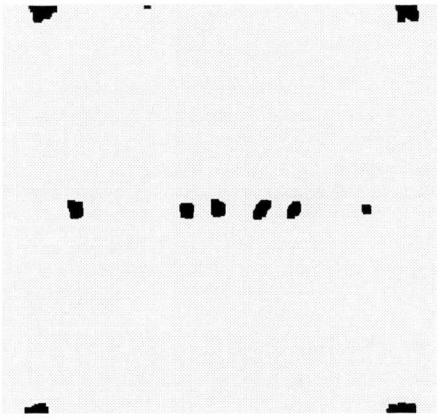


(b) Segmentation of Hough Space: Marr

Figure 9.7: Automatic Segmentation of Hough Space

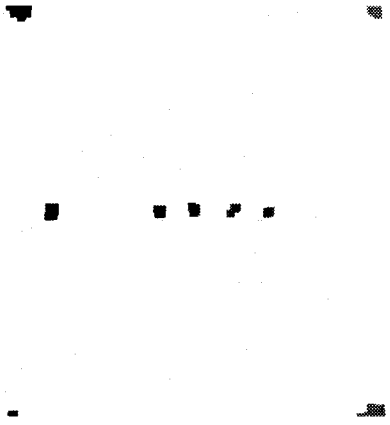


(a) Background Erosion : Sobel

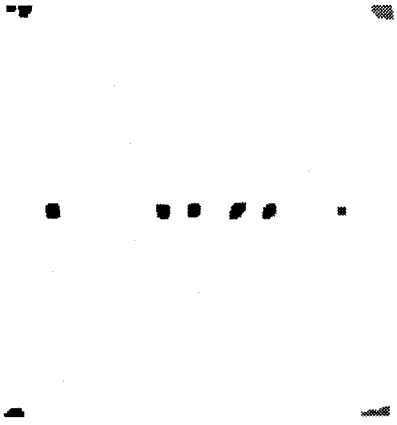


(b) Background Erosion: Marr

Figure 9.8: Binary Region Growth in Automatically Segmented Hough Space

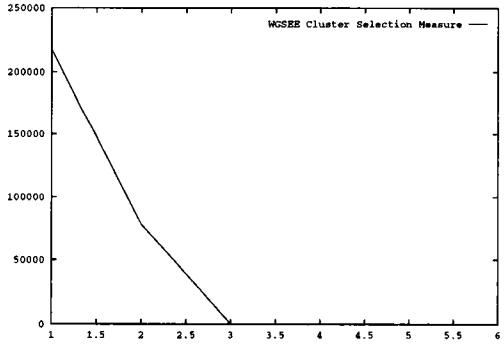


(a) Hough Space Regions: Sobel

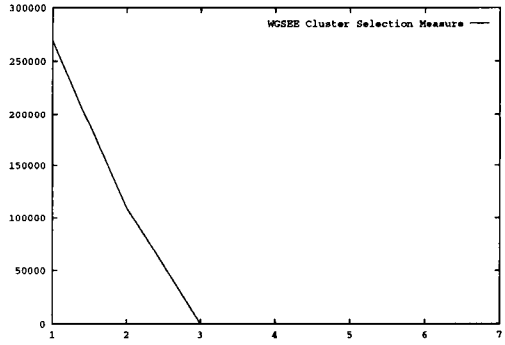


(b) Hough Space Regions: Marr

Figure 9.9: Labeled Hough Space

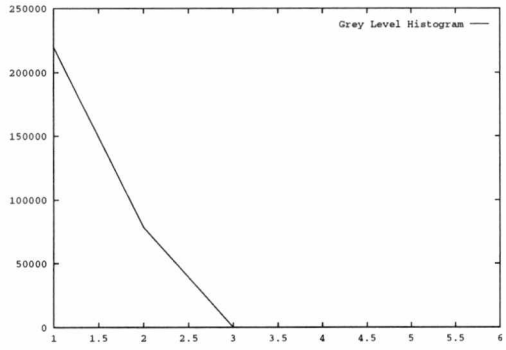


(a) WGSEE: Sobel

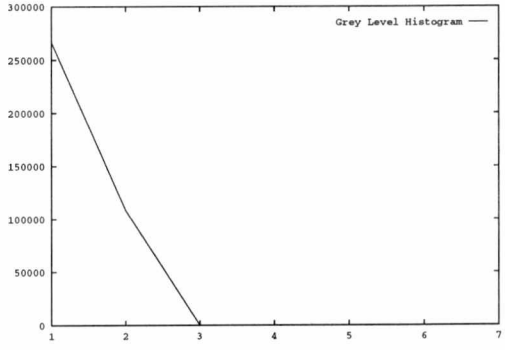


(b) WGSEE: Marr

Figure 9.10: Hough Space Cluster Selection Measure: Manual Segmentation

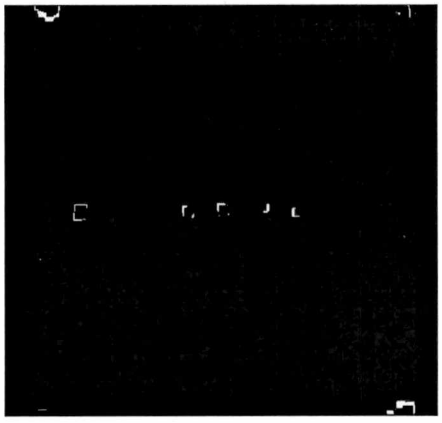


(a) WGSEE: Sobel

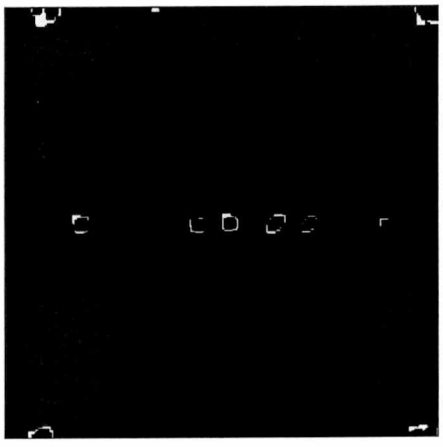


(b) WGSEE: Marr

Figure 9.11: Hough Space Cluster Selection Measure: Automatic Segmentation



(a) Error Image: Sobel



(b) Error Image: Marr

Figure 9.12: Manual vs. Automatic Segmentation of Hough Space

inherent error in a system by observing its response to different stimuli. Comparing the output with an understanding of the injected stimuli may reveal new information that can help in the performance assessment. Accordingly, the effects of segmentation will be observed on the Marr and Sobel generated Hough spaces. Also, it is assumed that the manual segmentation is representative of ideal information. Comparing the relative response of the segmentation algorithms with respect to each of these parameters should produce some insight into the repeatability of the segmentation.

A set of new images in Figure 9.12 was synthesized using the absolute value of the difference in the manually and automatically segmented images from Figures 9.5. Observing these "error" images will verify some of the initial assumptions that were used to formulate a processing plan. First, the ring-like nature of the resultant image regions suggests the preservation of the inertial centers of the blobs has been maintained. While some of the blobs do not exhibit completely uniform rings, the corresponding shift in the mass centers is miniscule with respect to the error tolerance. Finally, the morphological filtering has smoothed blob regions; consequently, the geometric quality of each of the blobs has been improved. While qualitative observations can be informative and more intuitive that numerical derivations, some sort of quantification must be produced that leads to a mathematical justification of the methodology.

An initial measure of the relative error in the process can be generated through a binary integration of each of the error images in Figure 9.12. Relative segmentation error, or RSE is defined as

$$\text{RSE} = \sum_{i=0}^{N} \sum_{j=0}^{M} e(i,j), \tag{9.9}$$

where $e(i,j)$ is an $N \times M$ error image. Table 9.4 sports the resultant cumulative error sum for the Marr and Sobel cases. Interestingly, the Marr operator reports the largest error sum. This may seem unusual at first glance. However, if we compare the annular rings that the Marr operator generated in Hough space with the more

Table 9.4: Relative Segmentation Error: Manual vs. Automatic Segmentation

	Case 1	Case 2	Case 3	Case 4
Sobel	0.0	0.0	0.0	554
Marr	0.0	0.0	0.0	868

compact rings from the Sobel, insight is gained into the measure. Marr edges are more spatially localized, giving rise to more fragmented cluster centers. Manual thresholding results in ideal conditions that replace these fragmented regions with smooth clusters. The automatic segmentation algorithm searches for a region of stability in the energy-spectral density of the error function. Since this is a linear search, stable threshold points by definition preclude the joining of these fragmented regions. Consequently, application of nonlinear morphological filters will generated large amounts of region growth due to the high surface area of segmented clusters. Increased region growth is directly proportional to the increase in the RSE measure. Although the RSE metric provides some insight into the nature of the relative differences between the automatic and manual segmentation procedures and may provide some vague idea of the absolute error if the manual segmentation is considered ideal; further analysis could be more revealing. Assume that two sets, A and B, representing the same binary image region are obtained from the manual and automatic segmentations respectively. An upper bound, or the maximum possible difference between the two sets, can be gained from

$$\epsilon_{\text{ub}} = \int A \bigcup B, \tag{9.10}$$

while the actual error between the sets is measured with

$$\epsilon = \int A \bigcup B - A \bigcap B. \tag{9.11}$$

The ratio of actual error to its upper bound is a scalar measure of the spatial and functional similarity between sets. If the ratio is small, the sets are similar; correspond-

ingly, a value near one implies dissimilar sets. We call this metric the *Region Similarity Measure* (RSM) and define it as

$$\mathrm{RSM} = \frac{\int A \bigcup B - A \bigcap B}{\int A \bigcup B} = \frac{\epsilon}{\epsilon_{\mathrm{ub}}}. \tag{9.12}$$

Expressed in its discrete form, Equation 9.12 becomes

$$\mathrm{RSE} = \frac{\sum_{i=0}^{N}\sum_{j=0}^{M} e(i,j)}{\sum_{i=0}^{N}\sum_{j=0}^{M} b(i,j)}. \tag{9.13}$$

Table 9.5 contains the binary integrations and the corresponding upper bounds derived from each image case. Table 9.6 reports the resultant RSM values. Amazingly, the values are nearly identical. Interpretation of the RSM values leads again to the confirmation of the initial assumptions. In both cases, segmentation produces clusters that respond with a linear increase in area to the region growing operation. Had the clusters been severely fragmented, we would expect the RSM to respond with widely varying numbers.

Although the RSM is a measure of the relationship between related image regions, it does experience some problems reporting the error between regions that have no intersection. In these cases, RSM returns the established upper bound on the error, completely ignoring the inter-cluster separation. Clearly, a measure is needed to reflect the increased error experienced when related regions do not overlap. Extending Equation 9.12 to include the inter cluster distance produces

$$\sum_{n=0}^{Q} \mathrm{d}(\overline{A}, \overline{B}), \tag{9.14}$$

where $\overline{A}$ and $\overline{B}$ represent the centroids of the related clusters A and B and Q is the number of cluster pairs. Real imagery may provide cases where a cluster A has no corresponding cluster B,. In these situations, the distance measurement is assumed to be zero by definition. Tables 9.7 and 9.8 contain the cluster center information for

Table 9.5: Integration of Segmented Hough Space(in pixels)

		Manual	Automatic	Sum
Case 1	Sobel	1792	2414	0.0
	Marr	495	1261	0.0
Case 2	Sobel	292	812	0.0
	Marr	1256	1352	0.0
Case 3	Sobel	2118	0.0	0.0
	Marr	0.0	0.0	0.0
Case 4	Sobel	2084	1530	3614
	Marr	2407	3275	5682

Table 9.6: Regional Similarity Metric

Sobel	0.153
Marr	0.152

Table 9.7: Sobel H-Space Lines: Manual vs Automatic Segmentation

Manual		Automatic	
Rho	Theta	Rho	Theta
49	7	50	5
42	475	43	476
87	241	87	241
215	240	215	240
255	238	256	239
302	239	302	239
343	241	345	241
469	472	471	474
470	6	469	6

each of the binary image regions located by a recursive labeling algorithm. Table 9.9 displays the corresponding Intra-Cluster Distance Similarity Metric, or ICDSM for each of the test cases. Clearly, the increase in the measured error from the Marr operation corresponds to the fragmented clusters found in its Hough space representation. It can be concluded that the morphological methodology is fairly sensitive to global cluster quality as indicated by the RSM. The ICDSM confirms initial suspicions that highly fragmented segmentation results in increased and non-uniform area growth which is a linear function of cluster surface area. While it is likely that fragmented clusters will be reconnected, the newly-joined region may experience spatial shifting of its centroid or it may exhibit abnormal area. Both phenomena correspond to the accumulation of error in the stereo processing phase. Now that unique Hough space cluster centers have been identified, a methodology must be employed to impose a structural grouping

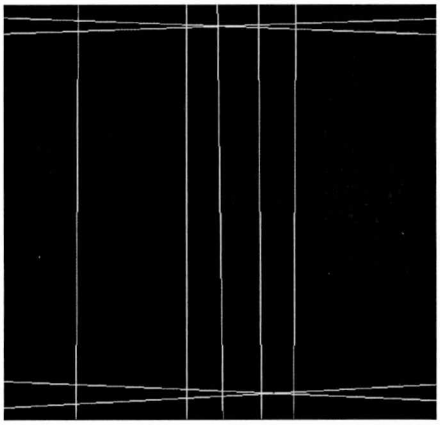
Table 9.8: Marr H-Space Lines: Manual vs Automatic Segmentation

Manual		Automatic	
Rho	Theta	Rho	Theta
42	473	41	473
36	3	46	8
52	6	86	239
86	240	170	2
217	240	217	241
253	239	253	238
304	239	304	239
342	240	342	240
427	240	427	240
470	474	469	473
475	7	474	9

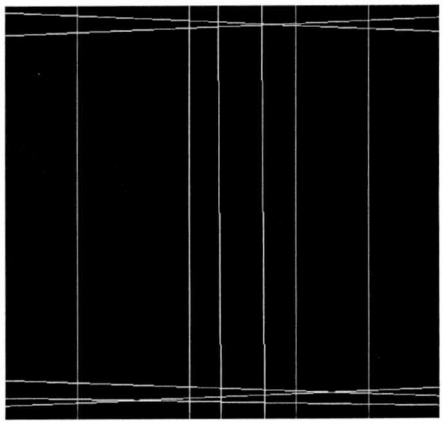
Table 9.9: Intra-Cluster Prototype Similarity Metric

Sobel	10.388
Marr	13.990

onto certain subsets of clusters that can be paired into possible cylinder candidates. Figures 9.13 and 9.14 display the line set generated by the linear feature extraction and the overlay of the features onto the original image set.

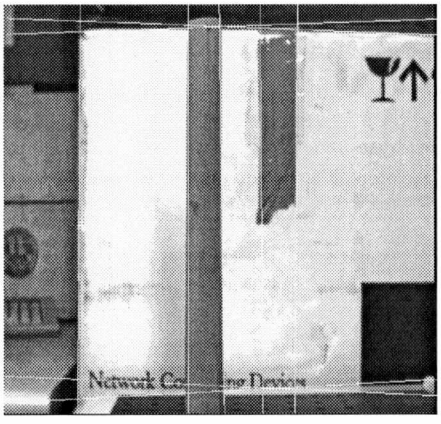


(a) Line Image: Sobel

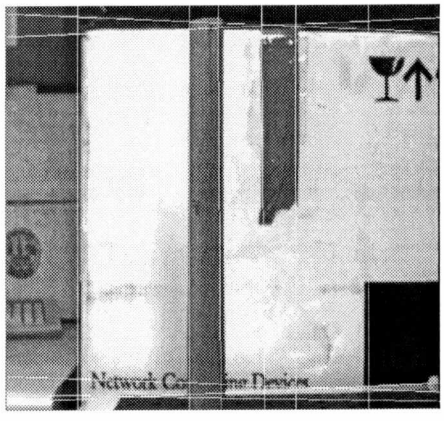


(b) Line Image: Marr

Figure 9.13: Synthesized Hough Space Line Images



(a) Overlay L: Sobel



(b) Overlay: Marr

Figure 9.14: Original Image with Synthesized Line Overlays

CHAPTER 10

Boundary Formation by Edge Line Clustering

Algorithms responsible for grouping individual Hough space clusters into related sets that represent possible pipe candidates must, by definition, display three important characteristics. First, the grouping algorithm must be tolerant to *foreshortening.* Foreshortening is a phenomenon associated with the perspective transformation that causes objects to appear smaller as their distance from the viewpoint increases. In candidate grouping, any object (especially cylinders) viewed at an angle with respect to the imaging surface will display foreshortened sections. If the objects are pipes, sets of parallel lines will be replaced by sets of lines converging at infinity. In Hough space, foreshortening generates a displacement of edge lines along the θ-axis. This displacement is accompanied by increased difficulty in the grouping process. In the ideal case a simple search for parallel sets of lines would produce the required sets; however, the θ shift preempts this type of primitive search operation. Secondly, the algorithm must uncover the optimal grouping strategy based only on the Hough space and any data it is able to derive from it. This statement implies that there is only one correct grouping of singleton clusters into related sets and the algorithm must discover it as a prerequisite to the success of the localization system. The final condition requires an unsupervised grouping technique. Unsupervised grouping techniques operate on raw input vectors and output a set of labelled vectors. Vectors that have the same label are thought of as having some structural relationship with respect to the grouping strategy. Less formally, examine the structure that can be derived from a collection of samples without being told their classification[DH73]. This type of procedure may require an iterative search

of the Hough space for this optimal grouping, perhaps trying all possible combinations and deriving the optimal grouping based on some error metric or criterion function.

10.1 Cluster Analysis

Serra [Ser82] relates that "to perceive an image is to transform it". Until now, the methodology has primarily been associated with processing and localization of edge lines in a set of input images. As Serra implies, some symbolic structure must be assigned to the input data set which permits higher levels of interpretation. Sorting individual clusters into related sets is the first step in the symbolic interpretation of the input data. Cluster analysis techniques will provide the required grouping.

10.1.1 Pattern Vector Extraction

Gonzalez claims that in any pattern analysis problem the primary task involves discovering an adequate set of discriminatory features from the observed data[TG74, DH73]. Or as Horn [Hor89] describes, the significant problem with clustering methods is the acquisition of enough information to allow generation of intelligent decision boundaries. Indeed, many indictments of cluster analysis techniques can be attributed to the inadequacy of the input feature vectors. A feature vector is defined as a group of measurements on the input data set. If the feature vectors reflect enough of the structural properties of the grouping that is to be imposed on the input data set, the cluster analysis techniques will perform well. The input data sets to the clustering algorithm will consists of the centroids of the clusters from the segmented Hough space, similar to that contained in Tables 9.7 and 9.8. The most critical step in the processing will be the selection of feature vectors from this centroid information.

Before extracting the required feature vectors from the input data set, a basic examination of what information the Hough space contains; what additional information,

if any, can be derived from the Hough space; and how the resultant data set relates to the grouping algorithm must be undertook. A good place to start this examination is with a basic definition of a successful grouping (clustering) algorithm. A clustering algorithm will be considered successful if it can operate within the constraints dictated by the four outlined conditions. For convenience, these conditions are

- Tolerance to foreshortening
- Tolerance to occlusions
- Uncovers an optimal grouping strategy
- Algorithm must be unsupervised

Of these conditions, only one, the first, is of concern in the feature vector selection stage of the algorithm. The second and third conditions are constraints placed on the actual clustering algorithm. If the clustering algorithms are to be tolerant to foreshortening, ideal feature vectors would be invariant under distortions caused by perspective foreshortening. Examination of Hough space, however, reveals three primary measurements that can be transformed into feature vectors. These are

- the θ coordinate
- the ρ coordinate
- the value of Hough space at the (θ, ρ) point

Unfortunately, each these measurements does exhibit change as the severity of the foreshorting increases. The clustering algorithm would gain great advantage if any of the foreshortening intolerant measurements were excluded from promotion to feature vectors. In image space, using a pair of lines to represent a cylinder does not force any

implied constraints on the ρ coordinate of either of the lines. This simply means that no boundaries other than physical are to be placed on matching individual lines into a cylinder candidate pair. In other words, no exclusions will be made based on inter-line pair distance; consequently, the grouping process is independent of ρ and it will not be made part of a feature vector. Similar arguments hold for the value of the Hough space function at the point (θ, ρ). Ideally, this measurement would reflect the average length of the edge line in image space pixels. While this is certainly an important measurement and will be used extensively in the later stages of processing, if exclusions were made based on average line length, occluded cylinders or partial lines caused by inadequate edge detection would be incorrectly paired. Only one measurement remains the θ coordinate. As mentioned, appropriate grouping would consist of a set of Hough space clusters that are contained within some narrow rectangular block of Hough space. This block, or region, would not be constrained with respect to the ρ coordinates of the clusters. Therefore, the task of the clustering algorithm is simplified. It must take a one-dimensional data set, the θ-coordinate of the clusters, and compile a sorted list consisting these rectangular regions and the Hough space clusters contained in them. Once this has been done, each of the regions can be individually analyzed for the presence of right circular cylinders by the region based stereo algorithm. Currently, the problem is to discover which clustering algorithm will produce the required discrimination of Hough space into these rectangular regions.

10.1.2 Graph-Theoretic Cluster Analysis

After extensive experimentation with the *Mean-Squared-Error* (MSE) clustering algorithm, and the consequent failure to assure convergence over the range of cluster centers, it was decided to change to graph theoretic cluster analysis. While clustering based on normal probability densities and minimum-error partitions tend to cluster in isolated

clumps of points, the inclusion of graph theory allows the exploration of more intricate structures[DH73]. A graph consists of a set of nodes, or points, with the dimensionality of the feature vector set, and a set of undirected edges with its associated lengths[Str86]. If the graph is connected to every point in the feature vector set from some initial seed point and there are no loops, the graph becomes a tree. A tree is said to span the graph of n feature vectors if it has $n-1$ edges[Str86]. Applying the parameter of edge length to the tree structure, a minimum spanning tree is a spanning tree such that the sum of all edge lengths is a minimum. The key idea in graph theoretic cluster analysis the generation of a minimum spanning tree. Once this tree is obtained, it can be broken into a set of individual trees, called a forest, by removing in succession the edges of greatest length. A forest, in this case, will represent the set of Hough space blocks where each block is a related set of edge-lines. Coupling the ideas of forests and minimum spanning trees, with the requirement of generating the set of related regions in Hough space, it would seem logical that the cluster analysis that is chosen would generate minimum forests. Generation of minimum forests will also be critical for the error measure used to provide completely unsupervised clustering capability.

Kruskal's algorithm displays the necessary components for the generation of minimum forests. The algorithm is stated as

- **Step 1**: List all arcs (x_i, x_j) in the order of increasing lengths.
- **Step 2**: Put the first arc (Shortest length) in the List $\mathcal{A}$.
- **Step 3**: Include the next arc from $\mathcal{A}$, unless its inclusion will give rise to a cycle, or loop, with the arcs already existing in $\mathcal{A}$.
- **Step 4**: If all nodes are connected with arcs from A then terminate; otherwise, repeat step 3.

- **Step 5**: Choose the desired number of clusters,k; and remove the kth largest edge lengths from $\mathcal{A}$.

Kruskal's algorithm is guaranteed to converge for any feature vector set[Str86]. It may seem at this point that the use of graph-theoretic cluster analysis is an ideal choice for generating the required Hough space regions; however, examination of Step 5 of Kruskal's algorithm will reveal the trouble spot. By definition, the algorithm must be told the number of trees, or clusters, that a given forest is to contain. Clearly, this is contrary to the constraint that the clustering must be completely unsupervised. An extension must therefore be produced that permits the cluster analysis, based solely on empirical observation to derive the appropriate number of cluster centers.

A constantly-recurring problem in cluster analysis is that of deciding just how many clusters are present in any given feature vector set[DH73]. An attempt to derive the actual number of clusters present in any data set can be made by repeating Step 5 of Kruskal's algorithm $n-1$ times and measuring the change in a criterion, or objective, function as a function of the number of cluster centers. If the feature vector set was well chosen, the criterion functions should inflect at the point where the appropriate number of clusters lies. As Duda relates in [DH73], if the n samples are grouped into $\hat{c}$ compact, well-separated clusters, one would expect to see the criterion function decrease rapidly until $c = \hat{c}$, decreasing much more slowly thereafter until it reaches zero at $c = n$. The criterion function referred to is the *Within-Group Sum of Squared Error* (WGSSE) measure. Mathematically, it is defined as

$$J(u,v) = \sum_{i=1}^{c} \sum_{x \in \mathcal{X}_i} ||x - V_i||^2, \tag{10.1}$$

where V_i is the mean of the feature vectors for the ith cluster, and $\mathcal{X}_i$ is the set of feature vectors associated with that cluster. Figures 9.10 and 9.11 display the WGSSE function for two test cases. From the graphs, a step drop in each metric is noted for $c = 3$.

Observation of Figures 9.7 and 9.5 confirms the correct cluster number of three. From here, a relatively simple analysis will reveal which of the candidate pairs are actually cylindrical objects within an assumed structural framework.

10.2 Simple Rules for Candidate Promotion

Synthesized line images in Figure 9.13 confirm the initial concern over ambiguity in the edge images. The overlay images in Figure 9.14 depict the original scene with the Hough space lines superimposed. Examination of this figure reveals which of the line pairs actually correspond to pipes, which arise from flat, rectangular images; and which are simply stray lines caused by background objects. Some rules must be imposed which operate within a structurally-based framework composed of certain assumptions about the scene. Once this framework is in place, regions that have a high probability of being cylindrical will be fed into a region-based stereo algorithm that will localize them in three-dimensional space.

CHAPTER 11

Region-Based Stereo

Projecting a three-dimensional scene onto an image plane via the perspective transformation is a many-to-one mapping[FGL87]. In other words, for any point in an image, there are an infinite number of possible corresponding world points. Given two images of this scene obtained from different viewpoints, it is possible to recover the missing coordinate information using binocular stereo. Binocular stereo uses the differences between the separate images of the scene to estimate the depth of objects captured in both images[Hor89]. A simple stereo geometry is shown in Figure 11.1. In the figure, (x, y, z) represent the location of the indicated corner of the box with respect to the coordinate frame located midway between the lens centers. (x_l, y_l) and $(x_r.y_r)$ represent the locations of the world point in the image space of the left and right camera systems respectively. Given these two image space points, the coordinates of the world point can be localized from

$$\begin{aligned} x &= b\frac{(x_l + x_r)/2}{x_l - x_r} \\ y &= b\frac{(y_l + y_r)/2}{x_l - x_r}, \text{ and} \\ z &= b\frac{f}{x_1 - x_r}. \end{aligned} \tag{11.1}$$

From Equation 11.2, b represents the distance between the lens centers and is called the *baseline* distance, while f is the effective focal length of the lens system.

While these simple stereo equations are useful for obtaining an intuitive picture of the binocular stereo process, they do not lend themselves to the accurate recovery of three-dimensional depth information. Much advantage would be gained using the results

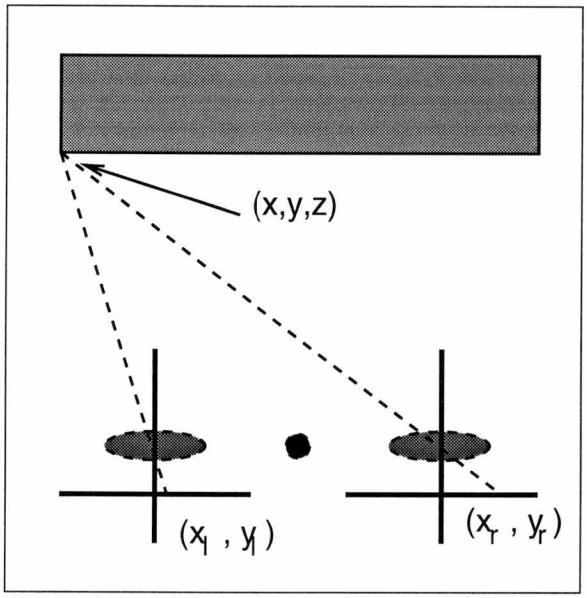


Figure 11.1: Basic stereo imaging geometry

of the calibration analysis from Section 2. Assuming that the system has generated a set of corresponding image points, one from each camera, that represent the same point from a scene, applying Equation 3.2 to both image coordinate vectors returns the three-dimensional coordinates of the projection of that image point onto a calibration plane. If information from two such calibration planes is generated, two vectors, each aligned with the actual world point and emanating from its respective camera is defined in the world reference space. The intersection of these vectors defines the location of the world point. Problems will arise because the synthesized vectors may not intersect. In this case, the point where the distance between the two vectors is a minimum is chosen as the location of the world point. Figure 11.1 graphically depicts the localization of a world point via this process.

It is assumed, at this point, that the system is given a set of corresponding points, one from image, and is expected to return the three-dimensional location of that point.

In the spirit of the process outlined above, Duda and Hart [DH73] suggest placing the location of the world point midway between the two vectors, $\vec{r}_1$ and $\vec{r}_2$, at the point where the distance between them is minimal. Two sets of coordinates vectors $(\vec{P}_{11}, \vec{P}_{12})$ and $(\vec{P}_{21}, \vec{P}_{22})$, returned from application of Equation 3.2 for both calibration planes join to form $\vec{r}_1$ and $\vec{r}_2$ through

$$\begin{aligned} \vec{r}_1 &= \vec{P}_{21} - \vec{P}_{11} \\ \vec{r}_2 &= \vec{P}_{22} - \vec{P}_{21} \\ \Delta &= \vec{P}_{12} - \vec{P}_{11} \end{aligned} \tag{11.2}$$

where $\vec{\Delta}$ is a vector in world reference space defining the relationship between the projections of image pixels from the left and right cameras with respect to the world coordinate space. Again, from Duda and Hart [DH73] the location of the world point should lie at the head of the vector, $\vec{v}$ defined by

$$\vec{v} = \frac{a_0\vec{u}_1 + (\Delta + b_0\vec{u}_2)}{2} + \vec{P}_{11}, \tag{11.3}$$

where a_0 and b_0 are defined as

$$\begin{aligned} a_0 &= \frac{\vec{u}_1 \bullet \vec{\Delta} - (\vec{u}_1 \bullet \vec{u}_2)(\vec{u}_2 \bullet \Delta)}{1 - (\vec{u}_1 \bullet \vec{u}_2)^2}, \text{ and} \\ b_0 &= \frac{(\vec{u}_1 \bullet \vec{u}_2)(\vec{u}_1 \bullet \vec{\Delta}) - (\vec{u}_2 \bullet \Delta)}{1 - (\vec{u}_1 \bullet \vec{u}_2)^2}. \end{aligned} \tag{11.4}$$

Figure 11.2 displays this solution in a more intuitive format.

11.1 Scan Line and Artificial Feature Generation

The only remaining task is the actual solution to the stereo problem. What is necessary for evaluation of Equation 11.3 is the generation of a set of feature points on the linear contours of the cylinder candidates. The set of feature points will be used

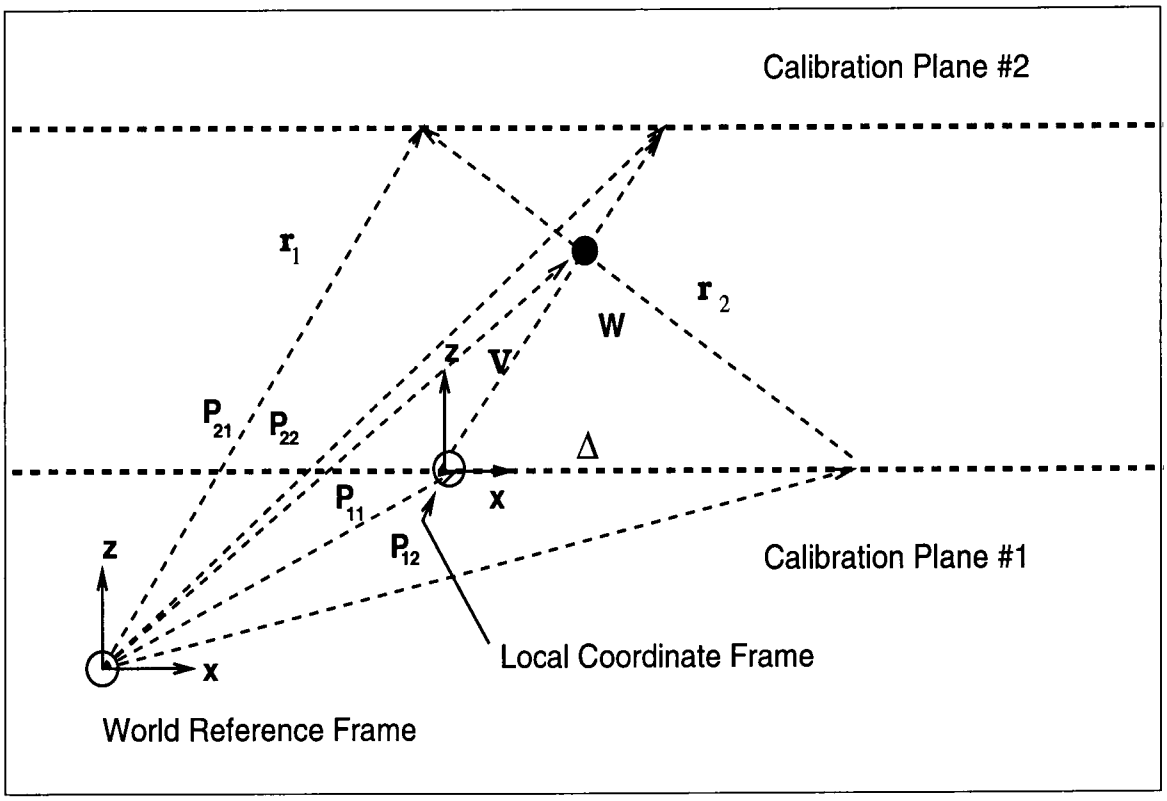


Figure 11.2: Solution to point of minimum distance stereo triangulation problem via two planes calibration procedure

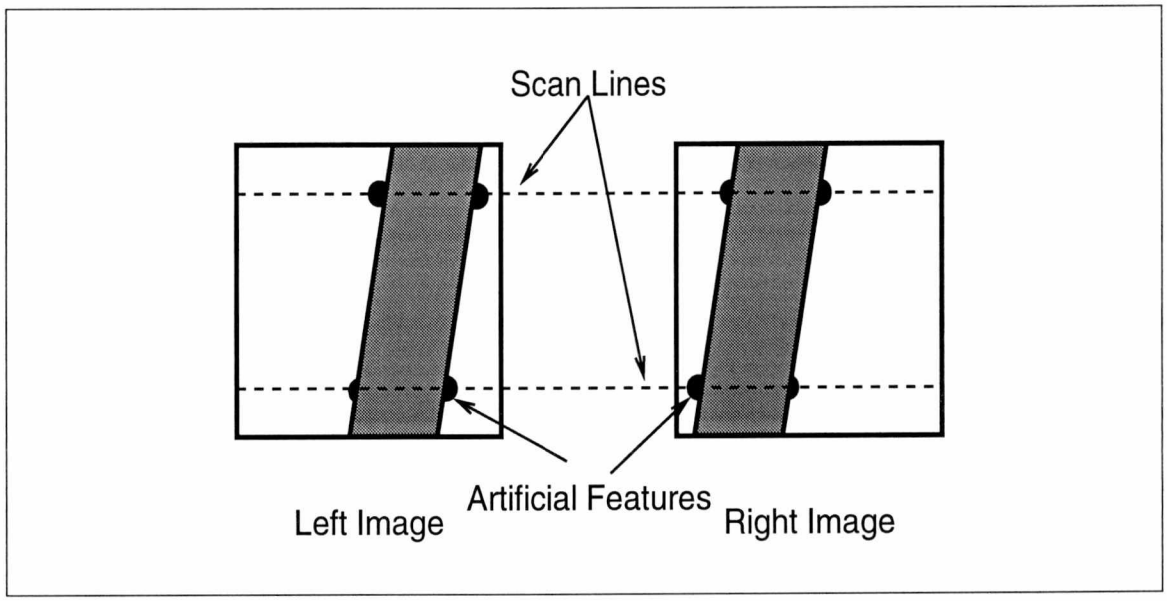


Figure 11.3: Solution of stereo correspondence problem showing scan-line and feature generation for foreshortened cylindrical object.

with the algorithms of the interactive stereo scheme of Chapter 7 to finalize the cylinder position. The methods of interactive stereo system require a human operator to isolate the feature points; however, in this system feature selection and correspondence will be autonomous.

Figure 11.3 displays an intuitive geometrical description of the scan-line generation problem. From the figure, it can be seen that two sets of these scan-lines must be produced. The intersection of these lines with the linear contours of the cylinder candidates will provide the initial set of features required for pose estimation. Scan-line generation begins in one image from the stereo pair. In this image, a seed scan-line is fixed perpendicular to the linear contour pair. The feature points on this line are located at the intersection of the scan-line with the linear boundaries of the candidate. Selection of the corresponding features in the remaining stereo image relies on the cross-correlation of a region built around the initial feature points and the corresponding set of linear

Table 11.1: Ground truth information for image I.

	X	Y	Z
Axis 1	-4.5	-6.0	14.5
Axis 2	-4.5	4.5	14.5

Table 11.2: Ground truth information for image II.

	X	Y	Z
Axis 1	1.5	-6.0	16.5
Axis 2	1.5	1.0	16.5
Axis 1	-4.5	-6.0	16.5
Axis 2	1.5	2.0	16.5

contours from the initial image. The isolated region is cross-correlated along a path that is parallel to the orientation of the linear contours of the candidate. Again, a match is found where the correlations responds with the largest numerical value. This process is repeated under the condition that a minimum distance between the scan-line pairs be maintained. When the final region has been matched, the feature points are selected from the scan-lines and feed into the stereo algorithms from Chapter 7.

11.2 Experimental Results

Four test case from Figures 8.1 and 8.2 were input the region-based stereo unit. Ground truth for each of the cases is contained in Tables 11.1, 11.2, 11.3, and 11.4. Tables 11.5, 11.6, 11.7,and 11.8 contain the range information returned from the stereo processing.

Table 11.3: Ground truth information for image III.

	X	Y	Z
Axis 1	-2.5	-6.0	16.5
Axis 2	-2.5	4.5	16.5

Table 11.4: Ground truth information for image IV.

	X	Y	Z
Axis 1	-4.5	-6.0	16.5
Axis 2	-4.5	1.0	16.5
Axis 1	-5.5	2.0	16.5
Axis 2	5.0	2.0	16.5
Axis 1	4.0	1.0	16.5
Axis 2	4.0	-6.0	16.5
Axis 1	-0.5	4.0	22.5
Axis 2	-0.5	-6.0	22.5

Table 11.5: Pose Estimation of image I.

	X	Y	Z
Axis 1	-4.43	-2.81	14.62
Axis 2	-4.45	3.04	14.58

Table 11.6: Pose Estimation of Image II.

	X	Y	Z
Axis 1	1.47	-3.79	16.59
Axis 2	1.53	2.79	16.67
Axis 1	FAIL	FAIL	FAIL
Axis 2	FAIL	FAIL	FAIL

Table 11.7: Pose Estimation of image III.

	X	Y	Z
Axis 1	-2.34	-3.79	16.67
Axis 2	-2.31	2.79	16.72

Table 11.8: Pose Estimation of image IV.

	X	Y	Z
Axis 1	-4.58	-3.8	16.61
Axis 2	-4.43	2.79	16.57
Axis 1	-5.51	2.0	16.59
Axis 2	5.11	2.0	16.51
Axis 1	4.11	2.79	16.42
Axis 2	4.15	-3.83	16.53
Axis 1	-0.57	2.79	22.59
Axis 2	-0.54	-3.83	22.65

CHAPTER 12

Conclusions

Analysis and design of two vision systems for three-dimensional cylinder localization were discussed. Development of each system began around a central design concept or philosophy. One unit, a human-machine system for interactive cylinder localization, attempted to build upon the existing perceptual power in human beings. Extending the concepts of telerobotics and human-machine systems to machine vision seemed a natural construct. Automatic vision for cylinder localization constituted the final design. Here, removing all forms of human interaction with the perception process was the fundamental motivation. Development of both systems was preceded by a discussion of some relevant issues common to both philosophical methods. This lead to the implementation of a calibration engine for three-dimensional object localization and some global derivations concerning the error induced by geometry assumptions.

Camera calibration using the Two-Planes algorithm was implemented and some extensions dealing with projections of vectors from the coordinate space of one imaging system into another were derived. Comprehensive error analysis techniques proved the numerical stability of the transformation under a variety of geometrical perturbations. Synthesized transformation matrices were recovered and found to be well-conditioned if their values are stored in long double precision format. Error intrinsic to the Two-Planes transformation within the defined region of operation was estimated to be 40 *mils* in the imaging space. Additional error was found to accumulate as a function of the uncertainty induced by the geometric assumptions used to simplify the system. The first assumption treated any projection of a cylinder in the image space as the projection

of the full diameter. The second dealt with error accumulated by assuming that two stereo views capture the same projection of the cylinder. Both of these error ratios were found to be (as a first-order approximation) functions of the diameter-to-standoff ratio. A general rule for controlling this error was derived. Lastly, spatial quantization error caused by the projection of the sensing element into the environment was examined. Within the region of operation, the maximum error introduced into the imaging system was found to be around 20 *mils*. A final estimate of the total transformation error was chosen as 60 *mils*.

Analysis of the human-machine system for cylinder localization involved both philosophical and theoretical considerations. First, a hierarchical analysis was performed on the localization requirements. From this, three specific strategies, or plans, were derived to serve as guidelines in the pose estimation process. Next, an overview of basic predictions on human performance in visual "pick-and-place" task was outlined. Also, theoretical and experimental extensions to human vigilance as it is concerned with these "pick-and-place" tasks were formulated. From these experiments, it was shown that within limits, the feature selection responds well under extreme lighting conditions. It was also shown that an inverse relationship exists between matching time and the accuracy of the recovered points. Statistical analysis of multiple experiments confirmed these results. Finally, algorithms for interactive stereo were developed. This resulted in generation of two different matching strategies. One relied on point-to-point feature matching, while the other employed a point-to-region strategy. A simple comparison of the matching algorithms was performed and each was found to produce acceptable results. Pose information was recovered with 3% error as a function of distance. The point-to-region algorithm returned the most consistent range information; however, it did experience some failing cases.

Implementation of the automatic machine vision system for cylinder localization

deviated significantly from the design principles involved in the interactive vision. Instead of relying on a human operator to extract relevant features, the automatic vision must have algorithms to extract these features without human intervention. This was accomplished through the use of edge-based analysis.

An algorithm for the segmentation of edge images was designed and implemented. This algorithm, used only on Sobel edge images, performed an iterative analysis of the energy spectral density as a function of threshold value. It searches for a region of stability in the energy of the edge image and chose a threshold at this point. The edge detection results via the Sobel or Marr operators were fed into the Hough transform.

In Hough space, morphological filtering was implemented to improve the geometric quality of the "feature" clusters. Several metrics were derived to measure the effects of morphological processing on the Hough space. These measures indicated that in all cases the morphological processing had shifted the upper levels of the histogram while minimizing the shift in the lower levels. This shift in grey level resulted in an improvement of Hough space cluster quality. Next, the Hough space was segmented via the automatic segmentation or by a simple "hard" threshold. After this, additional binary morphology was employed to cluster broken regions in the Hough space. A blob analysis produced points in the Hough space that represent lines from the image space. These points were grouped using an iterative graph-theoretic analysis. The pairs of pipe candidate points were then fed into the region-based stereo algorithm. The stereo algorithm worked well when the edge-maps recovered most of the significant lines from the input scene; however, if insufficient edge data is input to the blob detection unit, it may fail. Four test cases representing various degrees of difficulty were analyzed. Only one resulted in complete failure of the system due to inadequate edge detection. Range information was recovered at $4 - 6\%$ accuracy within the defined region of operation.

BIBLIOGRAPHY

BIBLIOGRAPHY

[Ati92] M. Atiquzzaman. Multiresolution Hough Transform - An Efficient Method of Detecting Patterns in Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1090–1095, November 1992.

[BG89] J. M. Bailey and G. K. Gorbet. Dynamic Modeling Used for the Addition of Robot Operation to the Advanced Servomanipulator. In *Proceedings of the IEEE Southeastcon*, 1989.

[CMN83] S. K. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction.* Lawerence Erlbaum Associates, Hillsdale, New Jersey, 1983.

[CSA89] M. S. Carey, R. B. Stammers, and J. A. Astley. Human-computer interaction design: the potential and pitfalls of hierarchical task analysis. In D. Diaper, editor, *Task Analysis for Human-Computer Interaction.* Ellis Horwood Limited, Chichester, GB, 1989.

[DH73] R. O. Dudu and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, New York, New York, 1973.

[FGL87] K. S. Fu, R. C. Gonzalez, and C. S. G. Lee. *Robotic Control, Sensing, Vision, and Intelligence.* McGraw-Hill, 1987.

[GW87] R. C. Gonzalez and P. Wintz. *Digital Image Processing.* Addison-Wesley Publishing Company, Reading, MS, 1987.

[Hal91] J. O. Hallset. Simple Visual Tracking of Pipelines for an Autonomous Underwater Vehicle. In *Proceedings of the IEEE Conference on Robotics and Automation*, volume 3, pages 2767–2772, Sacramento,California, 1991.

[Her91] C. J. Herget. Applications of Intelligent Telerobotic Control. In *Third IFAC International Workshop on Artificial Intelligence in Real Time Control*, September 1991.

[HJG71] W. C. Howell, W. A. Johnston, and I. L. Goldstein. Complex monitoring and its relation to the classical problem of human vigilance. In W. C. Howell and I. L. Goldstein, editors, *Engineering Psychology: Current Perspectives in Research.* Appleton-Century-Crofts, New York, New York, 1971.

[HN91] R. J. Holt and A. N. Netraval. Camera calibration problems: Some new results. *Computer Vision, Graphics and Image Processing*, 54(3):368–383, 1991.

[Hol71] J. G. Holland. Human vigilance. In W. C. Howell and I. L. Goldstein, editors, *Engineering Psychology: Current Perspectives in Research.* Appleton-Century-Crofts, New York, New York, 1971.

[Hor89] B. K. P. Horn. *Robot Vision.* The MIT Press, 1989.

[HST93] T. Hasegawa, T. Suehiro, and K. Takase. A Model-Based Manipulation System with Skill-Based Execution. *IEEE Transactions on Robotics and Automation*, 8(5), October 1993.

[IPS87] A. Izaguiree, P. Pu, and J. Summers. A New Development in Camera Calibration: Calibration of a Pair of Mobile Cameras. *International Journal of Robotics Research*, 6(3), Fall 1987.

[Jai89] A. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1989.

[KY91] A. Kaufman and R. Yagel. Toward a three-dimensional user interface. In A. Klingler, editor, *Human-Machine Interactive Systems.* Plenum Press, New York, New York, 1991.

[MH80] D. Marr and E. Hildreth. Theory of Edge Detection. *Proceedings of the Royal Society of London*, A(207):187–217, 1980.

[MSS92] O. Mohtadi, F. Safar, and J. L. C. Sanz. AGUILA: An Automatic Tube Detection System. In *SPIE Proceedings: Applications of Artificial Intelligence X: Machine Vision and Robotics*, volume 1708, pages 483–493, Orlando, FL, 1992.

[MT89] S. B. Marapane and M. M. Trivedi. Region-Based Stereo Analysis for Robotic Applications. *IEEE Transactions of Systems, Man, and Cybernetics*, 19(6):1447–1464, December 1989.

[MT93] S. B. Marapane and M. M. Trivedi. Mutli-Primitive Hierarchical (MPH) Stereo System. *Submitted to: IEEE Transactions on Pattern Analysis and Machine Vision*, Fall 1993.

[ND88] B. Noble and J. W. Daniel. *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[Nof85] S. Y. Nof. *Handbook of Industrial Robotics.* John Wiley & Sons, 1985.

[Nom92] Y. Nomura. Simple Calibration Algorithm for High-Distortions-Lens Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, November 1992.

[oE89] U. S. Department of Energy. *Environmental Restoration and Waste Management Robotics Technology Development Program: Five Year Program Plan*, volume II. 1989.

[OK93] M. Okutomi and T. Kanade. A Multiple-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:353–363, 1993.

[OW83] A. V. Oppenheim and A. S. Willsky. *Signals and Systems.* Prentice-Hall, Englewood Cliffs, NJ, 1983.

[Pla92] P. J. Plauger. *The Standard C Library.* Prentice-Hall, Englewood Cliffs, NJ, 1992.

[PLJ92] D. C. W. Pao, H. Li, and R. Jayakumar. Shape Recognition Using the Straight Line Hough Transform: Theory and Generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1076–1089, November 1992.

[Pre88] W. H. Press. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 1988.

[PZLB93] J. G. Postair, R. D. Zhang, and C. Lecocq-Botte. Cluster Analysis by Binary Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:170–180, 1993.

[Ron91] V. Ronchi. *Optics: The Science of Vision.* Dover Publications, Inc, 1991.

[Sea90] K. Storjohann and et. al. Visual Obstacle Detection for Automatically Guided Vehicles. In *IEEE Conference on Robotics and Automation*, 1990.

[Ser82] J. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, London, GB, 1982.

[She74] T. B. Sheridan. *Man-Machine Systems.* The MIT Press, Cambridge, MS, 1974.

[She89a] A. Shepherd. Analysis and training in information technology tasks. In D. Diaper, editor, *Task Analysis for Human-Computer Interaction.* Ellis Horwood Limited, Chichester, GB, 1989.

[She89b] T. B. Sheridan. Telerobotics. *Automatica*, 25(4):487–507, 1989.

[Str86] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley, MA, 1986.

[Swo84] E. W. Swokowski. *Calculus with Analytical Geometry.* Prindle, Weber and Schmidt, Boston, MA, 1984.

[SZ92] Y. C. Shiu and H. Zhuang. Pose Determination of Cylinders, Cones, and Spheres for Perspective Projections. In *SPIE Proceedings: Applications of Artificial Intelligence X: Machine Vision and Robotics*, volume 1708, pages 771–782, Orlando, FL, 1992.

[TG74] J. T. Tou and R. C. Gonzalez. *Pattern Recognition Principles.* Addison-Wesley Publishing Company, Reading, MS, 1974.

[Utt89] W. R. Uttal. Teleoperators. *Scientific American*, December 1989.

[VC92] V. Venkateswar and R. Chellappa. Extraction of Straight Lines in Aerial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1111–1114, November 1992.

[WCH92] J. Weng, P. Cohen, and M. Herniou. Camera Calibration with Distortion Models and Accuracy Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:1095–1099, November 1992.

VITA

Scott Mason Thayer was born in White Lake, South Dakota on April 21, 1969. He was awarded the Bachelor of Science degree in Electrical Engineering from the University of Tennessee, Knoxville in May of 1991. In August of 1993, he received the Master of Science degree in Electrical Engineering from UTK. Currently, he is enrolled in the doctoral program in Robotics at the University of Tennessee. His research interests include computer vision, sensor-based robotics, and real-time image processing.