OPTICAL RANGE-FINDING

FROM IMAGE FOCUS

By

PAUL REESE WECKLER

Bachelor of Science
California Polytechnic State University
San Luis Obispo, California
1982

Master of Science
Utah State University
Logan, Utah
1984

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
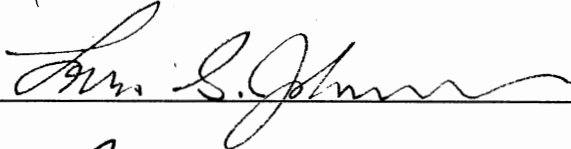the Degree of
DOCTOR OF PHILOSOPHY
December, 1989

OPTICAL RANGE-FINDING

FROM IMAGE FOCUS

Thesis Approved:

Glenn Kranzler

Thesis Advisor

Marvin L. Stone

Keith G. Teague

Rm. S. Johnson

Norman N. Durham

Dean of the Graduate College

ACKNOWLEDGEMENTS

Completing this dissertation elicits an emotional response akin to beating OU (preferably football). This section was possibly the most difficult part of my thesis to write, due to the many people who made my time in Stillwater so memorable.

A project of this nature required the assistance of many people. First and foremost, I am indebted to my advisor, Dr. Glenn Kranzler for his support, advice, and patience throughout my Ph.D. program. His guidance and attention to details were invaluable. I feel fortunate to have worked with him.

I'm continually amazed by my other committee member from Ag. Engineering. His diverse knowledge and infectious enthusiasm are remarkable. Without the assistance of Dr. Marvin Stone, this project would never have been completed. I also appreciate the consideration and patience of my other committee members, Dr. Louis Johnson and Dr. Keith Teague. Appreciation must also be extended to EG&G Reticon for the donation of a solid-state video camera, and to the UDSA National Needs Ph.D. Fellowship program which provided the financial support enabling me to pursue a Ph.D.

Numerous people played a significant role during my Ph.D. study. I would like to thank Bruce Lambert and Wayne Kiner for their help implementing the motorized camera-lens equipment, and to Mike Rigney for his assistance with the IRI computer. Thanks also go to Pamela Atherton

iii

for her typing, patience, and general good spirts. Her office has been my temporary home.

Professor Pat (Chicken) Lewis deserves special thanks for all his help. He and Dr. Bobby Clary helped keep me sane throughout this endeavor. Thanks to the many friends that made OSU such a rewarding experience, especially Cody Smith, who gave me a place to stay while I finished this dissertation. I would like to express my gratitude to Debbie White, who typed much of this thesis and spent many long evenings at the lab providing moral support.

Finally, I would like to dedicate this work to my Mom and Dad. At times when things looked bleak, their encouragement kept me going. Their support, understanding, and confidence in me was deeply appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of Problem

Three factors are significantly changing American business and industry. They are world competition, technological change, and consumer attitudes. Today, the marketplace for most items involves international competition, with the highest-quality, lowest-cost producer being most successful. Consumers want the best combination of high quality and low cost, regardless of national origin of the product. Adoption of new technology can reduce costs and increase the quality of products as compared with other producers. This result is true for agricultural products, as well as for manufactured items.

Agricultural crops such as grapes, apples, and oranges demand high labor requirements in production. Partially due to high labor costs in the United States, portions of these crops are now being imported from Chile, New Zealand, and Brazil, respectively (Krutz 1983). The status of intelligent machines and robotics in agriculture was reviewed by Sistler (1987), while Harrell, et. al. (1985) described a robotic tree fruit harvester designed to pick oranges. It has been widely noted that the adoption of robotics technology is one method of reducing costs and increasing the quality of the items produced.

1

Motivations for using robots vary by application, but a recent survey (Lewis 1983) lists the following reasons for adopting robotics technology:

1. Reduced labor costs

2. Elimination of dangerous and repetitious jobs

3. Increased output rate

4. Improved product quality

5. Increased product flexibility

6. Reduced materials waste

7. Reduced labor turnover

8. Reduced capital cost.

Much of the labor-intensive work in agriculture consists of reaching out, grasping an object, then placing the object in a desired position. This repetitious work exploits the unsurpassed hand-eye coordination in human beings. Substitution of machines for manual labor will require simulation of human hand-eye coordination. Most robots in agricultural applications will need the ability to recognize and manipulate three-dimensional objects. With present technology, this requirement makes agricultural robotic systems uneconomical, except for special applications (Pejsa 1983).

The image obtained by a machine vision system is a two-dimensional representation of a three-dimensional scene. This scene is usually captured by a planar solid-state image sensor which detects the intensity of the incident light. The third dimension (depth or range) must be obtained by another method.

Robotic depth perception (three-dimensional machine vision) may be broken into two basic techniques; triangulation and non-triangulation. These two methods may be subdivided into active and passive systems. Each method has advantages and disadvantages which dictate appropriate applications.

Most three-dimensional machine vision systems use two cameras for stereo triangulation or the projection of light patterns to obtain the third dimension. Jalkio, et. al. (1985) discuss the use of projected light patterns to obtain three-dimensional images. The paper examines the design and implementation of structured-light triangulation systems. Both stereo vision and structured lighting techniques require complicated software and extra equipment (second camera, laser lighting) to obtain depth information (McFarland 1983).

This research will examine a passive, non-triangulation technique to obtain the distance from the machine vision camera to the object of interest. A passive non-triangulation system has the simplest image aquistion requirements since it does not require a second camera, structured lighting, camera movement, or time-of-flight measurement equipment. This simplicity is balanced against the limited information obtained from a monocular view of the scene. For range measurement this method should provide adequate information for robot arm guidance. The simple hardware requirements are also an advantage when designing a rugged, cost-effective system for use in agriculture.

Objectives

The overall objective of this research was to investigate the feasibility of obtaining range data from a video camera using lens focus setting as an indication of object distance. This technique would yield information on the distance between the camera and object of interest without use of structured lighting, a second camera, camera movement, or time-of-flight measurement equipment.

The main objective was broken down into three sub-objectives. The first was development of software to control the equipment and to determine when the image is in focus. Calibration of the system to obtain the relationship between lens focus setting and object distance was the second sub-objective. The final sub-objective was to evaluate the system accuracy and investigate possible sources of error.

Key Assumptions

The research performed did not address the issue of object recognition. This is a very complex subject which currently is the topic of considerable research in computer vision and artifical intelligence. The research used a simple scene consisting of the object of interest. This constraint allowed the video camera to focus on the given object without having to search the image scene.

A practical robotic vision system will require object recognition ability. This research is only concerned with obtaining range data from the object of interest. Future work could use a zoom lens to allow a wide-angle

view of a scene at a short focal length. The object recognition software could then decide upon the object of interest, and the lens could zoom in on the object and perform the range measurements at a longer focal length. Object recognition should also be the subject of future research.

The camera lens used the largest possible aperture to minimize the depth-of-field. Depth-of-field is the distance interval in which an object appears to be in perfect focus. The smaller the depth-of-field, the more accurate the range data. Since depth-of-field is related to lens aperture size and focal length, the actual value depends on the lens used.

An important use of the range data would be the guidance of a robot arm. In this case the object distance from the video camera should be constrained to a length comparable to the robot arm's reach. Since an actual robot arm was not used, the maximum object distance was arbitrarily picked as three to four meters. The minimum object distance is constrained by the minimum focus distance of the lens. The minimum focus distance also is related to the lens aperture size and focal length. The focal length of the lens used required a compromise between being short enough to minimize the minimum focusing distance and long enough to give acceptable accuracy.

CHAPTER II

PRIOR RESEARCH

Introduction

Over the past ten years, hundreds of papers have been written on the subject of robotic vision and three-dimensional image processing. A very thorough discussion of three-dimensional object recognition and an extensive literature review have been performed by Besl and Jain (1985). They give a precise definition of object recognition, list qualitative requirements of recognition systems, and discuss emerging themes in various areas of three-dimensional imaging. Bajcsy (1980) reviewed the accomplishments and trends in the area of three-dimensional scene analysis. The paper covers current work in the following areas:

1. Three-dimensional data acquisition
2. Three-dimensional object representation
3. Software control structures for recognition of three-dimensional objects.

A general review of three-dimensional images for robot vision is given by McFarland (1983). Hall and McPherson (1983) have performed a review of machine vision techniques. They gave special emphasis to three-dimensional perception and methods for non-contact measurement of the coordinates or surface normals of objects using stereo, shading, and

projection techniques. A review of optical methods of three-dimensional sensing for machine vision was reported by Strand (1985).

Range-Finding and Distance Measurement

Various methods of noncontact distance measurement are being investigated by researchers in robotics and computer vision. An excellent review of noncontact distance sensor technology is given by Koenigsberg (1983). He classifies the sensors into three categories, depending on mode of operation; mechanical, electromechanical, and electro-magnetic. Various techniques of noncontact range-finding are considered in a survey article by Wolpert (1987), while methods for mobile robots are reviewed by Everett (1987).

Optical ranging systems for robots have been developed recently by Ozeki, et. al. (1986), Okada (1982), Pipiton and Marshall (1983), and Kinoshita, et. al. (1986). The first three papers describe projected light systems using triangulation to obtain range data, while the last paper discusses a novel projected light ring and focusing technique. Kanade and Sommer (1983) describe a proximity sensor for robotics use based on active illumination and triangulation. Another triangulation-based, three-dimensional sensing scheme claims to reduce the inherent trade-off between resolution and depth of focus (Bickel, et. al. 1985).

A distance sensing method that measures the phase shift of an array of sinusoidally modulated laser diodes is described by Cathey and Davis (1986). It has a range resolution of 15 centimeters at a distance of 13 meters and is intended mainly for automated space-station docking. A novel technique of range-finding using diffraction gratings or holograms

has been developed by De Witt (1988). A comprehensive review of range-finding techniques for computer vision was given by Jarvis (1983). Some of the optical ranging methods described in the previous paper have been applied to autofocus 35-mm cameras (Orrock, et. al. 1983).

Three-Dimensional Machine Vision

Researchers in three-dimensional machine vision have proposed several generalized categories for the techniques used to obtain the third dimension of an image. One such categorization (McFarland, 1983) suggests three basic approaches:

1. Stereo views
2. Range images
3. Structured light projections.

Another categorization scheme disscussed by McFarland (1983) groups methods into systems that use triangulation techniques and systems that do not. Further classifying identifies whether the system is active or passive. The paper by McFarland (1983) ultimately groups three-dimensional imaging techniques into four categories:

1. Passive triangulation systems
2. Passive non-triangulation systems
3. Active triangulation systems
4. Active non-triangulation systems.

He discusses each category. One of his conclusions is that the passive non-triangulation techniques give limited data, but offer the advantage of the simplest image acquisition requirements.

Image Focusing and Autofocus

One passive non-triangulation technique frequently mentioned utilizes focus-derived measurements to obtain depth information. Distance to an object may be estimated from an in-focus image of that object, if the focal length of the lens and the distance to the image plane are known. The process of focus measurement for depth data is being used in microscopy (Jarvis 1976). Krotkov (1986, 1987) investigates the use of lens focusing to compute the absolute distance from the lens to a sharply imaged object. He performs a diffraction analysis and a geometric optical analysis of image defocus, and proposes nine different criteria functions for measuring the quality of focus. The type of lens model, windowing of the image, and the effect of lens zoom setting are discussed.

There has been much research in the area of camera autofocusing. Schlag, et. al. (1983) investigated various algorithms for automatic focusing of a computer vision system. Jarvis (1983) describes automatic focus television cameras that evaluate scene sharpness and adjust the focus to maximize contrast. A simple focusing system built around a 256-element photodiode array has been developed by Selker (1983), while Honeywell Corp. (Shazzer and Harris 1985) has developed monolithic image processing chips to implement their autofocus technique used in FLIR (Forward Looking Infrared) imaging.

A companion research area is that of determining image quality. Schade (1975, 1987) produced the definitive work on the comparison of photographic and television image quality. His pioneering research at RCA Laboratories during the 1940's and 1950's led to many of the common concepts used today for evaluating imaging parameters.

Fourier Transform and Other Transforms

Geometrical and diffraction analyses show that the condition of defocus results in the attenuation of high spatial frequencies. This well-known fact reveals that the degree of defocus varies inversely with the amount of high spatial frequency energy present in the spatial frequency spectrum (Born & Wolf 1983). For measuring the quality of focus, Krotkov investigated criteria functions which respond to high frequency content in the image.

The Fourier transform of an image directly provides the spatial frequency distribution of that image. In 1968, Horn used the Fourier transform to investigate automatically focusing a vidisector. Work by Schlag, et. al. (1983) considered using the Fourier transform as an autofocusing algorithm, but after analyzing the number of computer floating-point additions and multiplications required, the idea was abandoned.

The Fourier transform was not considered for implementation by Krotkov because of computational complexity, superfluous data (ie. magnitude and phase), and lack of guidance regarding procedures when little high-frequency energy is present in the frequency spectrum. The other eight focus measuring methods proposed by Krotkov were spatial domain techniques. Of these eight, six were implemented and evaluated (Krotkov, et. al. 1986).

Other orthogonal transformations such as the Hartley, Discrete Cosine, and Walsh-Hadamard have not been used for the purpose of determining image focus quality. This omission is most likely due to their

complexity and to the amount of computational time required. Fundamentals of Walsh functions and Hadamard matrices are covered in the paper by Ahmed, et. al. (1971), and in the books by Harmuth (1977) and Beauchamp (1984). Application of the Hadamard transform in image processing was proposed by Pratt, et. al. (1969) and is now a standard topic in most image processing books.

VLSI Image Processing Chips

With increases in computer performance, the development of digital signal processing chips, and the ability to put dedicated algorithms into silicon, the use of various orthogonal transformations should not be overlooked. An example is a commercially available digital signal processing chip that can compute a complex 1024 point fast Fourier transform (FFT) in 2.4 milliseconds.

Many very large-scale integration (VLSI) chips are being developed specifically for image processing applications. Sugai, et. al. (1987) have developed an image processing chip that implements the FFT, Affine transform, spatial filtering, and histogram operations. It can be connected in parallel to increase the processing speed on certain algorithms such as the FFT. A chip that computes area, center of gravity, orientation, and size on gray-scale images in real-time (60 Hz) was implemented at AT&T Bell Labs (Anderson 1985). The chip was tested in a vision system connected to a robot arm that caught ping-pong balls as they were rolled across a table. An image processor used for image coding and video bandwidth compression was reported by Hein and Ahmed (1978). It implemented the Discrete Cosine, or Walsh-Hadamard transform, in real time.

Two novel techniques for performing Fourier and Walsh-Hadamard transforms are discussed by Kowel, et al. (1979) and Yarlagadda and Hershey (1981). The first paper describes an image sensor called a "Direct Electronic Fourier Transform" (DEFT) device. The sensor uses a photoconducting film deposited on a piezoelectric substrate. Two orthogonally directed surface acoustic wave transducers modulate the image light-induced electric charges, creating an electronic signal in the general form of the two-dimensional Fourier transform of the image intensity.

The second paper (Yarlagadda and Hershey 1981) describes a method of performing the Walsh-Hadamard and Fourier transform with charge transfer devices. The authors state that charge transfer devices are suited for many signal processing tasks where cost and simplicity, rather than speed, are the premier factors. This conclusion tends to exclude image processing from consideration due to the high data rate of information contained in images.

CHAPTER III

THEORETICAL BACKGROUND

Image Formation and Optics

When a scene is viewed from a fixed location, the light received by the observer varies in color and brightness, and may be expressed as a function of direction. Scene brightness and color are resultants of the illumination, reflectivity, and geometry of the scene.

In an optical image produced by a lens, light rays from each scene point in the field of view are collected by the lens and brought together at the corresponding point in the image. Figure 1 shows a simple thin-lens imaging system.
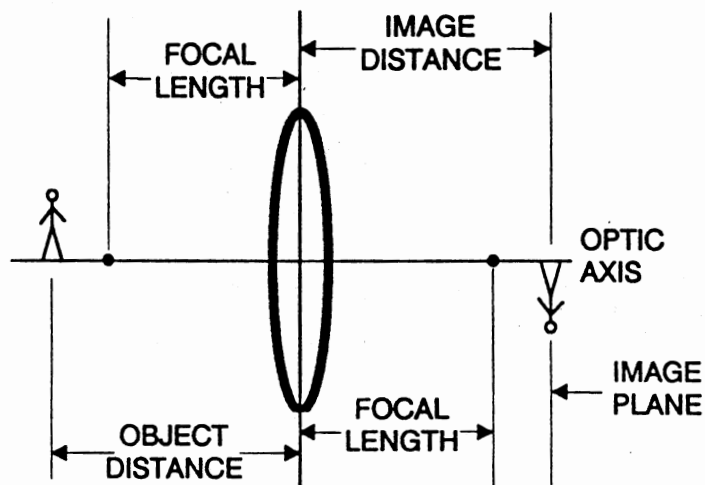
Figure 1. Simple Thin-Lens Imaging System

Scene points at different distances from the lens give rise to image points at different distances. From geometrical optics, the basic equation for a thin (Gaussian) lens is;

$$\frac{1}{f} = \frac{1}{S_i} + \frac{1}{S_o} \tag{3-1}$$

where, S_i is the distance from the image point to the lens

S_o is the distance from the object point to the lens

f is a constant called the focal length of the lens.

If S_o is large, ie. the scene points are all relatively far from the lens, $1/S_o$ is negligible, and $S_i \approx f$. This condition places the image points at about the same distance from the lens, near the lens focal length. Depending on the focal length of the lens and the distance the scene points are from the lens, the "image plane" and "focal plane" are at approximately the same location. Thus, the image formation process converts the scene information into an illumination pattern in the image plane. This illumination pattern is a function of two variables which are the coordinates of the plane. A general discussion of optics and imaging can be found in Hecht and Zajac (1979) and Ballard and Brown (1982), while detailed analysis and derivations have been performed by Born and Wolf (1983).

The light received by an optical system produces a two-dimensional image of what most likely is a three-dimensional scene. This

image must be converted into an electrical signal form by a sensor or recorded photographically as a picture, to be of practical use. Ballard and Brown (1982), as well as Gonzales and Wintz (1977), briefly describe various image sensing devices, while Tseng, et. al. (1985) describe the evolution of the solid-state image sensor. Operation of the many image sensing devices available will not be further discussed here.

Camera Models and Calibration

Once the image has been captured by the sensing device, it can then be used for some practical purpose. If the image is to be used in a three-dimensional machine vision or robotics application, there must be some form of camera model developed from calibration data (Shafer 1989).

A camera model defines the relationship between a point in the real world (global point) and the corresponding point in the acquired image, with respect to the location of the camera in the real world. Using the camera model, the position of objects in a scene can be determined by measurement of the features in the image. Ballard and Brown (1982) mathematically describe a calibration procedure, while Sobel (1970) uses a simple photogrammetry camera model to calibrate a computer-controlled, moveable camera on a pan-tilt head.

Most camera models use the traditional pinhole-camera projection geometry, in which the lens is modeled as an infinitesimally small aperture. A camera model developed by Potmesil and Chakravarty (1982) approximates the effects of the lens and aperture functions of a real

camera. This model adds certain optical characteristics of a lens, such as the effects of diffraction, to the usual ray-tracing (pinhole) camera model.

Tsai (1987) has developed a versatile camera calibration technique for three-dimensional machine vision systems that uses standard imaging components. It includes effects due to image scanning electronics, effective focal length, and radial lens distortion.

Image Quality

The focusing of light onto an image may be interpreted through diffraction theory as the convolution of the actual scene radiance with the transmission properties of the lens. The end result is to "smear" or "blur" the image. Geometric optics rely upon ray-tracing to explain the blurring, whereas diffraction theory can elegantly explain the cause of the blurring. Diffraction theory gives results that can easily be explained using the mathematical tools of Fourier analysis.

When an image is formed by a perfect lens (a lens with performance limited only by diffraction effects) from a point source of light, the image intensity pattern is broadened as a result of diffraction. This pattern consists of a central bright spot and alternating light and dark concentric rings radiating from the central spot. This form is generally referred to as the Airy pattern, with the central bright spot called the Airy disk (Hecht and Zajac 1979).

For a circular lens the radius of the Airy disk is;

$$r = 1.22 \frac{f\lambda}{d} \qquad (3\text{-}2)$$

where, r = radius of Airy disk

f = focal length of lens

λ = wavelength of light

d = aperture diameter.

For a typical human eye in bright sunlight;

$d = 2$ mm   (pupil diameter)

$\lambda = 550$ nm   (green light)

$f = 20$ mm.

The resulting radius is 6.7 um, or roughly twice the mean distance between photoreceptor cells (Hecht and Zajac 1979).

A point source of light is imaged on the retina as a circular disk and is called the "circle of confusion". Due to the limited resolving power of the eye, it cannot distinguish between a sharp image point and an out-of-focus point until the diameter of the latter exceeds the diameter of the circle of confusion. Thus, there are distance zones extending from either side of the object plane in which the image appears to be equally sharp. These distance zones are called the depth-of-field. The depth-of-field for a thin lens is shown in Figure 2.
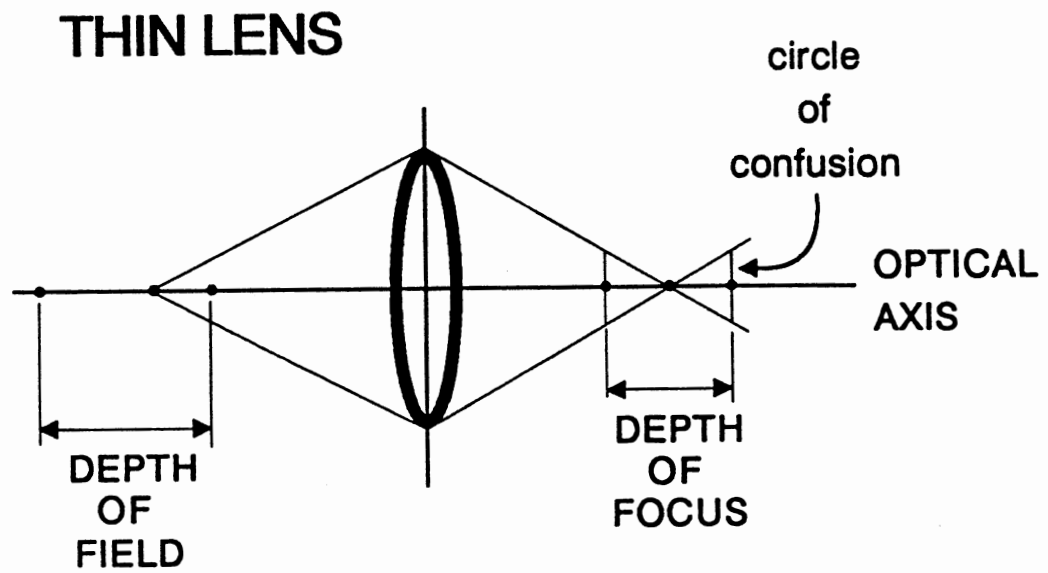
Figure 2. Depth-of-field for a Thin Lens

The depth-of-field may be thought of as the distance interval in which an object appears to be in perfect focus. From simple geometric optics, the depth-of-field can be calculated;

$$D = \frac{X}{1 - XCN/f^2} - \frac{X}{1 + XCN/f^2} \tag{3-3}$$

Where;   D = depth-of-field

N = f-stop number setting of the lens

f = focal length

C = diameter of circle of confusion

X = distance to the object.

The distance interval is a function of lens focal length, aperture, and object distance. Depth-of-field is not a linear function. For a 135 mm lens with the f-stop at f/4, the depth-of-field is approximately 40 mm for an object at 1500 mm, and 110 mm for an object at 2500 mm. This relationship is shown in Figure 3.
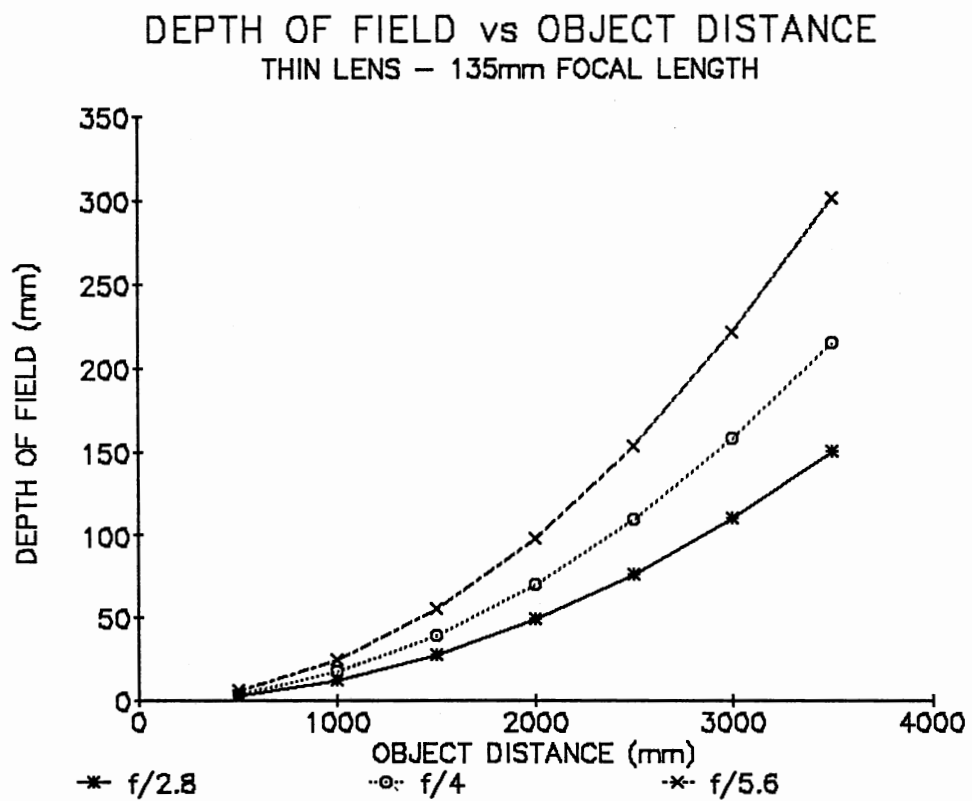
Figure 3.   Depth-of-field Versus Object Distance for a 135mm
Focal Length Thin Lens.

Fourier Analysis

One of the major advances in the field of optics during the past 30 years has been the application of system concepts and information theory to optical imaging. Optical devices consisting of lenses, mirrors, prisms, etc. may be considered to provide a deterministic transformation of an input light distribution to some output light distribution. Linear system theory and Fourier analysis are very useful tools to explain the behavior of optical systems (Gaskill 1978).

Sometimes it is easier to analyze a function by representing it as a set of numbers which, when expressed in terms of a properly chosen coordinate space, specify the function uniquely. Fourier analysis decomposes a given function into a set of orthogonal functions, using sine and cosine functions as the basis vectors. This is a well-known technique in communication theory.

The Fourier transform is most commonly used to transform a time-varying function into the frequency domain, although its use is not limited to time/radian frequency variable pairs.

$$t \iff \omega \qquad \omega = 2\pi \cdot (\text{frequency}) \qquad (3\text{-}4)$$

For a function $f(x)$, the Fourier transform is;

$$F(u) = \int_{-\infty}^{\infty} f(x)\ e^{-j2\pi ux}\ dx \qquad (3\text{-}5)$$

Any pair of variables can be used to form a Fourier transform pair, provided the product is dimensionless (ie. time <------> frequency = 1/time). The Fourier transform can also be extended to more than one dimension by choosing appropriate variable pairs in each dimension.

The brightness, or intensity, of an image is equal to the radiant power incident on the sensing surface. A typical image formed by an optical system is a function of two spatial variables. Therefore, the image intensity function can be expressed as;

$$\text{Intensity} = f(x,y) \qquad (3\text{-}6)$$

The Fourier transform can be used to transform the image intensity function into the domain of "spatial" frequency;

$$F(u,v) \;=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\; e^{-j2\pi(ux+vy)}\; dx\, dy \qquad (3\text{-}7)$$

or in shorthand notation, $F(u,v) = T\{\, f(x,y)\, \}$ $\qquad (3\text{-}8)$

where, $\quad T\{\ \} \;=\;$ the Fourier transform operator

$F(u,v) =$ the Fourier transform of $f(x,y)$

$u\ \&\ v \;=\;$ the spatial frequency variables.

Since an image covers a two-dimensional area, the Fourier transform variables are in dimensions of length and 1/length. The (1/length) term is called the spatial frequency variable, and is sometimes expressed as cycles/length or lines/length. High spatial frequency means many lines per unit length are visible in an image. Therefore the image "resolution" is related to the spatial frequency content of the image.

Image Resolution

Resolution is a loosely used term when trying to describe visual system parameters. It may be defined in terms of modulation transfer functions, optical line pairs, spot size, television lines, or Rayleigh criterion. Furthermore, each definition is internally consistent. However, correlations among the various definitions must be clearly stated.

Hall (1979) describes the different techniques of defining resolution and relates them to the human visual system. He characterizes the human visual response as a bandpass filter in which both high and low spatial frequency components are attenuated. The most rigorous definition of resolution is in terms of spatial frequency content and the modulation transfer function of the system (Hall 1979).

To this point, the image produced by an optical system has been considered to be a continuous function of two spatial variables. To be of practical use, the image must be captured or recorded electronically or photographically. This captured image is no longer a continuous function, but is now a discrete function having been sampled by some

sensor. This sampling process also has an effect on the resulting image resolution.

The resolution in a photograph is related, in part, to the grain size of the photographic film. In a video camera, it is related to the image sensor design. Solid-state image sensors have individual photosensitive elements which sample the incident illumination (Tseng, et al. 1985).

The conversion of a continuous image function into a discrete representation of that image by sampling is covered in most image processing textbooks. The mathematics involved in sampling and the "sampling theorem" are derived in all digital signal processing textbooks.

After the continuous image has been sampled, it is represented by a discrete image function. The continuous Fourier transform integral is no longer applicable, therefore the discrete Fourier transform must be applied. The relationship between the continuous Fourier transform and the discrete Fourier transform is covered in detail by Bracewell (1986) and Brigham (1988). The discrete Fourier transform of the function $f(x)$ is;

$$F(u) \; = \; \frac{1}{N} \sum_{x=0}^{N-1} f(x) \; e^{-j2\pi ux/N} \qquad (3\text{-}9)$$

For an N x N discrete function;

$$f(x,y): \; x = 0, \, 1, \, 2, \, \ldots, \, N\text{-}1$$
$$y = 0, \, 1, \, 2, \, \ldots, \, N\text{-}1$$

The discrete two-dimensional Fourier transform is;

$$F(u,v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y)\ e^{-j2\pi(ux+vy)/N} \qquad (3\text{-}10)$$

The number of complex multiplications and additions required to calculate an N-point discrete Fourier transform is proportional to N^2. In 1965, a method of computing discrete Fourier transforms suddenly became widely known (Cooley and Tukey 1965). By properly decomposing the discrete Fourier transform equation, the number of multiply-and-add operations can be made proportional to $N\log_2 N$. The decomposition procedure is called the fast Fourier transform (FFT) algorithm. The derivation of this algorithm can be found in most digital signal processing textbooks and is thoroughly covered in the classic book by Brigham (1988).

As previously stated, the formation of an image may be interpreted as the convolution of the actual scene radiance with the transmission properties of the lens. Using Fourier analysis and linear system theory, it can be shown that convolution in one variable domain is the same as multiplication in the Fourier transformed variable domain (Gaskill 1978).

The transmission properties of a lens can be examined in both the spatial domain and the (spatial) frequency domain. In the spatial domain, the impulse response of the lens or optical system is called its point spread function (Hecht and Zajac 1979). For a perfect

(diffraction-limited) lens, the system response to a single point of light (impulse response) is the Airy pattern.

Therefore the image intensity is;

$$i(x,y) = r(x,y) * h(x,y) \qquad (3-11)$$

where; $i(x,y)$ = image intensity

$r(x,y)$ = object radiant intensity

$h(x,y)$ = point spread function of the lens

* = convolution operator.

Because convolution in one domain is equivalent to multiplication in the transformed domain, the transmission properties of a lens can have an alternate description. Using the Fourier transform, the preceding equation may be written in the (spatial) frequency domain as;

$$T\{i(x,y)\} = T\{r(x,y)\} * T\{h(x,y)\} \qquad (3-12)$$

$$I(u,v) = R(u,v) \times H(u,v) \qquad (3-13)$$

The Fourier transform of the point spread function (spatial impulse response) is the frequency response, or the frequency transfer function of the system. Since the frequency response, in effect, transfers the object spectrum into the image spectrum, it is usually called the unnormalized Optical Transfer Function (OTF). The modulus or magnitude of the unnormalized OTF is known as the Modulation Transfer Function (MTF), and is a widely used means of specifying the performance of imaging elements.

Modulation Transfer Function

The MTF is plotted as modulation versus spatial frequency, with various units for the modulation and spatial frequency. Spatial frequency is sometimes given in terms of normalized frequency, cycles/mm, TV lines/mm, optical lines/mm or line pairs/mm, with the latter two being equivalent and most common.

Modulation is defined as the maximum image intensity minus the minimum image intensity divided by the sum of the two.

$$\text{modulation} \quad = \quad \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \qquad (3\text{-}14)$$

The most common plotting method is to normalize the spatial frequency and modulation of the system in question to that of a diffration-limited lens of the same kind. The spatial frequency is normalized to the limiting resolution frequency, and the modulation to the maximum value obtained for a perfect lens.

When the transmission properties of a lens are evaluated, many aberrations can cause the MTF to vary from that of a perfect lens. It is well-known that defocusing a diffraction-limited lens causes aberration. The effect can be shown on an MTF plot (Goodman 1968, Born and Wolf 1983). The MTF can be described in terms of cutoff frequency, bandwidth, and roll-off, as with any other linear filter function. The effect on the MTF due to defocusing can be seen in Figure 4.
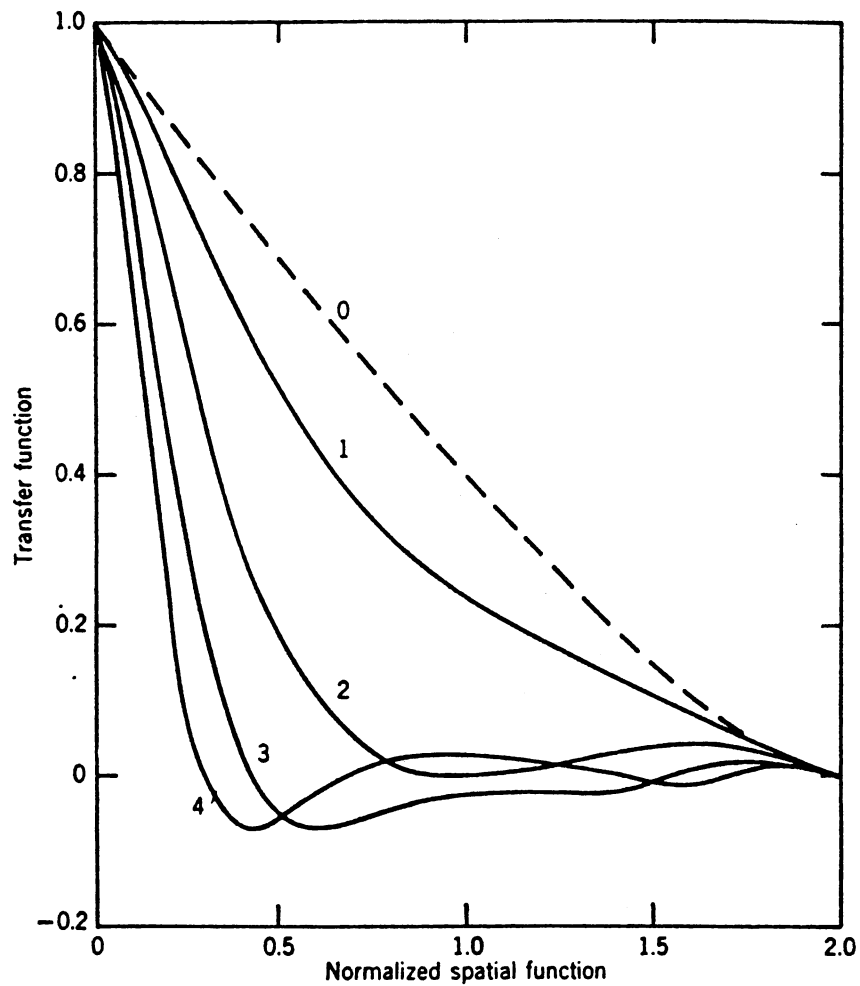
Figure 4. Modulation Transfer Function of a Defocused
Diffraction-limited Lens.

Each curve corresponds to n/π wavelengths of defocus, where n is

the number on the curve. The MTF shows the attenuation of the higher

spatial frequencies and a reduction in the cutoff frequency. This

reduction in frequency content accounts for the lack of resolution or

blurring that occurs in a defocused image.

Walsh Functions

Fourier analysis may be described as the representation of a function by a set of orthogonal sinusoidal waveforms. The coefficients of this representation are called frequency components, and the waveforms are ordered by frequency. In the two-dimensional case, the Fourier transform may be viewed as a special case of a sequence of matrix multiplications of the given function to be transformed by a general matrix multiplier kernel (Pratt, et. al. 1969).

There are many other matrix multipliers that could be used to transform the given function. Instead of using sinusoidal waveforms, a given function could be represented using square waves with values of only +1 and -1. Walsh functions are a complete set of orthogonal square wave functions which can be used to represent an arbitrary function. Spectral analysis may be performed with Walsh functions analogous to the use of the Fourier transform (Hall 1979). Walsh functions can be ordered by the number of zero crossings in the open interval (0,1). This property has been called "sequency" (Harmuth 1977). The coefficients of this representation are called the sequency components, with the same interpretation as given to frequency.

Computation of the Walsh transform is much simpler than the Fourier transform, because Walsh functions are real, rather than complex. Furthermore, they take on only the values of +1 and -1. A fast algorithm identical in form to the successive-doubling method used to calculate the FFT can be used to compute a fast Walsh transform (FWT). The only

difference between the two transforms is that all the exponential terms in the FFT algorithm are set equal to 1 (Gonzalez & Wintz 1977).

A discrete set of Walsh functions may be developed by sampling the continuous functions at equally spaced points in the interval [0,1]. The number of samples should be a power of 2, to preserve the even-and-odd function pairing and to permit a sample in each function interval. If Walsh functions with the number of zero crossings less than, or equal to, 2^n-1 are sampled at $N = 2^n$ uniformly spaced points, a square matrix is produced with elements of values +1 and -1. These matrices are orthogonal, and the rows are ordered with increasing number of zero crossings (Ahmed, et. al. 1971, Hall 1979).

Hadamard Matrices

A Hadamard matrix is an orthogonal matrix with elements of value +1 and -1, only. Although the values of the elements in the Walsh matrix and Hadamard matrix are identical, the order of rows and columns is different. In fact, when both NxN matrices are of size N = 2^n, the ordering of the rows is the only difference between the Walsh and Hadamard transforms.

When N is not equal to an integer power of 2, this difference is more important. While the Walsh transform can be formulated for any positive integer value of N, existence of the Hadamard transform for values of N other than integer powers of 2 has been shown only up to N = 200 (Pratt, et. al. 1969). Since most applications of transforms in image processing are based on N = 2^n samples per row or column of an image,

the use (and terminology) of Walsh and Hadamard transforms is
intermixed in the image processing literature. The term Walsh-Hadamard
transform is commonly used to denote either or both transforms
(Gonzalez & Wintz 1977).

CHAPTER IV

HARDWARE

Introduction

Investigation into obtaining range data from a video camera using lens focus required the combination of various components of equipment and extensive software development. This chapter describes the hardware components used and their integration into a complete system. The main hardware components are an image processing computer, video camera, motorized lens, and lights.

Image Processing Computer

The image processing computer used for this project was an International Robomation/Intelligence (IRI) D256 machine vision system. Resolution is 256 horizontal pixels by 240 vertical pixels, with 8 bits of gray level per pixel. The system provides four frame buffers for image processing tasks or for multiple camera input. The IRI-D256 incorporates a hardware coprocessor which performs computationally intensive operations such as arithmetic functions, histograms, convolutions, run-length encoding, and moments calculations.

The computer uses a real-time Unix-type operating system and includes image processing functions which may be called from programs written in FORTH or C. The resident Iconic Kernal System (IKS) is the

library of available image processing system functions. An IKS interpreter is provided for interactive processing and for fast prototyping of IKS function calls before the actual FORTH or C code is written.

The IRI-D256 uses an 8 Mhz Motorola 68010 microprocessor as CPU and has 512K bytes of RAM. A 40-megabyte hard disk is included, along with a 5 1/4 inch floppy disk drive.

Video Camera and Lens

An EG&G Reticon Model MC9256 camera was used for image acquisition. The EG&G Reticon MC9256 is a solid-state camera using a photodiode array image sensor. Each photodiode in the matrix can be clocked out serially and processed individually. The camera is capable of high-speed operation with image rates up to 105 frames per second. The two-dimensional, self-scanned optical sensor array is composed of 65,535 discrete photodiodes arranged in a square 256 x 256 matrix with each photodiode spaced 40×10^{-6} meters apart.

Unfortunately, the IRI-D256 image processing computer requires an RS-170 format video input. The IRI-D256 also must supply the horizontal and vertical video synchronization (sync) signals to the camera, externally. These requirements made the EG&G Reticon MC9256 unuseable without interface circuitry.

An EG&G Reticon Model MB9000 Video Data Formatter was used to format the camera output into a RS-170 compatible video signal. This video data formatter converts the analog video data from the camera into various output data configurations. The formatter may be equipped with 10 different user options to accommodate specific applications.

Power Supply and Sync Driver

The EG&G Reticon Video Formatter required electrical power with +15 volts dc at 0.3 amps, +5 volts dc at 2.0 amps, and -15 volts dc at 0.3 amps. Two modules were installed in the EG&G Reticon MB9000 Video Data Formatter to give both the horizontal and vertical scan timing for RS-170 compatible video output.

The IRI-D256 computer supplies the horizontal and vertical sync signals in single-ended form, whereas the EG&G Reticon MB9000 Formatter requires the signals in RS-422 differential line-driver form. RS-422 balanced lines allow long cables to be driven reliably in noisy electromagnetic environments. An interface circuit was designed and built to convert the IRI-D256 single-ended sync signals to RS-422 differential line-driver form. An uA9638 line-driver chip was used and installed in the MC9256 and MB9000 external power supply. High-quality shielded cables were used to connect the IRI-D256 sync signals to the circuit and the video formatter.

During the design of the interface circuit, it was discovered that the IRI-D256 computer did not supply the sync signals in standard RS-170 format, or even at TTL signal levels. This was in part due to the design of the video circuitry in the IRI-D256 computer, and possibly due to the requirements of the Hitachi KP-120 CCD video camera normally used with the system. The Hitachi camera did not output a standard RS-170 video signal, but the computer still functioned properly. Because of the large variations allowed by the uA9638 line-driver chip in the input signal level,

the video data formatter operated properly despite the non-standard sync signals supplied by the IRI-D256.

To implement an autofocusing technique, some method was required to change the focus and zoom of the video camera lens. Commercially available 35mm and video camera lenses were investigated, including motorized focus and zoom lenses.

Most motorized focus and zoom lenses for video cameras used servo motors with open-loop control to alter focus and zoom settings. Their minimum cost was over $700. Motorized autofocus lenses for 35mm cameras were less expensive, but had many additional features that were unnecessary and required complicated mechanical interfacing and control.

The camera lens design broke down into the following options;

a) open-loop control vs. closed-loop control

b) servo motor vs. stepper motor

c) zoom lens vs. fixed focal length lens

d) IRI-D256 computer control vs. separate

microprocessor control.

Since a limiting assumption was the use of a fixed focal-length lens, and a simple low-cost system was desired, the design choices became fairly obvious.

Stepper Motor and Driver Board

A fixed focal length lens was used with a stepper motor driven by the IRI-D256 computer using open-loop control. A Vivitar 135mm f2.8 lens was used with a C-mount adaptor. A small 6-volt stepper motor was attached to drive the lens using miniature sprockets and cable chain. Sprocket sizes yielded a 4:1 drive reduction ratio. The stepper motor was driven by an AMSI Corp translator-driver board which included its own power supply. The translator-driver board was connected to the IRI-D256 parallel printer port. Step and direction signals were required from the IRI-D256 computer.

The IRI-D256 computer contains a parallel printer port which can be configured for parallel I/O. Software was supplied by IRI to initialize the port for parallel read-and-write operations. Modifications to this code were used to control the stepper motor. A simple aluminum bracket was fabricated to mount the stepper motor and the camera lens. A block diagram of the system hardware is shown in Figure 5.
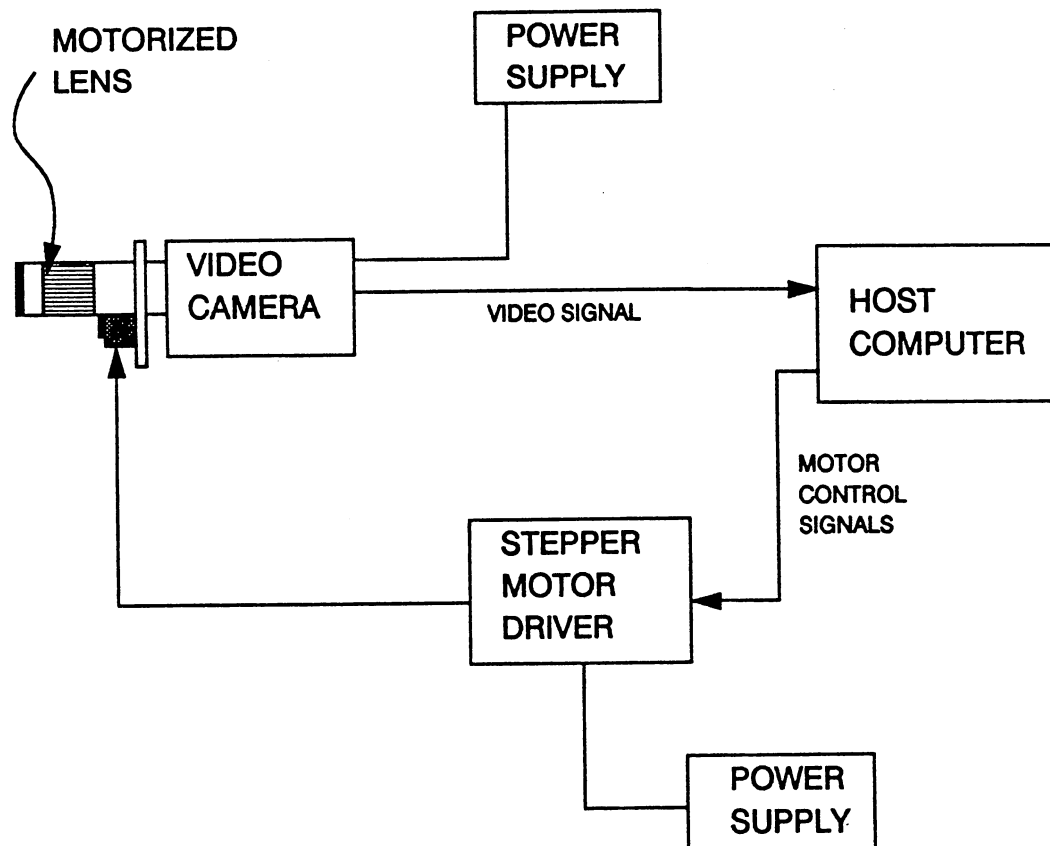
Figure 5. Block Diagram of System Hardware

CHAPTER V

SOFTWARE

Introduction

This chapter describes the software developed to control the equipment and to perform the required tasks. The system software is divided into three modules; MAIN, FOCUS QUALITY, and MOTOR MOVE. The MAIN module searches for optimum focus position by calling the FOCUS QUALITY function and the MOTOR MOVE function. It also calculates the object distance from the optimum focus position and stores the data to disk. The FOCUS QUALITY function and MOTOR MOVE functions are combined with a header file using the Unix Makefile utility to obtain the executable program. The three modules will be described in greater detail in the following sections.

Focusing Algorithms

The focus quality module uses either the Fourier or the Walsh-Hadamard transforms to obtain a measure of focus quality. The function returns a single number which increases in value as focus quality increases. Whenever the lens is moved to a new location during the search routine, the FOCUS function is called.

Image Window

When the FOCUS function is called by the MAIN module during the first iteration, it displays a live image on the video monitor and prompts the user to snap the image. The user is then asked for a desired window size in a range from 8 x 8 to 128 x 128 pixels. This windowing is required, because the IRI-D256 does not have sufficient RAM to perform a Fourier transform without writing intermediate data values to the hard disk. The IRI-D256 contains 512K bytes of RAM, of which the Regulus operating system takes 151K bytes, leaving 361K bytes of memory available for executable code, global data, stack, and heap. Executable code for this project is almost 100K bytes in size, leaving about 269K bytes available for global data and the stack and heap.

Three of the four 64K byte frame buffers are used to hold the original image and the real and imaginary parts of the Fourier transform. The magnitude is then displayed using frame buffer #3.

After selecting a window size, the user is asked to position the window within the 256 x 256 pixel image. The user is asked for the desired x and y coordinates of the upper left-hand corner of the window. This position is referenced to the upper left-hand corner of the image at x=0, y=0 and the lower right-hand corner at x=256, y=256.

The window positioning function contains error-checking to prevent any part of the window from extending beyond the image edge. For example, a 64 x 64 sized window could not be located at image position x=200, y=200.

If the location of the upper left-hand corner of the window is within the image boundary, the software asks the user if the location is acceptable, or if another location is desired. The outline of the window is drawn on the video monitor using frame buffer graphics functions. This procedure allows the user to verify that the window contains the region of interest. The window size and positioning are performed only during the first call of the focus function by the MAIN module. During all subsequent calls, the image is snapped and the same window location is maintained. This operation is performed automatically by declaring the upper left-hand corner coordinates of the window as global variables.

Memory Requirement

After the image is snapped and stored in frame buffer #1 and the window size and location selected, the FOCUS function must allocate enough RAM for the window data. This RAM is allocated using the C function, calloc(). An array of double-precision, floating-point pointers is used, with one pointer for each row in the window. Each pointer is allocated adequate memory for the row of pixels in the window.

Because the IRI-D256 digitizes the video input to 256 gray levels, one byte (8 bits) of memory is required for each pixel value. Calculation of the Fourier transform requires the use of floating point functions and numbers, which the Regulus operating system automatically changes to double-precision. The pixel values in the frame buffer are stored as an unsigned character variable (8 bits wide), whereas the FFT results are double-precision float variables (32 bits wide). This arrangement required an index

offset scheme and variable type-casting when reading pixel values from the frame buffer into RAM, and from RAM back into the frame buffer for storage. Because of these changes in the number of bits required to represent data, more memory was required than apparent. A 64 x 64 pixel window required 4K bytes of frame buffer memory, but 32K bytes of RAM were required to hold the intermediate FFT values.

64 pixels x 64 pixels = 4096 pixels = 4K pixels in window
4K pixels x 1 byte/pixel = 4K bytes integer data

4K bytes x 4bytes/double precision = 16K bytes data in RAM
16K bytes data x 2 (for complex FFT result) = 32K bytes required.

The Walsh transform required only 8K bytes of memory, since it uses only real numbers and can be executed with long integer (2 bytes/long integer) arithmetic.

Data Taper Function

After memory has been allocated for the window data and the values read into RAM, the image contents within the window are displayed on the video monitor, with the rest of the image turned black. When caculating Fourier transform coefficients, image data within the window area are multiplied by a Tukey-Hanning taper function which smooths the pixels values to zero at the edges of the window. Taper functions such as the Tukey-Hanning are called "window functions" in the digital signal processing literature.

If the sampling interval is fine enough to cope with the highest frequencies present in the data, there is no aliasing error in the Fourier transform. Other errors exist, due to the finite length of the data. Truncation of a function to a finite length introduces smoothing error (a reduction of fine detail in the Fourier transform) and leakage error.

Leakage error tends to falsify the higher frequencies in the spectrum, whereas smoothing error is distributed differently. Leakage error may be reduced at the expense of increased smoothing error, by use of a tapered truncation (window) function. A thorough discussion of taper functions may be found in Brigham (1988). After tapering the image data, the software calculates the FFT.

Transform Calculation

The two-dimensional discrete Fourier transform and Walsh-Hadamard transform equations can be expressed in separable form. The separability property allows the calculation of the two-dimensional transform by successive applications of the one-dimensional transform. The two-dimensional transform is calculated by first taking the one-dimensional transform along the rows of the data matrix, then taking the one-dimensional transform down the columns of the resulting matrix obtained by transforming the rows.

An image is composed of real data values (imaginary component is zero). After taking the first one-dimensional Fourier transform along the rows, the result contains both real and imaginary components. Taking the one-dimensional transform down the columns of the resulting matrix now

requires complex additions and multiplications. The final result is a matrix of real and imaginary components. The software implements several additional functions to put the resulting spectral data in a more suitable form for display.

It is difficult to interpret results of the Fourier transform if the spectral components are retained as complex numbers. The most common technique is to calculate the magnitude of the spectrum. To display the spectrum on the video monitor, some form of dynamic range compression is required. Normalizing the Fourier coefficient values to the interval [0,1] or expressing them in units of decibels are two methods of reduction. The new values are then scaled from 0 to 255 for display as gray levels on the monitor. Both methods were used to visualize the transformed image.

The one-dimensional discrete Fourier transform has been defined for N data values in the interval [0, N-1]. In the frequency domain this formulation yields two half-periods of N/2 points back-to-back in the interval [0, N-1]. The spectrum is symmetrical about the point N/2, with points greater than N/2 being negative frequencies. To display the spectrum in the conventional manner, it is necessary to move the origin of the transform to the point N/2. This translation is accomplished by multiplying the original data $f(x)$ by $(-1)^x$ prior to taking the transform. The Walsh-Hadamard transform does not possess such translation properties, therefore must be evaluated differently.

The same situation holds true for the two-dimensional Fourier transform, except that results are considerably more difficult to interpret if the origin of the spectrum is not shifted to the point (N/2, N/2). The software centers the spectrum by multiplying every image point $f(x,y)$ by

$(-1)^{x+y}$ prior to taking the transform. The multiplication by $(-1)^{x+y}$ takes place after the Tukey-Hanning taper function is executed. Brigham (1988) discusses rearranging or centering the two-dimensional spectrum for conventional viewing. One additional adjustment must be made to the spectrum for the display to resemble an optically generated Fourier transform (diffraction pattern). The magnitude matrix generated by the FFT has symmetry when viewed by quadrants. Quadrant 3 is a positive reflection of quadrant 1, and quadrant 4 is a positive reflection of quadrant 2. Therefore only half of the values are unique.

The spectrum appears slightly skewed when the origin of the spectrum is centered at point (N/2, N/2). Point (N/2, N/2) is the zero frequency point and is actually in quadrant 3 of the rearranged spectrum. To restore complete symmetry for display, column 0 must be repeated as column N and row 0 must be repeated as row N. This procedure is equivalent to repeating the Nyquist spatial frequency sample values in each quadrant.

As discussed in Chapter III, Walsh functions may be generated by various methods. Each method implies a specific ordering of the functions. Walsh functions ordered in terms of increasing sequency, is only one possible form of ordering. Another possible ordering is known as "dyadic" or Paley ordering. Dyadic or Paley ordering results from a specific method of generating the Walsh functions. This ordering is important because the algorithm used to calculate the Fast Walsh transform (based on the Cooley-Tukey FFT algorithm) returns the Walsh coefficients in dyadic order. Although this result is of little consequence for applications such as image encoding for bandwidth compression, it is inconvenient for spectral analysis

work where sequency ordering of Walsh coefficients is usually desired. Dyadic or Paley ordered Walsh functions are sometimes called regular or normal ordered Walsh functions.

Fortunately, there is a simple correspondence between normal (dyadic) and sequency-ordered Walsh functions. Given a sequency-ordered Walsh function $W(n)$, one can express the decimal index number "n" in Gray code, then interpret the Gray code binary number as the binary equivalent of a decimal number. Defining this new decimal index number as "m", the normal-ordered Walsh function $N(m)$ is equivalent to the sequency-ordered Walsh function $W(n)$.

The fast Walsh-Hadamard transform returns the transform coefficients in normal order. Changing the coefficients to sequency-ordering is accomplished easily using the exclusive-or and barrel-shift bitwise operators included in the programming language C. The reordering software code adds very little to the overall calculation time, since bitwise operators execute at high speed.

The Walsh-Hadamard transform of an image does not possess a convenient physical analogy as does the Fourier transform. It is unknown if the same elegant mathematical techniques of Fourier optics can be used to describe image focus quality. Specifically, do high-sequency Walsh coefficients yield sharp, high contrast images? The effect of coefficient ordering with regard to image quality is also unknown. To answer these questions, both normal-ordered and sequency-ordered transform coefficients are calculated and evaluated by the software.

Focus Quality Value

To obtain a measure of focus image quality, the Fourier transform coefficient (spectrum) values and Walsh-Hadamard transform coefficient values must be analyzed. Sharp focus conditions yield higher frequency components in the spectrum. This assumption will be the assumption used with the Walsh transform coefficients and the "sequency spectrum". Therefore, the desired result is to maximize the high-frequency content of the spectrum or the high-sequency coefficients of the Walsh-Hadamard transform.

The software evaluates the frequency content of the spectrum by performing a calculation analogous to the moment of inertia in mechanics. Each pixel in the spectrum has a coefficient value that is analogous to the mass value of an incremental mass element. In mechanics, the mass element is multiplied by the square of the lever arm length. The distance from the mass element to the axis of rotation is the lever arm. The moment of inertia equation for discrete mass elements is,

$$\Phi = \sum_{i=n/2}^{N} \sum_{j=1}^{N} m_{ij}\, r_{ij}^2 \qquad (5\text{-}1)$$

where, Φ = moment of inertia

m_{ij} = mass element (or pixel value) in the ith row and j^{th} column

r_{ij} = radial distance to the ith row and j^{th} column element.

Because the spectrum of an image is symmetrical, only half of the Fourier coefficients need be included in any calculations. The zero frequency point (N/2, N/2) may be viewed as the axis of rotation. The radial distance from the zero frequency point to each pixel is used as the lever arm. See Figure 6.

FOURIER COEFFICIENTS

16 * 16
WINDOW

MOMENT
CALCULATION
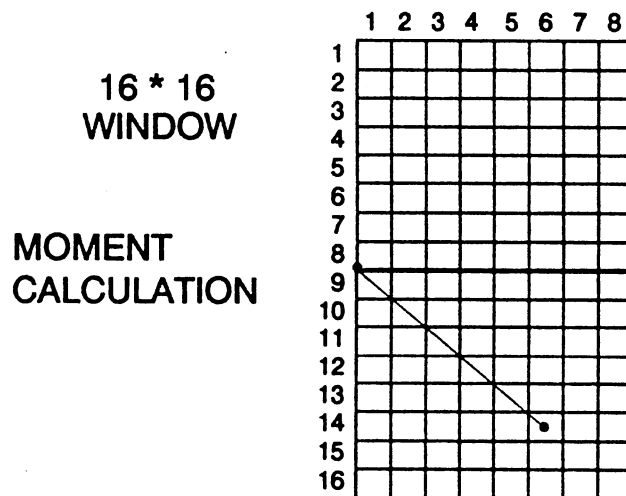
Figure 6. Moment of Inertia Calculation using Fourier Cofficients

Higher frequency components are analogous to mass distributions away from the axis of rotation. Therefore, as the moment of inertia gives a measure of the mass distribution about an axis of rotation, the focus quality can be judged by the spectral distribution about the zero frequency point. A single numerical value is then obtained as an indication of focus quality. The larger the numerical value, the sharper the focus should appear. In

order to compare focus quality values for differing data window sizes, a normalization routine is contained in the software.

As a geometric object increases in scale, the moment of inertia increases by a factor to the fourth power. Given two objects, one twice as large as the other, the moment of inertia for the larger object is 16 times that of the smaller object. Therefore, various NxN size windows can be compared, provided the focus quality value is normalized by N^4.

For an N x N pixel image, the Walsh-Hadamard transform yields an N x N matrix of real numbers. When placed in sequency order, coefficient (0,0) is the zero sequency (dc or no zero-crossing) component and coefficient (N-1, N-1), the highest sequency component. To calculate a focus quality value using Walsh coefficients, a moment-of-inertia type calculation is performed similar to that used with the Fourier coefficients. In this case, all N x N Walsh coefficients must be included, since the Walsh-Hadamard transform does not possess the symmetry of the Fourier spectrum. The Walsh coefficient value is multiplied by the radial distance squared, with the radial distance from the coefficient (0,0) to each pixel component used as the lever arm. As before, the coefficient (0,0) may be viewed as the axis of rotation. See Figure 7.

$$\Phi = \sum_{i=1}^{N} \sum_{j=1}^{N} m_{ij} \, r_{ij}^2 \qquad (5\text{-}2)$$

where, Φ = moment of inertia

m_{ij} = mass element (or pixel value) in the ith row and j^{th} column

r_{ij} = radial distance to the ith row and j^{th} column element.

WALSH COEFFICIENTS

16 * 16
WINDOW

MOMENT
CALCULATION
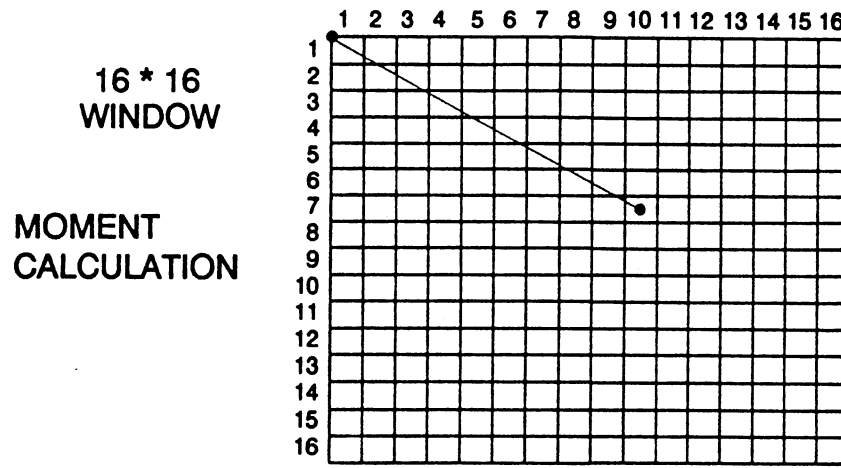
Figure 7. Moment of Inertia Calculation using Walsh Coefficients

The final task performed by the FOCUS module is to free all the memory allocated by the calloc() function. Each time the FOCUS function is called, memory is allocated for two arrays of pointers. Failure to free the memory at the end of each call would cause the system to seek a new block of memory each time the module was used. Eventually, all the memory in the system would be filled. Freeing the memory also leaves a block of memory available for the next time the FOCUS function is called. Program execution speed increases, because the system does not search for unused memory during subsequent function calls. After the free memory routine, the software returns the focus quality value and program control to the MAIN software module. A flow chart outlining the FOCUS QUALITY function is shown in Figure 8.

Figure 8.   Flow Chart for FOCUS QUALITY Function

Search Routine

The SEARCH routine forms a large part of the MAIN module. This algorithm locates the stepper motor position that gives optimum focus quality. The resulting stepper motor position is used in the prediction equation to obtain object distance from the camera.

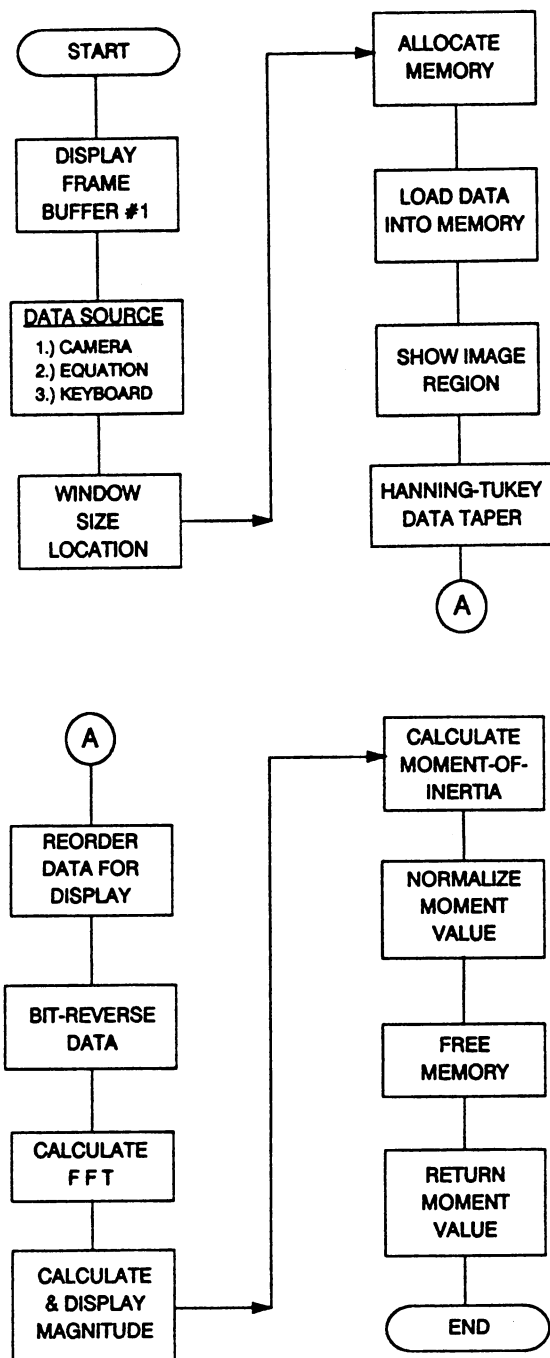Focus quality is a function of stepper motor (or focus) position for a given object distance, f-stop setting, scene lighting, etc. This quality function reaches a maximum at the point of optimum focus. Therefore, the search routine must employ a numerical method of finding the maximum value of a function. Common methods of finding the maximum or minimum of a function require taking the first and second derivatives of the function. Focus quality is not an analytic function, therefore derivatives cannot be used.

Many algorithms are available to find the maximum or minimum of a function without using derivatives. Cheney and Kincaid (1980) describe a Fibonacci search algorithm and the Golden Section search method, along with advantages and disadvantages. The Golden Section search algorithm, parabolic interpolation, and Brent's method are discussed by Press, et al. (1986).

Each method has advantages and disadvantages with respect to ease of programming, number of function evaluations required, rate of convergence, and apriori knowledge needed. The universal requirement is that the function in question must be continuous, unimodal, and monotonic on either side of the maximum value. The Golden Section search method was chosen for this project because of its ease of programming and

integration into existing software, and because no apriori knowledge about the function is required.

The stepper motor required 1200 steps to move the camera-lens focus ring from one endpoint to the other. Stepper motor position #0 corresponded with the lens completely withdrawn (focused at infinity). The lens was completely extended, or focused on the minimum focusing distance (approx. 4 feet), at motor position #1200. The Golden Section search required two endpoints of the independent variable (stepper motor positions), and an arbitrary point (motor position) somewhere between the endpoints. A tolerance value was also required to limit the search. This tolerance value specified an interval of the independent variable (motor position), which caused the routine to end when the search distances became less than the interval. This technique allowed control over how closely the search routine could approach the true maximum value.

Knowing the two endpoints (motor positions) and an arbitrary point X in between defines two intervals of stepper motor position; #0 to X, and X to #1200 These points permit the search routine to calculate two new positions, x1 and x2, within the two intervals. (see Figure 9)

```
●————————————●————————————————●————————————●————————————————●
0            X1               X             X2               1200
```
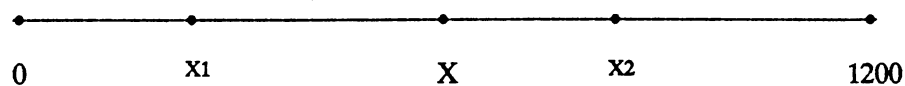
Figure 9. Stepper Motor Search Intervals

These intervals are obtained by using the Golden Ratio employed by Egyptian pyramid builders and ancient Greek architects. The Golden Ratio satisfies the equation;

$$R^2 = 1 - R \qquad \text{or} \qquad R = \frac{1}{2} (\sqrt{5} - 1) \approx 0.618 \qquad (5\text{-}3)$$

The Golden Ratio search subdivides the intervals 0 to X and X to 1200, into sections approximately 1/3 and 2/3 original size. A detailed derivation of this algorithm is given by Press, et al. (1986).

The search algorithm passes the value of x1 to the MOTOR MOVE function, which moves the stepper motor (and lens) to position x1. The FOCUS quality function is then called and its value returned to the MAIN module and stored in an array. This first focus evaluation is considered LOOP #1. The same procedure is repeated for position x2, and this is considered LOOP #2. The motor position number is also stored in an array, with the array subscript for both variables being the loop number.

The two endpoints, the arbitrary point X, and the two calculated points x1 and x2 define four intervals of stepper motor position. Based on the two focus function evaluations at positions x1 and x2, the algorithm determines in which of the four intervals to perform the next focus quality evaluation. The next evaluation is called LOOP #3. This procedure continues until the algorithm converges on a given interval which contains the maximum focus value. The tolerance specification gives a minimum convergence interval of 2 steps of the stepper motor. This interval may be larger, depending on the direction from which it is approached (ie., the 0.382 segment or the 0.681 segment).

The final results are arrays of focus quality values and stepper motor positions for each evaluation or LOOP #. The maximum focus quality value is found, along with the corresponding stepper motor position. The stepper motor position is then entered into the prediction equation to obtain the object distance. The MAIN module then asks the user if the arrays for stepper motor position and focus quality values are to be saved to the hard disk. If affirmative, the software requests a filename and checks for correct entry. The final task is to write the data to the hard disk, then exit to the operating system.

Stepper Motor Control

The motorized video camera lens uses a stepper motor to adjust lens focus position. The stepper motor is actuated by a translator-driver board connected to the IRI-D256 computer parallel printer port. The MOTOR MOVE function controls the direction of motor rotation and number of steps executed. This module also initializes and writes data to the parallel I/O (printer) port. Software to move the stepper motor is the smallest of the three modules. The functions that configure the parallel printer port for parallel I/O and read-and-write data to the port were supplied by IRI.

The stepper motor position value is passed to the MOTOR MOVE module by the search algorithm in the MAIN module. This is an absolute motor position between 0 and 1200. The MOTOR MOVE function compares this number with the previous motor position value. It then determines the relative number of steps to rotate the motor and the direction of rotation. The first time the function is called, it assumes that the previous motor

position is 0. This constraint means that the camera lens must be completely withdrawn (at motor position #0) each time the program is started.

The motor control requires the use of only two data lines of the parallel port. Data line D0 is toggled between TTL logic levels 0 and 1 to step the motor, while data line D1 is taken high to move the lens out (clockwise) or low to move the lens in (counter-clockwise). A printf statement is used to provide a delay to avoid stepping the motor too quickly. After moving the proper number of steps, the MOTOR MOVE function returns to the main program. The flow chart in Figure 10 displays the overall design of the system software.
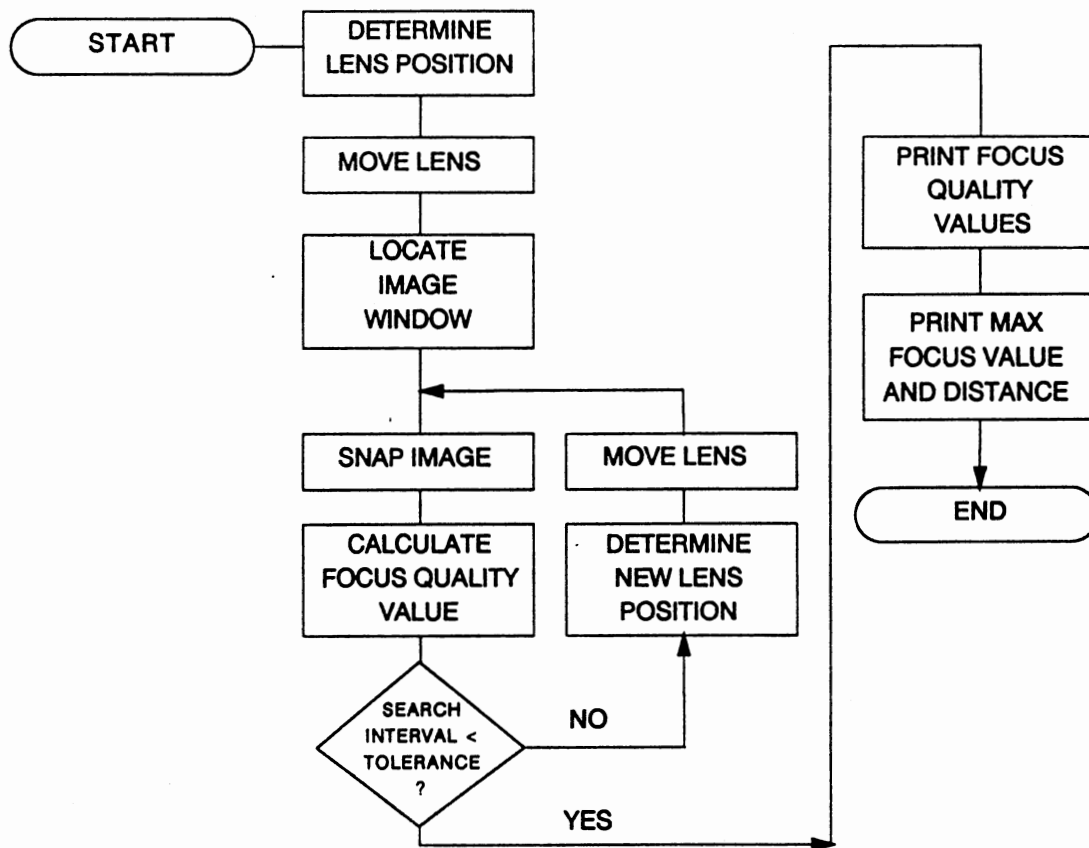
Figure 10.   Overall System Software Design

CHAPTER VI

TESTING AND VERIFICATION

Focus Module

Each function and module was individually tested during software development. The FFT and Walsh-Hadamard algorithms were tested with one and two-dimensional sample data from the literature to verify results. Several versions of the software were configured, depending on the function being tested. One version allowed two-dimensional data input from a data equation, the camera, or individual values entered from the keyboard.

An exponentially damped sine-wave equation was taken from a digital signal processing textbook to test the one-dimensional FFT algorithm. The calculated real and imaginary Fourier coefficients agreed with the answers in the textbook to six decimal places. In the book by Hall (1979), an example contains the Fourier and Walsh transforms of an 8 x 8 data matrix. Real and imaginary coefficients and magnitude of the Fourier transform along with the Walsh coefficients are presented in the text. The two-dimensional Walsh and Fourier transform software was tested with the same 8 x 8 data matrix entered from the keyboard. Calculated transform coefficients agreed with the textbook results to five decimal places. The textbook contained examples of sequency-ordered and normal-ordered Walsh coefficients.

To test the focus quality function, the camera was focused on a standard USAF-1951 resolution target. Image focus was manually adjusted to the best setting as viewed by a human observer. Repeated calculations of the focus quality function were performed and evaluated. The values calculated had a range of variation of less than +/- 2%. Readings were taken at several distances from the resolution target to determine any correlation between variation in focus quality value and range. Variation was essentially the same at all distances.

In order to observe fixed pattern and dark current noise of the image sensor, images were snapped with the camera-lens aperture closed and lens cap in place to eliminate all incident light. The IRI-D256 histogram coprocessor was used to display the pixel variations. A small bias noise level was observed with zero light, however, it did not affect camera operation under normal light conditions.

The IRI-D256 computer has the capability of snapping successive images with the video camera. Instead of snapping a single image frame for processing, the system can snap multiple frames and average them into a single image. Averaging multiple frames of the same image reduces any random noise introduced by the image sensor, video circuitry, A/D conversion, or other system electronics.

The software was modified to snap ten frames and average them into an image. The camera was again focused on the resolution target and the focus quality value calculated. The variation in values was reduced to less than +/- 1.5%. This demonstrated that approximately +/-0.5% of the variation was due to random noise during image acquisition. The remaining variation was due in part to fixed pattern noise of the image sensor.

Snapping ten frames instead of one required more than ten seconds to obtain the averaged image. The decrease in variation of focus quality value did not justify the increase in execution time.

Theoretically, sharper image focus should yield larger numerical values of the focus quality function. For verification, the lens focus setting was varied and focus quality value calculated for each position. In all cases, sharper focus produced larger values. The actual numerical value of the focus quality function depends on the number of edges present in the image. More edges contribute more Fourier coefficients to the spectrum.

Stepper Motor Control

Implementation of the lens-stepper motor system required verification of hardware and software operation. Each component was evaluated individually and then tested again when connected to the lens-stepper motor system.

The stepper motor and translator-driver board were first tested for proper operation using a square-wave generator and an oscilloscope. The translator-driver board can sequence in half-step or full-step mode. Half-step mode doubles the number of motor steps per revolution, at the expense of reducing motor torque. The motor did not produce sufficient torque in half-step mode to turn the lens focus ring. Torque required to rotate the lens focus ring varied as focus ring position varied.

Tests performed using a signal generator showed 40 steps per second to be the maximum stepping rate in full-step mode. Faster stepping rates caused the motor to stall and skip steps halfway through the focus ring travel. A printf() statement was used as a delay in the control software to

slow the step rate below 40 step per second. The actual step rate used was about 25 steps per second.

Proper voltage levels on the parallel port data lines were confirmed before connecting to the driver-translator board. Finally, the entire stepper motor software module was tested before adding it to the main software module.

Search Routine

The Golden Section search algorithm was developed as a separate module and later added to the MAIN module. Several quadratic equations were used to verify the ability of the algorithm to find the function maximum. With proper operation confirmed, the search routine was incorporated into the MAIN software module and tested with the FOCUS function.

The looping structure of the MAIN module was written and tested before adding any functions required for system operation. Since the software is user interactive, many sections of code were required to prompt the user for input. A set of reusable, generic input functions was developed to interact with the user. The user reply is checked for improper input (ie. entering a zz for a yes or no question). If input is of proper form, it is echoed back to the monitor screen and the user is given one last opportunity to change the response. The same input procedure was used for the code which saves test data to the hard disk.

CHAPTER VII

COMPARISON OF FOCUS QUALITY FUNCTIONS

The performance of the focus quality functions will be investigated in this chapter. The Fourier transform, normal or dyadic-ordered Walsh-Hadamard transform, and sequency-ordered Walsh-Hadamard transform will be examined and compared. Focus quality value versus focus position will be plotted and evaluated.

There are several important requirements for a focus quality function. First, it should reach a maximum at the optimum focus condition (this might seem obvious). It should also be unimodal and monotonic. This requirement would allow the maximum value to be found. It was shown in Chapter III that the Fourier transform contains information on the sharpness or contrast in an image. Theoretically, the Fourier transform could be used to determine optimum focus of a camera-lens system. The Walsh-Hadamard transform (in its various forms) has never been used as a focus quality function; and therefore, its performance was previously unknown.

The premise that focus quality value increases numerically as focus quality improves was tested for each algorithm. This was accomplished by incrementing the camera lens to a fixed interval and calculating the focus quality value. The image on the monitor was observed and the optimum

focus condition noted. Each algorithm appeared to yield a maximum value at the optimum focus setting.

To further test the focus quality algorithms, the Golden Ratio search routine was added to the software. The focusing software was executed to determine the maximum focus quality value. Focus quality value data for each stepper motor position were recorded. The maximum focus quality value always occurred at the optimum image focus condition. Several tests were performed to compare the behavior of the Fourier transform and both Walsh-Hadamard transform versions. Plots of focus quality value versus stepper motor position were made at several different object distances. The focus function values were normalized to facilitate comparison of the different algorithms. Figure 11 shows the behavior of the normal-ordered Walsh-Hadamard and Fourier transforms. The normal-ordered Walsh-Hadamard transform is compared to the sequency-ordered Walsh-Hadamard transform in Figure 12. A plot of all three focusing algorithms is given in Figure 13.

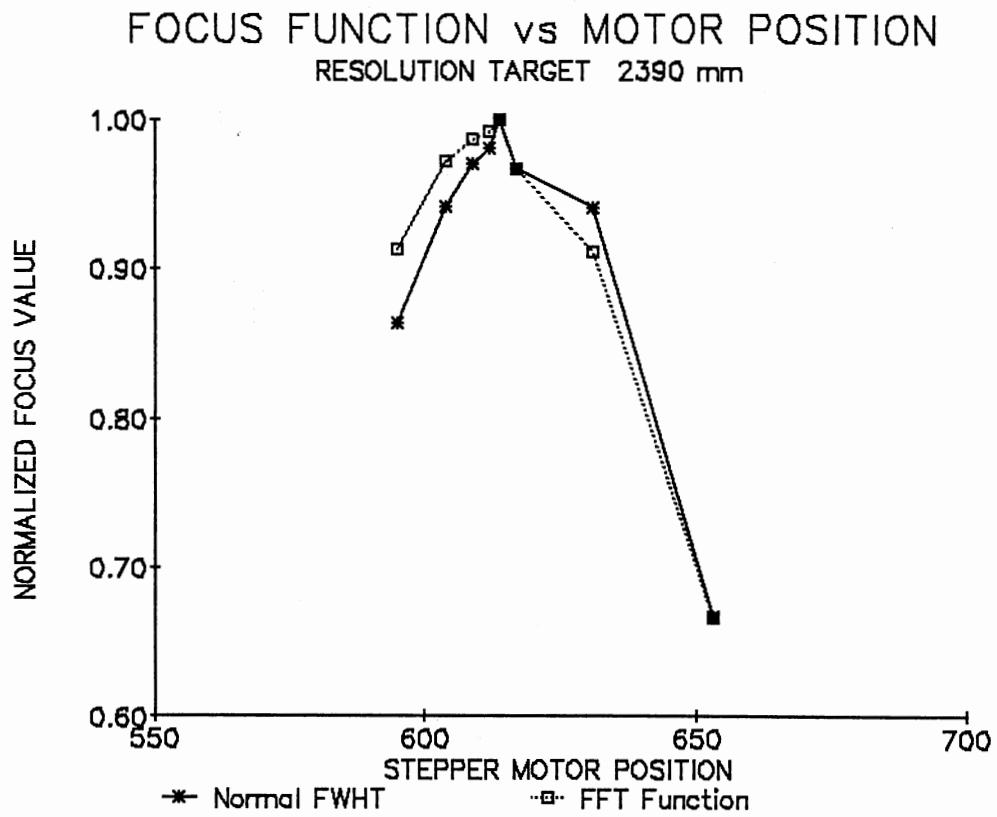FOCUS FUNCTION vs MOTOR POSITION
RESOLUTION TARGET 2390 mm

Figure 11.   Normalized Focus Quality Value Versus Lens Focus Position for
the Fourier and Normal-ordered Walsh-Hadamard
Transforms.
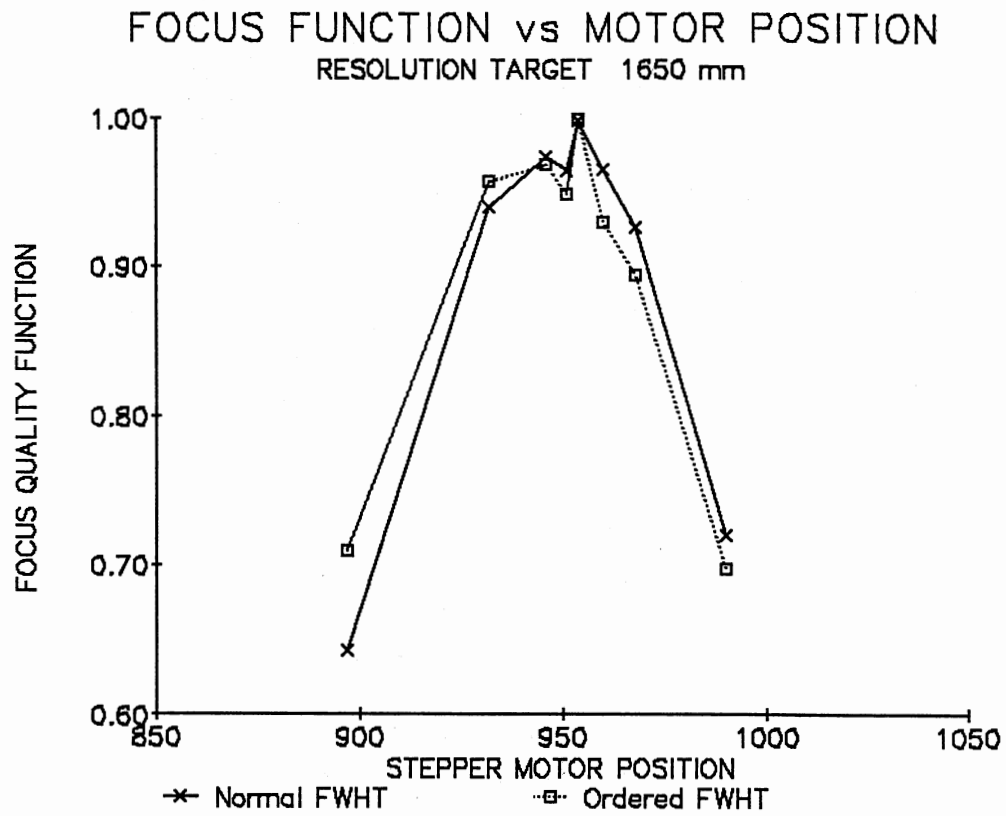
Figure 12.   Normalized Focus Quality Value Versus Lens Focus Position for the Normal-ordered Walsh-Hadamard and Sequency-ordered Walsh-Hadamard Transforms.

FOCUS FUNCTION vs MOTOR POSITION
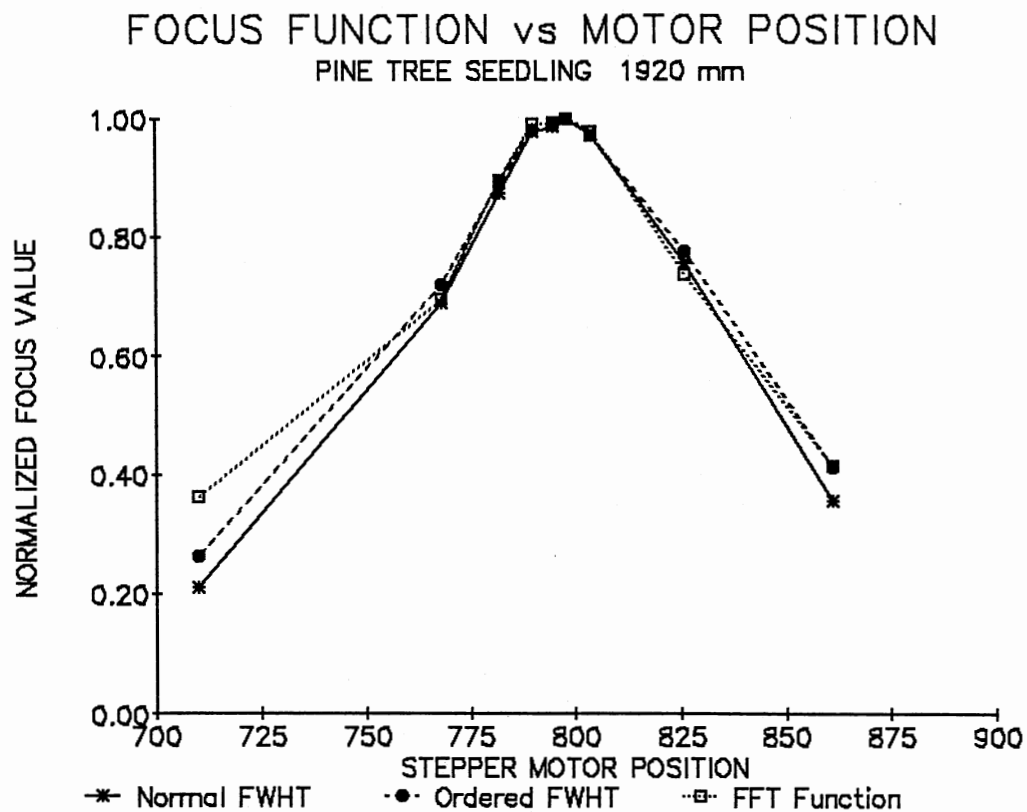### PINE TREE SEEDLING   1920 mm

Figure 13.   Normalized Focus Quality Value Versus Lens Focus Position for the Fourier, Normal-ordered Walsh-Hadamard, and Sequency-ordered Walsh-Hadamard Transforms.

CHAPTER VIII

CALIBRATION OF SYSTEM

Procedure

From geometrical optics, the basic equation for a thin (Gaussian) lens is;

$$\frac{1}{f} = \frac{1}{S_i} + \frac{1}{S_o} \qquad (8\text{-}1)$$

where, f = lens focal length

S_i = distance from lens to image plane

S_o = distance from lens to object plane.

By knowing lens focal length and image plane distance, the distance from the lens to the actual object can be calculated. This relationship is shown in Figure 14. Applying this relationship to a machine vision system becomes more complicated, because most lenses for video cameras have multiple lens elements instead of a single thin (Gaussian) lens. The image distance (ie. distance from the rear of the actual lens to the image sensor) is usually fixed. Focusing is achieved by the translation of internal lens elements.

OBJECT DISTANCE vs IMAGE DISTANCE
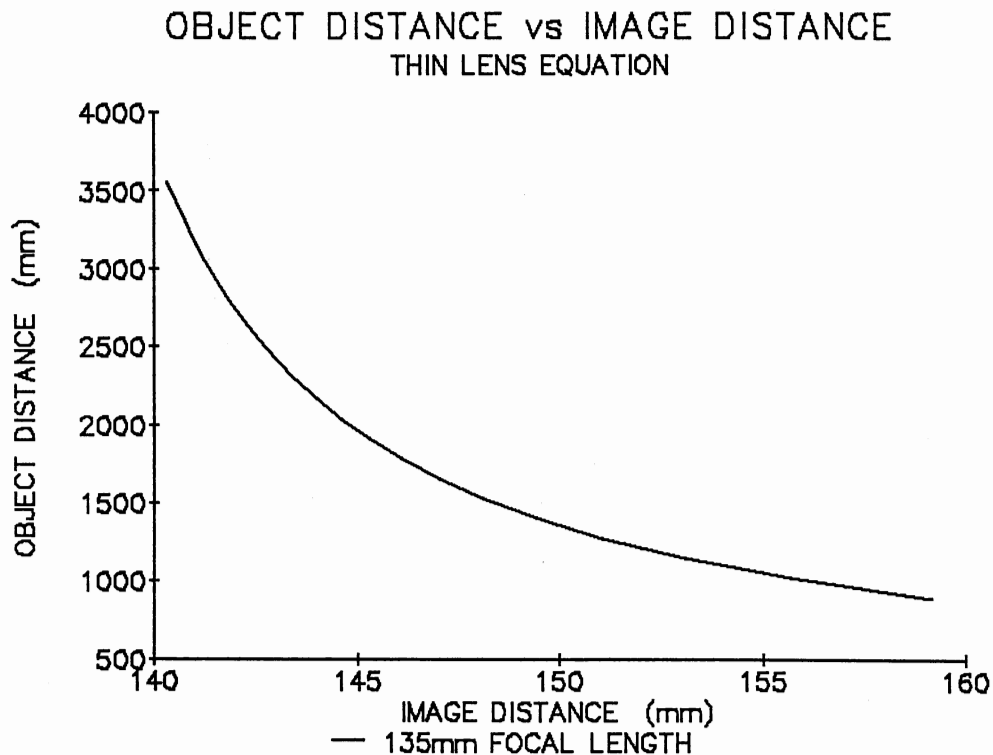### THIN LENS EQUATION

Figure 14. Object Distance Versus Image Distance for a 135mm Thin Lens.

This problem was circumvented by calibrating the camera-lens system to various known object distances. Calibration of the camera lens system was accomplished by first locating an object at a known distance from the camera lens. The software was then executed to find the maximum focus quality function value (optimum image focus). Object distance and stepper motor position at optimum focus were recorded in a data table. The object was placed at a new location and the previous steps were repeated. This procedure was continued until the complete range of object distances was measured. Three calibration curves were developed using three different

algorithms for determining image focus quality. The Fourier transform, normal (dyadic) ordered Walsh-Hadamard transform, and sequency-ordered Walsh-Hadamard transform were the algorithms used.

Results

Fourier Transform

The first calibration data used the Fourier transform as the method to determine image focus quality. Data obtained during calibration were fitted to a curve, plotting stepper motor position against object distance. A total of 51 data points were taken at 8 different distances from the resolution target. The data points were analyzed using statistical and curve-fitting software. An equation of the form inverse X;

$$Y = b0 + b1/X \qquad (8\text{-}2)$$

was fitted to the data with a coefficient of determination,

$$R^2 = 0.9987. \qquad (8\text{-}3)$$

The curve obtained by plotting the prediction equation is similar to a plot of image distance versus object distance using the simple lens equation and a fixed focal length. Figure 15 presents a plot of the prediction equation obtained from the calibration data. A plot of the difference between measured and predicted values is shown in Figure 16.

CALIBRATION CURVE
INVERSE X EQUATION — 51 DATA PTS.

Figure 15. Object Distance Versus Stepper Motor Position for 135mm Lens, FFT Algorithm.

RESIDUAL PLOT
FFT EQUATION — 51 DATA POINTS

Figure 16.   Difference Between Measured and Predicted Distances for Calibration Curve in Figure 15.

Normal-ordered Walsh Transform

The normal-ordered Walsh-Hadamard transform was investigated as a method to determine image focus quality. Data obtained during calibration were fitted to a curve, plotting stepper motor position against object distance. A total of 33 data points were taken at 8 different distances from the resolution target. The data points were analyzed using statistical and curve-fitting software. A power curve equation of the form;

$$Y = aX^b \qquad\qquad (8\text{-}4)$$

was fitted to the data with a coefficient of determination,

$$R^2 = 0.9987. \qquad\qquad (8\text{-}5)$$

Figure 17 presents a plot of the prediction equation obtained from the calibration data. A plot of the difference between measured and predicted values is shown in Figure 18.
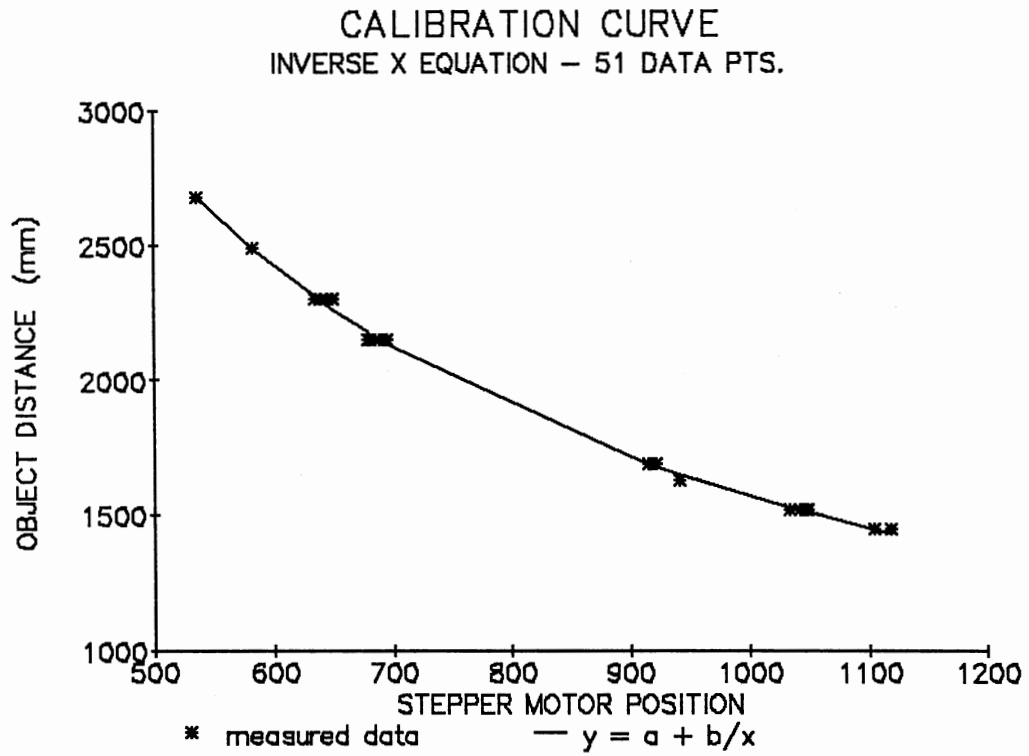
CALIBRATION CURVE
NORMAL FWHT — 33 DATA PTS.

Figure 17. Object Distance Versus Stepper Motor Position for 135mm Lens, Normal-ordered FWHT Algorithm.

RESIDUAL PLOT
NORMAL FWHT — 33 DATA PTS.
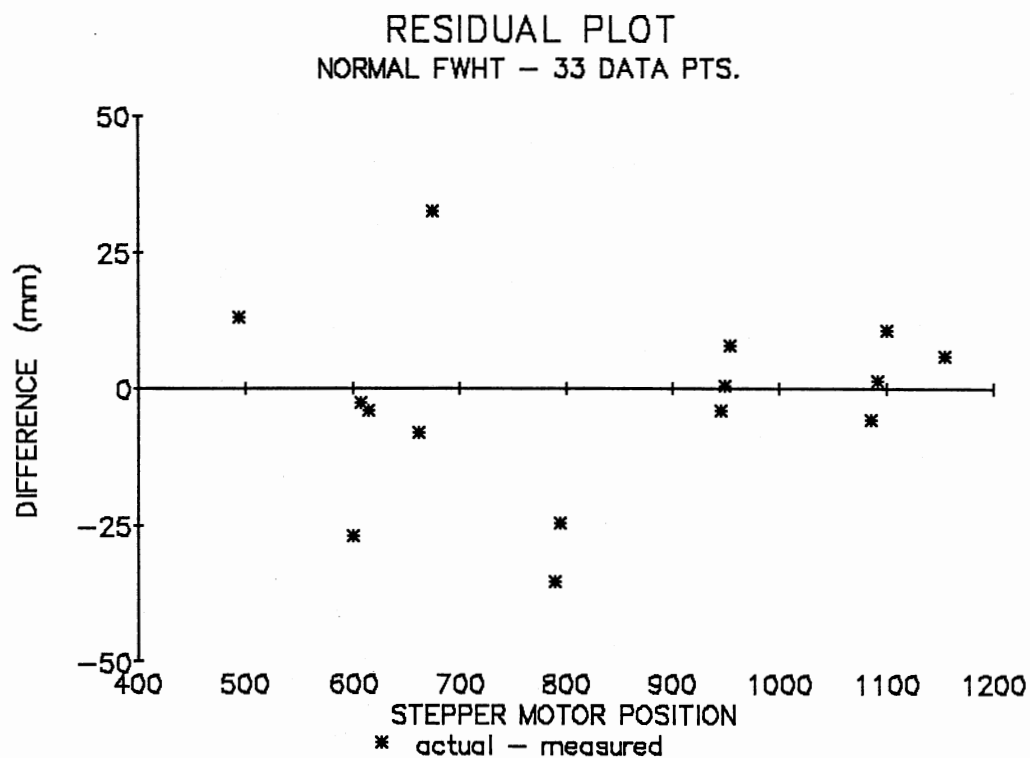
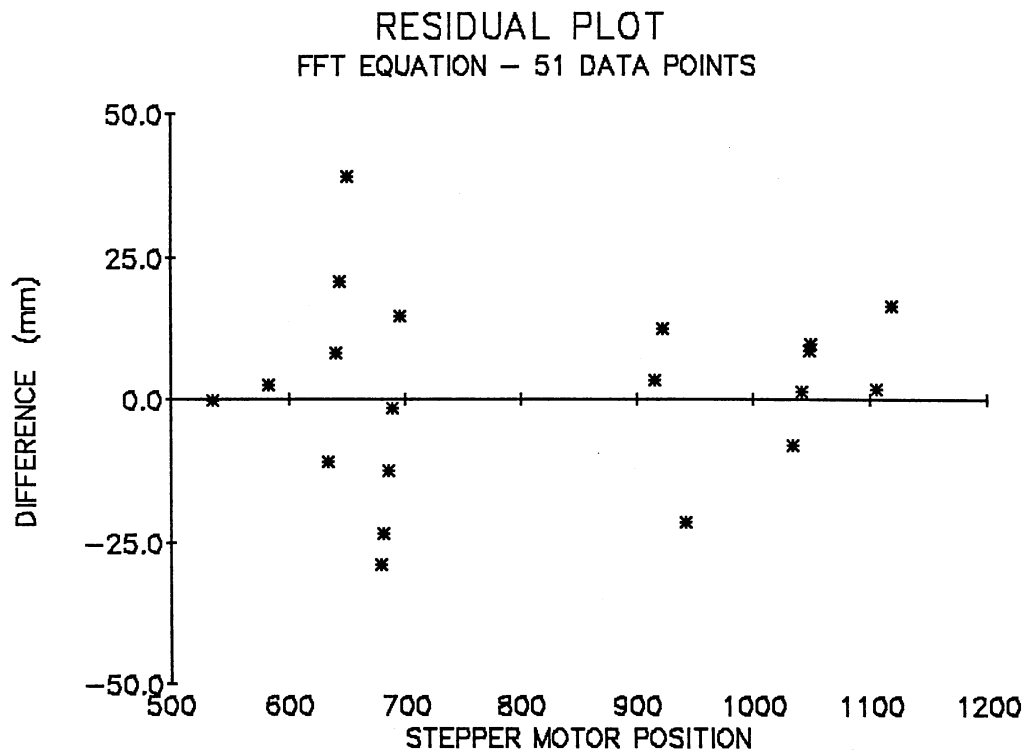Figure 18.   Difference Between Measured and Predicted Distances for
Calibration Curve in Figure 17.

Sequency-ordered Walsh Transform

The sequency-ordered Walsh-Hadamard transform was also
investigated as a method to determine image focus quality. Data obtained
during calibration were fitted to a curve, plotting stepper motor position
against object distance. A total of 34 data points were taken at 8 different
distances from the resolution target. The data points were analyzed using
statistical and curve-fitting software. A power curve equation of the form;

$$Y = aX^b \qquad\qquad (8\text{-}6)$$

was fitted to the data with a coefficient of determination,

$$R^2 = 0.9979. \qquad\qquad (8\text{-}7)$$

Figure 19 presents a plot of the prediction equation obtained from the calibration data. A plot of the difference between measured and predicted values is shown in Figure 20.
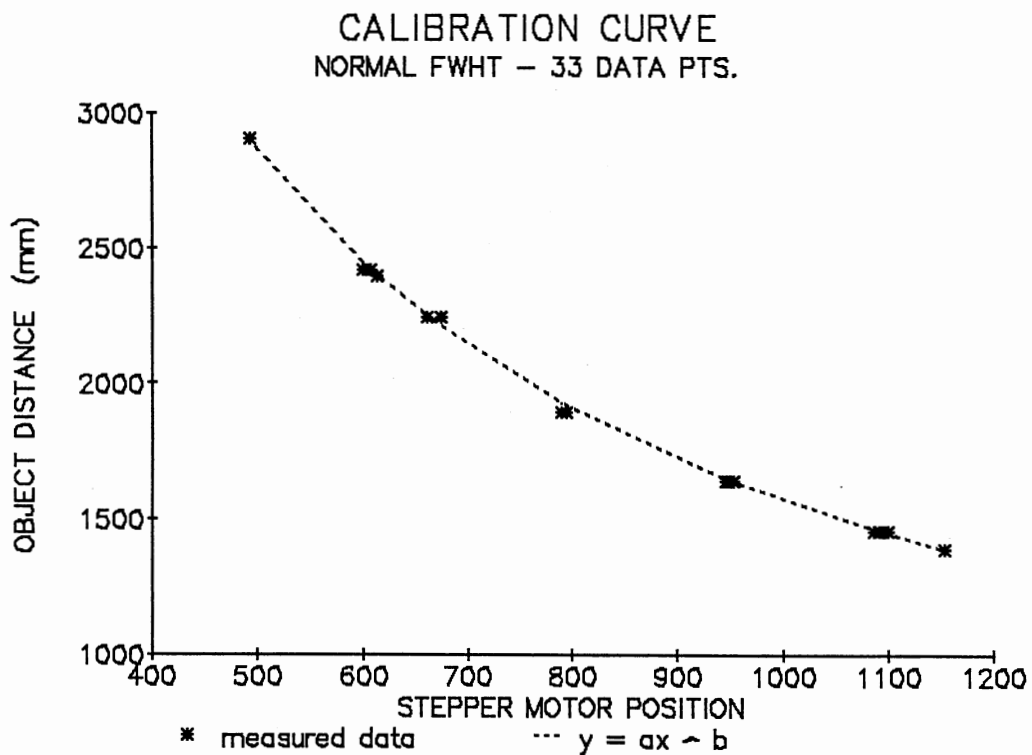
CALIBRATION CURVE
ORDERED FWHT — 34 DATA PTS.
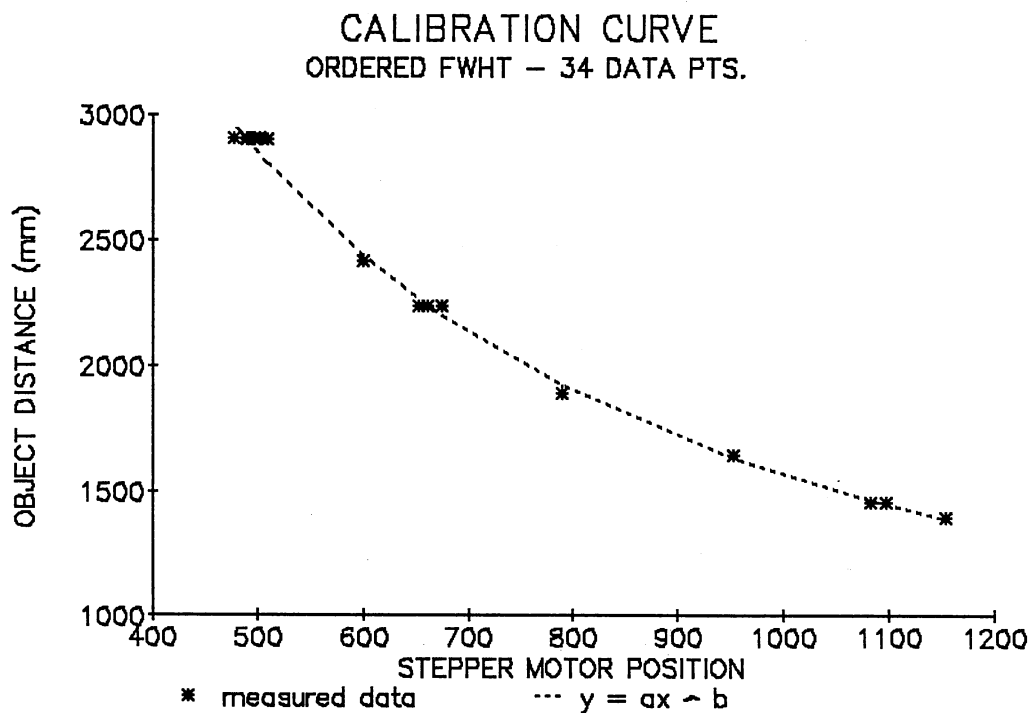
Figure 19. Object Distance Versus Stepper Motor Position for
135mm Lens, Sequency-ordered FWHT Algorithm.

RESIDUAL PLOT
ORDERED FWHT — 34 DATA PTS.

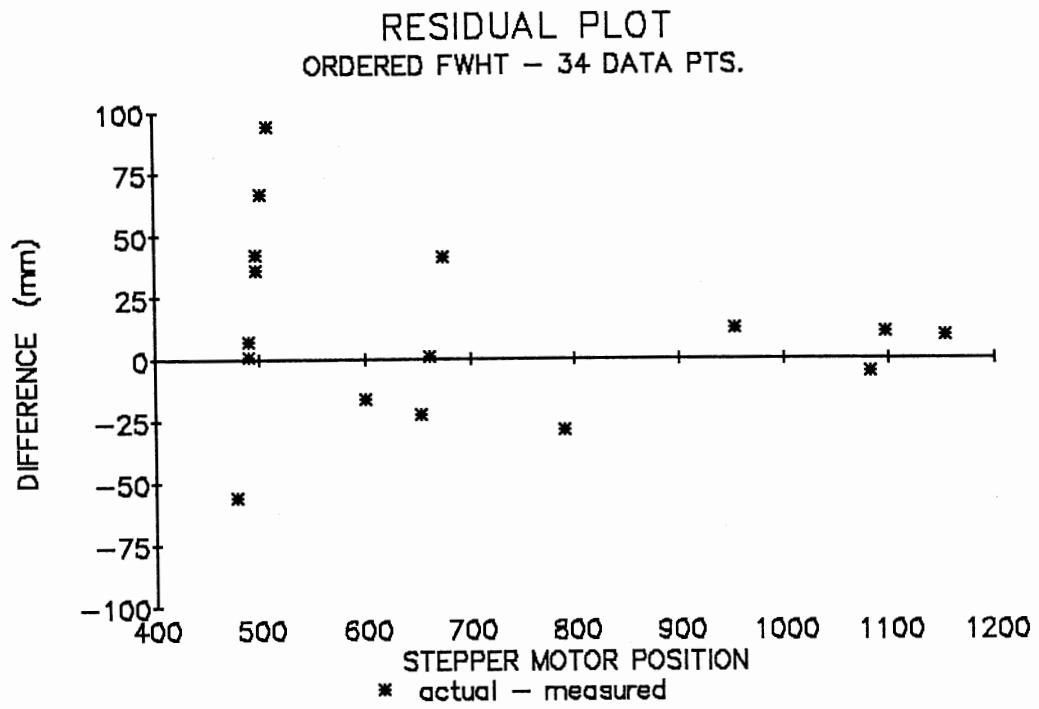Figure 20.   Difference Between Measured and Predicted Distances for
Calibration Curve in Figure 19.

CHAPTER IX

RANGE MEASUREMENTS

Introduction

The calibration equations were added to the MAIN MODULE to enable the system to make range measurements. Stepper motor position corresponding to maximum focus quality value was determined, then entered into the prediction equation to calculate range. Except for this additional step, the software used for range measurement was identical to that used for calibration.

Two different range measurement experiments were performed. The first test used only the Fourier transform to determine image focus quality. A pine tree seedling and an orange were used to test the system with actual three-dimensional objects. Range values from the video camera to the pine tree seedling were calculated at 4 distances, with a total of 17 data points. Distance to the orange was calculated at 2 locations with a total of 11 data points. Lighting conditions were kept constant. An aperture of f/4 was used for all tests. To ensure background uniformity, a Kodak 18% Gray Card was used.

The second test used all three focusing algorithms to compare their accuracy. A pine tree seedling was the object of interest. Range values from the video camera to the pine tree seedling were calculated at 3

distances, with a total of 21 data points (22 data points for normal-ordered Walsh transform). Constant lighting conditions were used, along with a Kodak 18% Gray Card background. An aperture of f/4 was used for all tests.

Results

Test 1

In all but one case, the optically measured distance differed from the actual distance by less than 1%. The average percent difference is less than 0.5%. Table I displays results of the Test 1.

TABLE I

ACTUAL VERSUS MEASURED DISTANCES FOR TEST 1

	Actual Distance (mm)	Average Optically Measured Distance (mm)	Average Percent Difference	Plus & Minus 3 Std Dev (mm)
PINE	1407	1398	0.64%	9.8
TREE	1429	1421	0.56%	6.7
SEEDLING	1892	1876	0.85%	7.8
	2267	2250	0.75%	20.3
ORANGE	1864	1860	0.21%	15.5
	2508	2511	0.12%	25.9

Column 2 of the table lists the mean of the optically measured distances for each location. The last column shows plus-and-minus 3 standard deviations from the mean.

Linear regression analysis was performed on the measured-versus-actual distances for both data sets. Seedling range measurements fit the straight-line equation;

$$Y = 4.93 + 0.99X \qquad (9\text{-}1)$$

with a coefficient of determination,

$$R^2 = 0.9998. \qquad (9\text{-}2)$$

A plot of the data points and the regression equation are shown in Figure 21. The difference between actual and predicted distances is shown in the residual plot of Figure 22.

Figure 21. Actual Versus Measured Distance for a Pine Tree Seedling.

RESIDUAL PLOT
PINE TREE — 17 DATA PTS.

Figure 22. Difference Between Measured and Predicted Distances
for Regression Curve in Figure 21.

Distances from the camera to the orange fit the straight line

equation;

$$Y = -22.94 + 1.01X \qquad\qquad (9\text{-}3)$$

with a coefficient of determination,

$$R^2 = 0.9994. \qquad\qquad (9\text{-}4)$$

A plot of the data points and the regression equation are shown in Figure

23. The difference between actual and predicted distances is shown in the
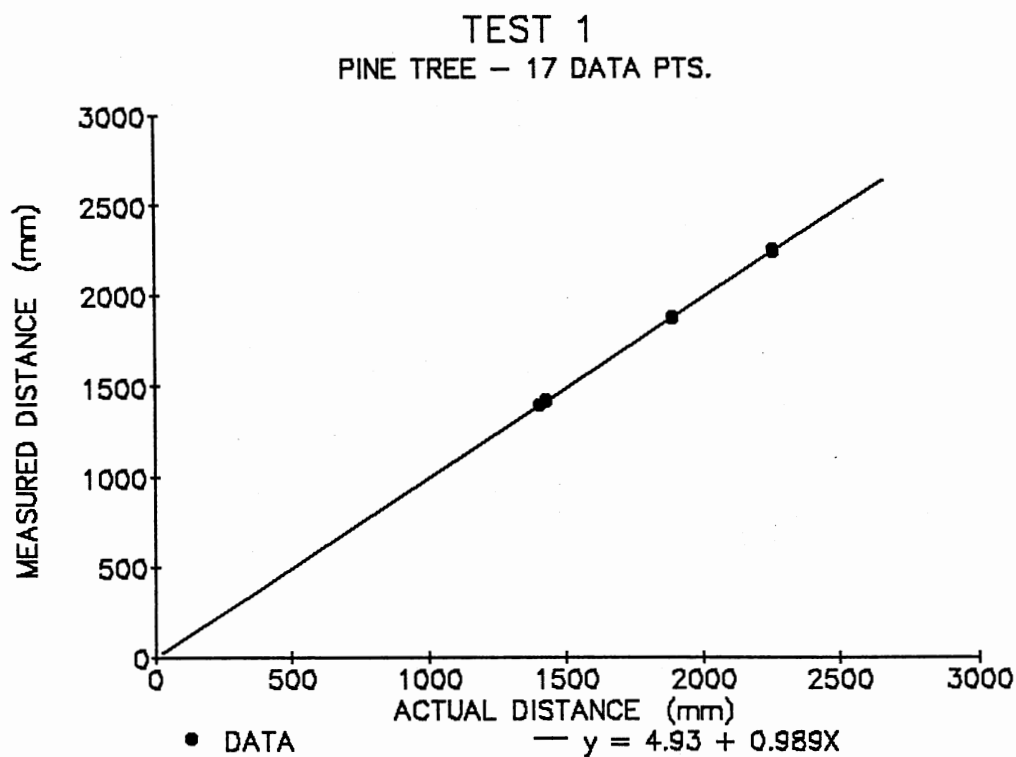
residual plot of Figure 24.
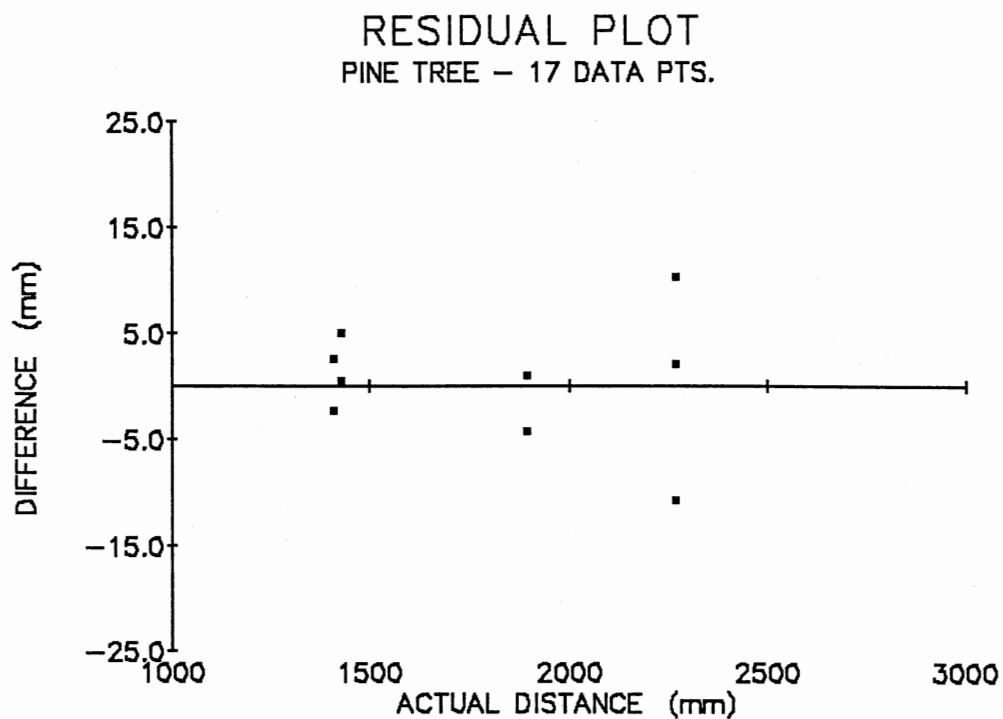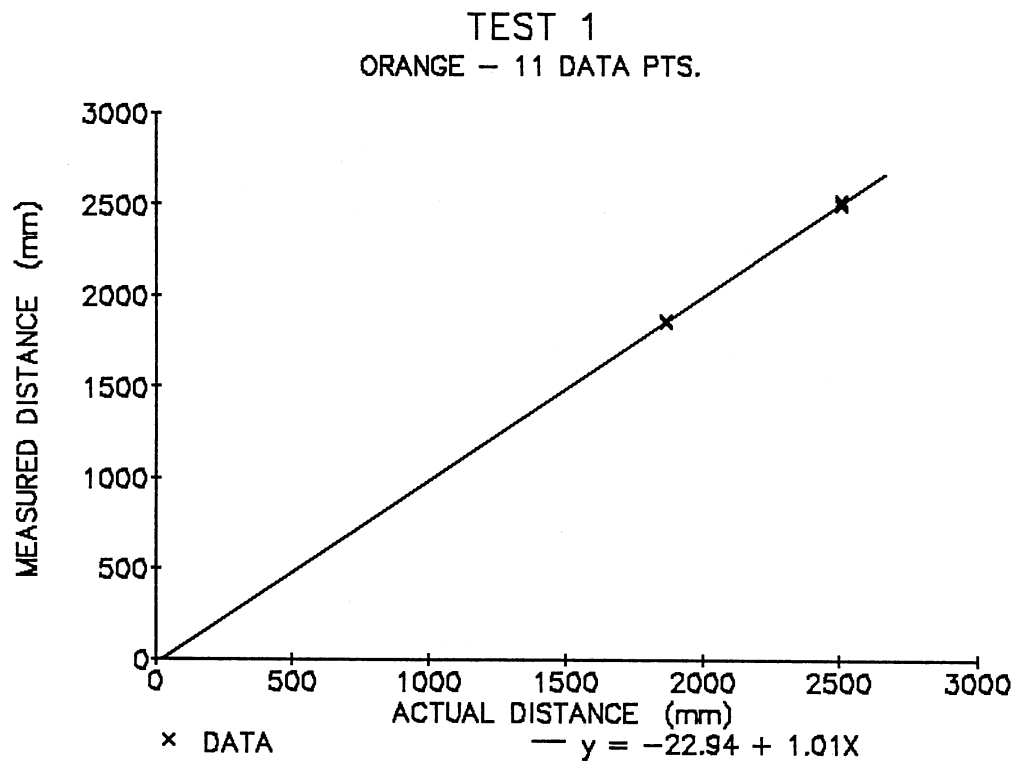
Figure 23. Actual Versus Measured Distance for an Orange.
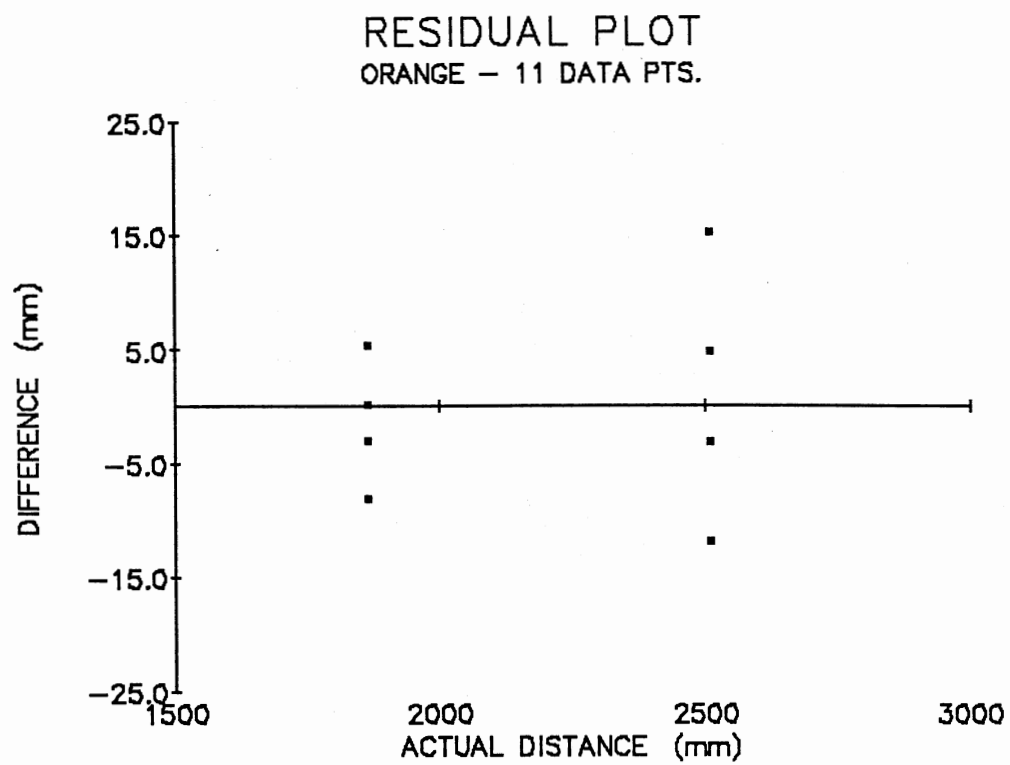
Figure 24. Difference Between Measured and Predicted Distances
for Regression Curve in Figure 23.

Test 2

The second test compared distances calculated by the Fourier, normal-ordered Walsh-Hadamard, and sequency-ordered Walsh-Hadamard transforms. In all but one case case, the optically measured distances differed from the actual distance by less that 1.2%. The average percent difference is less than 0.55%. Table 2 displays the results of Test 2. Column 2 of the table lists the mean of the optically measured distances for each location. The last column shows plus-and-minus 3 standard deviations from the mean.

TABLE II

ACTUAL VERSUS MEASURED DISTANCES FOR TEST 2

	Actual Distance (mm)	Average Optically Measured Distance (mm)	Average Percent Difference	Plus & Minus 3 Std Dev (mm)
FOURIER	1486	1481	0.31%	15.5
TRANSFORM	1918	1901	0.89%	23.4
	2585	2566	0.72%	17.5
NORMAL	1486	1478	0.53%	14.0
ORDERED	1918	1919	0.07%	25.1
FWHT	2585	2577	0.29%	18.0
SEQUENCY	1486	1476	0.67%	14.2
ORDERED	1918	1904	0.70%	0.0
FWHT	2585	2572	0.48%	18.0

Linear regression analysis was performed on the measured-versus-actual distances for all three data sets. Optical range measurements using the Fourier transform fit the straight-line equation;

$$Y = 12.020 + 0.9879X \tag{9-5}$$

with a coefficient of determination,

$$R^2 = 0.9997. \tag{9-6}$$

A plot of the data points and the regression equation are shown in Figure 25. The difference between actual and predicted distances is shown in the residual plot of Figure 26.

PINE TREE — 21 DATA PTS.
FOURIER TRANSFORM

Figure 25. Actual Versus Measured Distance for a Pine Tree Seedling.

Figure 26. Difference Between Measured and Predicted Distances
for Regression Curve in Figure 25.

Distances from the camera to the pine tree seedling using the normal-ordered Walsh-Hadamard transform algorithm fit the straight line equation;

$$Y = -2.840 + 0.9986X \qquad (9\text{-}7)$$

with a coefficient of determination,

$$R^2 = 0.9998. \qquad (9\text{-}8)$$

A plot of the data points and the regression equation are shown in Figure 27. The difference between actual and predicted distances is shown in the residual plot of Figure 28.

Figure 27. Actual Versus Measured Distance using Normal-ordered Walsh-Hadamard Transform.

RESIDUAL PLOT
REG. ORDERED FWHT — 22 DATA PTS.
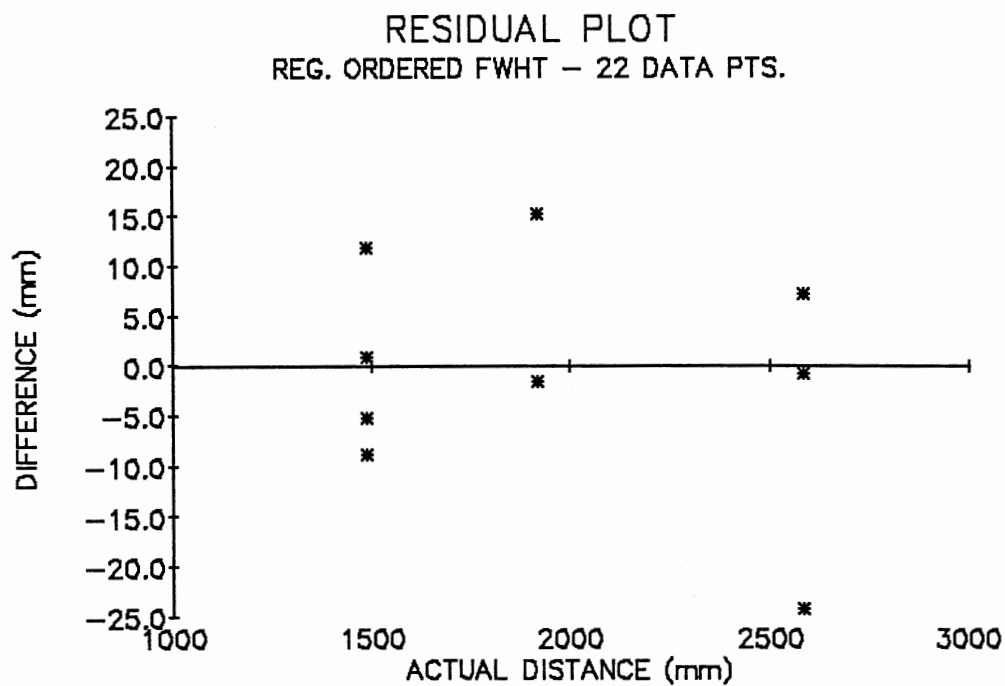
Figure 28. Difference Between Measured and Predicted Distances for Regression Curve in Figure 27.

Distances from the camera to the pine tree seedling using the sequency-ordered Walsh-Hadamard transform algorithm fit the straight line equation;

$$Y = -7.142 + 0.9978X \qquad (9\text{-}9)$$

with a coefficient of determination,

$$R^2 = 0.9999. \qquad (9\text{-}10)$$

A plot of the data points and the regression equation are shown in Figure 29. The difference between actual and predicted distances is shown in the residual plot of Figure 30.
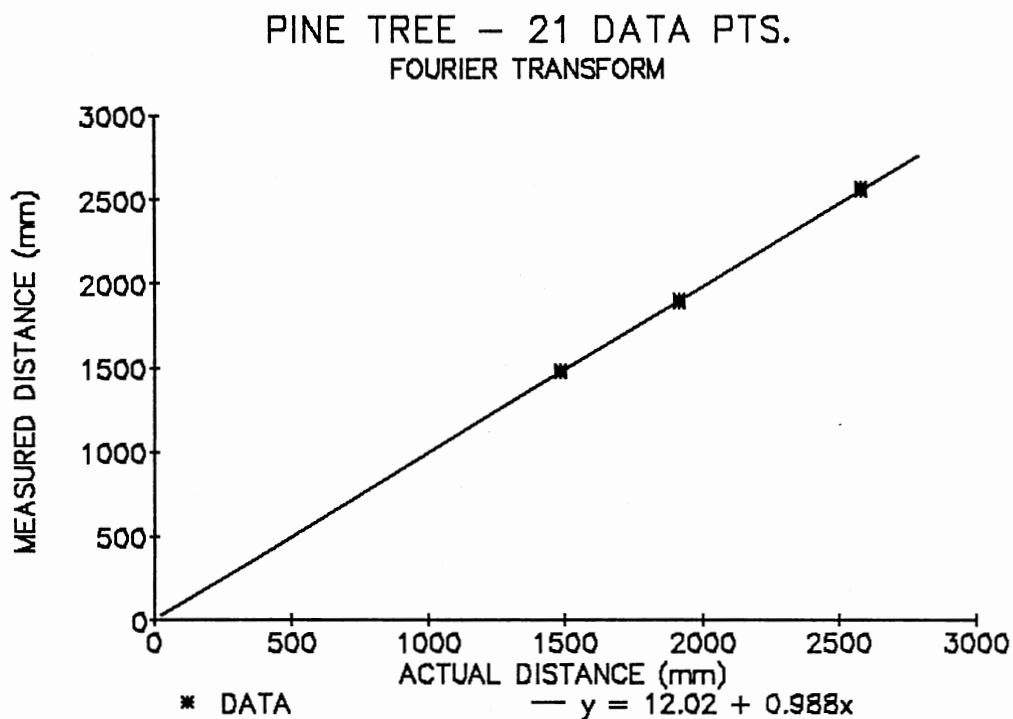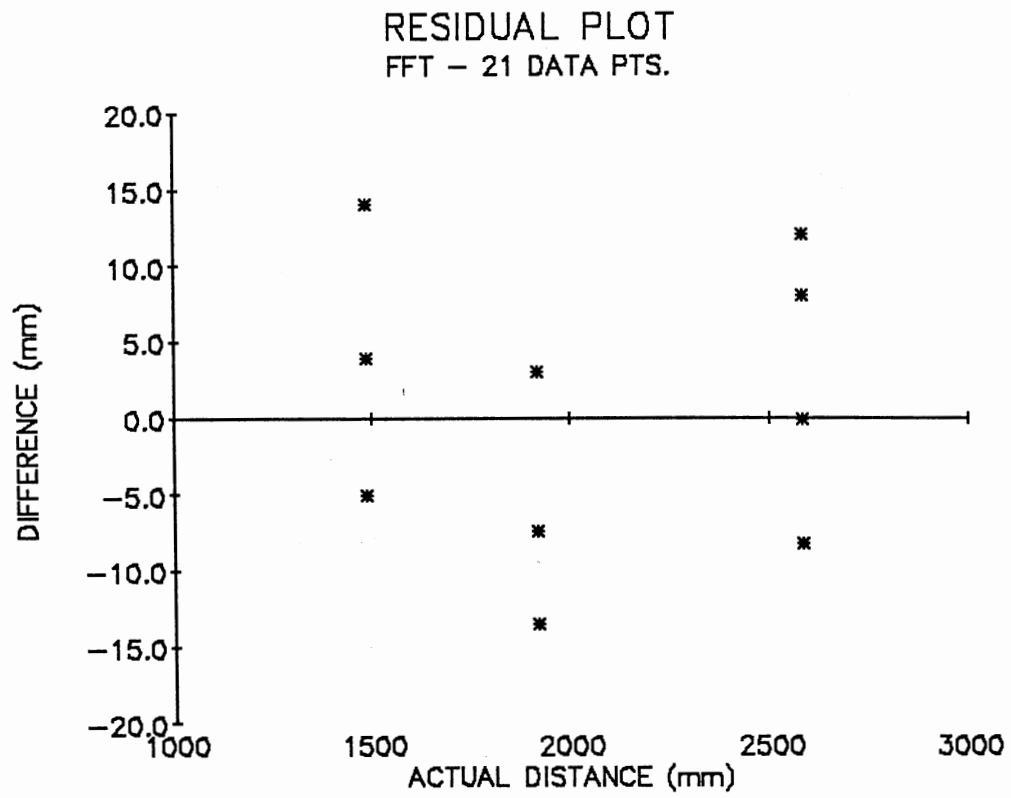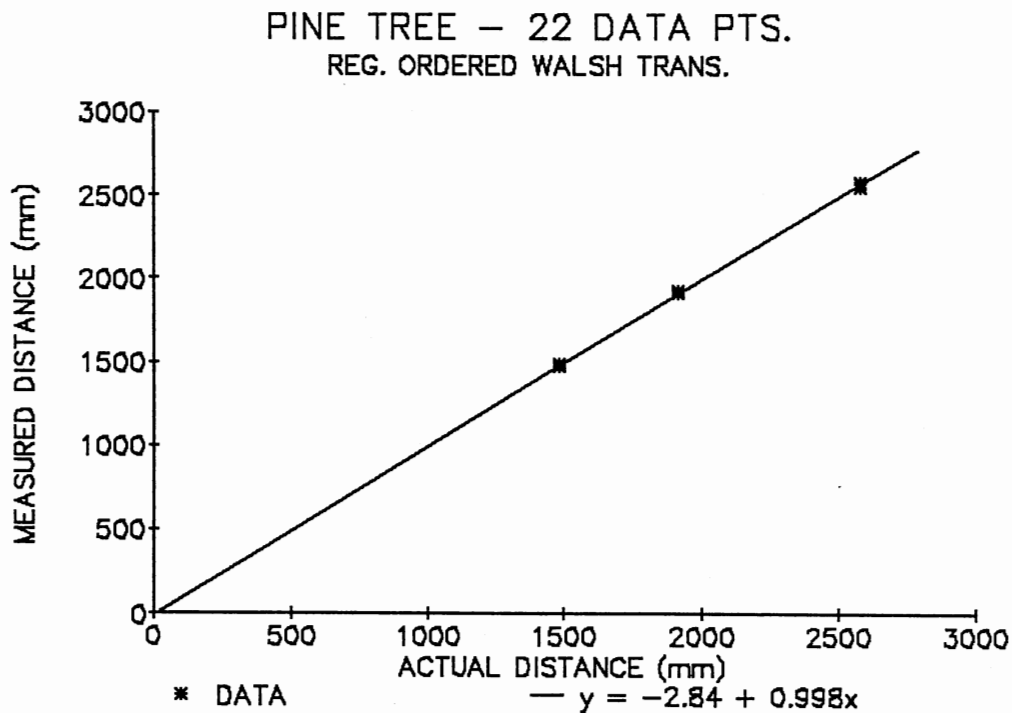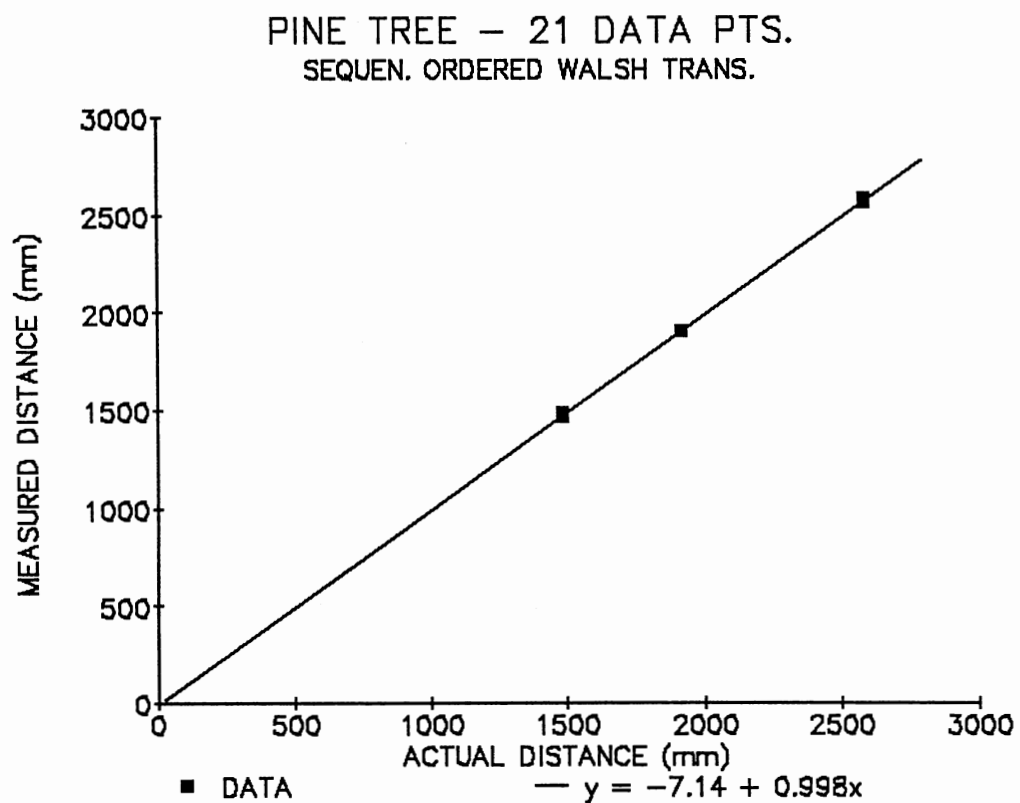
Figure 29. Actual Versus Measured Distance using Sequency-ordered
Walsh-Hadamard Transform.

RESIDUAL PLOT
SEQ. ORDERED FWHT — 21 DATA PTS.
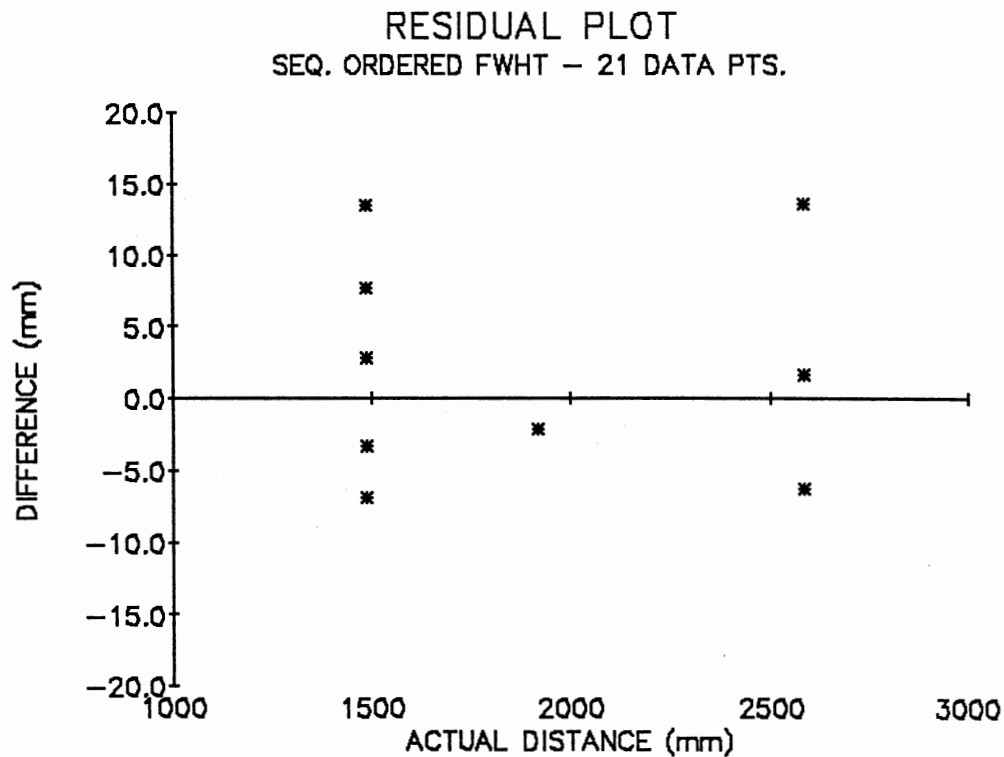
Figure 30. Difference Between Measured and Predicted Distances
for Calibration Curve in Figure 29.

Due to equipment constraints, range measurements were limited to the interval between 1400 and 2600 mm. Minimum focusing distance was slightly less than 1400 mm for the lens used. Power supply and video cable length, as well as room size restricted measurements to less than 3000 mm. A compromise was required between execution speed and a window size containing sufficient area of the object of interest. The image quality function was evaluated within a 64 x 64 pixel window.

CHAPTER X

CONCLUSIONS AND RECOMMENDATIONS

Summary

A technique based on autofocusing was developed to optically measure the distance from a video camera to an object of interest. Software was written to determine optimum image focus, and to control a motorized camera lens. The system was calibrated to obtain a relationship between stepper motor position and object distance. Distance measurements were performed using an orange and pine tree seedling as target objects.

This system used a passive, non-triangulation technique to obtain the distance from the machine vision camera to the object of interest. A passive, non-triangulation system has the simplest image acquisition requirements, since it does not require a second camera, structured lighting, camera movement, or time-of-flight equipment. This simplicity must be balanced against the limited information obtained from a monocular view of the scene. For range measurement, this method should provide adequate information for robot arm guidance. The simple hardware requirements are also an advantage when designing a rugged, cost-effective system for use in agriculture.

Conclusions

Comparison of the focus quality functions showed very little difference in ability to determine optimum image focus. The "moment-of-inertia" calculation used to evaluate transform coefficients proved to be an effective means of determining focus quality. As shown by the plots in Chapter VII, the three transform algorithms provided nearly identical results.

Both Walsh-Hadamard transform algorithms appeared to perform more effectively than the FFT when there was limited information content in the image. This low-contrast condition occurred when the camera lens was grossly defocused. Focus quality values generated by the FFT demonstrated larger fluctuations than those generated by the Walsh transform under harsh image defocus. These qualitative observations should be investigated in more detail.

Accuracy

In the worst case, calculated distance varied from actual distance by almost 30 mm, representing a difference of just over 1%. One possible source of error is the depth-of-field of the lens. Depth-of-field is the distance interval in which an object can reside while maintaining a sharply focused image. The distance interval is a function of lens focal length, aperture, and object distance. Depth-of-field is not a linear function. For a 135-mm lens with f-stop at f/4, the depth-of-field is approximately 40 mm for an object at 1500 mm, and 110 mm for an object at 2500 mm. This

relationship suggests an inherent error of 2.7% for an object at 1500mm, and an error of 4.4% for an object at 2500mm. These values drop to 27.7mm (1.9%) at 1500mm and 76.8mm (3.1%) at 2500mm when the f-stop is reduced to f/2.8. All distance measurement errors were well below the depth-of-field error.

At first glance, the demonstrated degree of measurement accuracy appears highly unlikely. Re-examination of the calibration procedure explains the apparent contradiction. During calibration, the camera was positioned at various known distances from the resolution target. The focus quality software was executed and the stepper motor position yielding optimum image focus determined. This procedure was extensively repeated, producing a distribution of motor positions for a given distance. All motor positions were at optimum image focus, ie. within the depth-of-field.

The statistical and curve-fitting software package fit a least-squares regression curve to the calibration data. As shown by the residual plots in Chapter VIII, the regression curve minimizes the difference between measured and predicted data values. This procedure corresponds to placing the regression curve somewhere near the middle of the depth-of-field interval. When the calibration equation is used to predict the distance to an object, the error should be less than one-half the depth-of-field distance. This condition was true in almost all cases.

Sensitivity of the focus quality function is also a possible source of error. The function converges on the portion of the image with the greatest number of edges. Therefore, if the desired measurement point has fewer edges than a nearby location, the function focuses on the position with the greatest number of edges, and the distance calculated is compromised. The

pine tree seedling measurements contain this type of error. Actual measured distances were from the camera to the main stem, whereas the software tended to focus on needles slightly ahead of the stem. The calculated range was always slightly less than the actual distance to the stem.

Another problem can occur when there are insufficient edges or not enough contrast in the sharply focused image. Sufficient high-frequency components must be generated by actual edges in order to rise above the noise floor of the spectrum. This requirement was evident for example, when the image window contained only a small section of the orange surface. The dimples on the orange rind surface often failed to present sufficient contrast to enable the software to converge on sharp focus.

Test 1

The results of Test 1 are shown in Figures 17, 18, 19, 20, and Table I. In all but one case, the optically measured distance differed from the actual distance by less than 1%. Linear regression analysis of the measured versus actual distances for both the orange and pine tree show very close agreement. The residual plots show an increasing difference between measured and predicted values as the actual object distance increases. This was expected, since depth-of-field increases as object distance increases.

For scenes with edges and contrast, the system performed flawlessly and with greater accuracy than indicated by the depth-of-field limitation. Distance errors averaged less than 0.5%. The Golden Section search algorithm required between 7 and 11 iterations to locate the maximum focus quality value. Calculation of the 64 x 64 point FFT took approximately 55

seconds, while the balance of the software required between 30 and 90 seconds to execute, depending on the number of iterations performed by the search routine. Each focusing cycle totaled approximately 20 to 30 minutes to complete.

Test 2

The objective of the second series of measurements was comparison of the Fourier, normal-ordered, and sequency ordered Walsh-Hadamard transforms. A pine tree seedling was used as the object of interest. Table II and Figures 21 through 26 summarize the data.

Results of the data are somewhat inconclusive. All three focus quality algorithms performed equally well. In all but one case, the optically measured distance differed from the actual distance by less than 1.2%. The average percent difference was less than 0.55%. Again, all measurements were well within the depth-of-field error. Distribution of the measurement errors displayed no conclusive pattern, as shown in Table 5. Regression analysis confirmed a highly linear relationship between measured and actual distances. The residual plots show a somewhat random distribution of data points. In contrast to the residual plot from Test 1, the difference values do not increase as actual distance increases. Additional experiments are needed to clarify the error distribution. More measurements at each selected distance, and several more test distances should be used to provide for a statistically valid analysis.

Slow execution speed of the system prevented additional data collection. Reducing the processing time would increase the amount of data

collected during a given test period. The Walsh-Hadamard transform calculation took approximately 17 seconds for a 64 x 64 pixel image. This time could have been reduced by rewriting the code using integer variables and arithmetic instead of floating-point values. The Walsh transform calculation was still more than three times faster the FFT.

Optimizing the hardware and software could reduce the focus cycle time by almost two orders of magnitude. The use of a more efficient search algorithm such as Brent's method (Press, et al. 1986) could reduce by about one-third the number of iterations needed to locate the maximum focus quality value. Implementing the transform calculations in silicon could dramatically decrease execution time. Various chips have been designed to calculate the Walsh-Hadamard transform (Clarke 1985, Yarlagadda and Hershey 1981). With the availability of low-cost image processing boards capable of computing the FFT on a 128 x 128 image in less than 0.4 seconds, this technique should prove valuable for autofocusing and optical range-finding.

Recommendations for Future Research

The objectives of this project have been successfully completed and have laid a foundation for future work. In the preceding section some suggestions were made for further study. These will be incorporated into the following recommendations.

1. Increase speed of execution.

   As previously mentioned, the system required 20 to 30 minutes for one distance measurement. This pace limits the amount of data that can be collected. A more efficient search algorithm and taking advantage of increased hardware processing power could significantly reduce execution time. Further investigation into the behavior of the focus quality algorithms would be facilitated.

2. Investigate behavior of focus quality algorithms.

   The focus quality functions should be tested with a variety of different objects. Sufficient data points should be gathered to clarify the distribution of errors and the factors contributing to the error. Fluctuations of the focus quality value, when limited image information (harsh defocus) is present, should be studied. Noise level in the focus quality value affects the performance when the lens is grossly defocused. Possible sources of noise in the system should be found and corrected.

3. Zoom lens control.

> A practical system using this technique would most likely utilize a zoom (variable-focal-length) lens. Since depth-of-field is a function of focal length, zoom would allow control over the error due to depth-of-field. The focal lengths used would depend on the relative size of object in question and its range of distances from the camera.

4. Implement object recognition

> In conjunction with the previous recommendation, the addition of object recognition would be required for a general-purpose robot system. The final system would take a global view of a scene with a short focal length, then decide on the object of interest within that scene, and zoom in on that object with a longer focal length to make the distance measurement.

REFERENCES

Ahmed, N., Rao, K. R. and Abdussattar, A. L., "BIFORE or Hadamard Transform." IEEE Transactions of Audio - Electroacoustics, Vol. AU-19, 1971, pgs. 225-234.

Anderson, P. L., "Real-Time Gray-Scale Video Processing using a Moment-Generating Chip." IEEE Journal of Robotics and Automation, Vol. RA-1, No. 2, June 1985, pgs. 79-85.

Bajcsy, R., "Three-dimensional scene analysis." Proceedings of the Fifth International Conference on Pattern Recognition, IEEE Computer Society, Vol. 2, December 1980, pgs. 1064-1074.

Ballard, D. H. and Brown C. M., "Computer Vision." Prentice-Hall Inc., New Jersey, 1982.

Beauchamp, K. G., "Applications of Walsh and Related Functions." Academic Press Inc., Orlando, Florida, 1984.

Besl, P. J. and Jain, R. C., "Three Dimensional Object Recognition." ACM Computing Surveys, Vol. 17, No. 1, March 1985, pgs. 75-145.

Bickel, G., Hausler, G. and Maul, M., "Triangulation with Expanded Range of Depth." Optical Engineering, Vol. 24, No. 6, Nov-Dec. 1985, pgs. 975-977.

Born, M. and Wolf E., "Principles of Optics - 6th ed." Pergamon Press Inc., New York, 1983.

Bracewell, R. N., "The Fourier Transform and its Applications - 2nd ed." McGraw-Hill Book Co., New York, 1986.

Brigham, E. O., "The Fast Fourier Transform and its Applications." Prentice-Hall Inc., New Jersey, 1988.

Cathey, W. T. and Davis, W. C., "Vision System with Ranging for Maneuvering in Space." Optical Engineering, Vol. 25, No. 7, July 1986, pgs. 821-824.

Cheney, W. and Kincaid, D., "Numerical Mathematics and Computing." Brooks/Cole Publishing Company, Monterey, California, 1980.

Clarke, R. J., "Tranform Coding of Images." Academic Press Inc., London, 1985.

Cooley, J. W., and Tukey, J. W., "An Algorithm for the Machine Calculation of Complex Fourier Series." Mathematics of Computation, Vol. 19, No. 90, April 1965, pgs. 297-301

De Witt, T., "Range Finding With Diffraction Gratings." Advanced Imaging Magazine, No. 12, July-Aug. 1988, pgs. 50-56.

Everett, H. R., "Noncontact Ranging Systems for Mobile Robots." SENSORS Magazine, April 1987, pgs. 9-19.

Fuma, F., Krotkov, E. and Summers, J., "The Pennsylvania Active Camera System." Dept. of Computer and Information Science, University of Pennsylvania, T.R. CIS-86-15 Grasp Lab 62, 1986.

Gaskill, J. D., "Linear Systems, Fourier Transforms and Optics." John Wiley & Sons, New York, 1978.

Gonzalez, R. C. and Wintz, P., "Digital Image Processing." Addison-Wesley Publishing Co., Reading, Massachussets, 1977.

Goodman, J. W., "Introduction to Fourier Optics." McGraw-Hill Book Co. Inc., New York, 1968.

Hall, E. L., "Computer Image Processing and Recognition." Academic Press, New York, 1979.

Hall, E. L. and McPherson, C. A., "Three Dimensional Perception for Robot Vision." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol. 442, 1983 pgs. 117-142.

Harmuth, H. F., "Sequency Theory - Foundations and Applications." Academic Press, New York, 1977.

Harrell, R. C., Adsit, P. D. and Slaughtter, D. C., "Real-Time Vision-Servoing of a Robotic Tree Fruit Harvester." presented at the 1985 Winter Meeting of the American Society of Agricultural Engineers, Chicago, IL, 1985, paper no. 85-3550.

Hecht, E. and Zajac, A., "Optics." Addison-Wesley Publishing Co. Inc., Reading, Massachusetts, 1979.

Hein, D. and Ahmed, N., "On a Real-Time Walsh-Hadamard /Cosine Transform Image Processor." IEEE Transactions on Electromagnetic Compatibility, Vol. EMC-20, No. 3, August 1978, pgs. 453-457.

Horn, B., "Focusing." MIT Project MAC, A.I. Memo 160, Massachussets Institute of Technology, May 1968.

Jalkio, J. A., Kim, R. C. and Case, S. K., "Three-Dimensional Inspection using Multistripe Structured Light." Optical Engineering, Vol. 24, No. 6, Nov-Dec. 1985, pgs. 966-974.

Jarvis, R. A., "Focus Optimisation Criteria for Computer Image Processing." Microscope, Vol. 24, No. 2, 1976, pgs. 163-180.

Jarvis, R. A., "A Perspective on Range Finding Techniques for Computer Vision." IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No. 2, March 1983, pgs. 122-139.

Kanade, T. and Sommer, T. M., "An Optical Proximity Sensor for Measuring Surface Position and Orientation for Robot Manipulation." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol.449, 1983, pgs. 667-674.

Kinoshita, G., Idesawa, M. and Naomi, S., "Robotic Range Sensor With Projection of Bright Ring Pattern." Journal of Robotic Systems, Vol. 3, No. 3, 1986, pgs. 249-257.

Koenigsberg, W. D., "Noncontact Distance Sensor Technology." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol. 449, 1983, pgs. 519-531.

Kowel, S. T., Kornreich, P. G., Mahapatra, A., Mehter, M. and Szebenyi, T., "Image Feature Analysis using DEFT Sensors." Remote Sensing of Earth from Space: Role of "Smart Sensors." Breckenridge, R. A., Editor, Vol. 67, Progress in Astronautics and Aeronautics, American Institute of Aeronautics and Astronautics, 1979.

Krotkov, E., "Focusing." Dept. of Computer and Information Science, University of Pennsylvania, T.R. CIS-86-22, Grasp Lab 63, 1986.

Krotkov, E.P. and Martin, J., "Range from Focus." Dept. of Computer and Information Science, University of Pennsylvania, T.R. CIS-86-09, Grasp Lab 59, 1986.

Krotkov, E., "Focusing." International Journal of Computer Vision, Vol. 1, No. 3, 1987, pgs. 223-237.

Krutz, G. W., "Future Use of Robotics in Agriculture." Proceedings of the First International Conference on Robotics and Intelligent Machines in Agriculture, October 2-4, 1983.

Lewis, A., "Investment Analysis for Robotics Applications." 13th ISIR Robotics Proceedings Conference, Chicago, April 1983, pgs. 4-128 to 4-139.

McFarland, W.D., "Three Dimensional Images for Robot Vision." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol. 442, 1983, pgs. 108-116.

Okada, T., "Development of an Optical Distance Sensor for Robots." The International Journal of Robotics Research, Vol. 1, No. 4, Winter 1982, pgs. 3-14.

Orrock, J. E., Garfunkel, J. H. and Owen, B. A., "An Integrated Vision/Range Sensor." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol. 449, 1983, pgs. 419-425.

Ozeki, O., Nakano, T. and Yamanoto, S., "Real-time Range Measurement Device for Three Dimensional Object Recognition." IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 4, July 1986, pgs. 550-554.

Pejsa, J. H. and Orrock, J. E., "Intelligent Robot Systems: Potential Agricultural Applications." Proceedings of the First International Conference on Robotics and Intelligent Machines in Agricultural, October 2-4, 1983.

Pipitone, F.J. and Marshall, T.G., "A Wide-field Scanning Triangulation Rangefinder for Machine Vision." The International Journal of Robotics Research, Vol. 2, No. 1, Spring 1983, pgs. 39-49.

Potmesil, M. and Chakravarty, I., "Synthetic Image Generation with a Lens and Aperture Camera Model." ACM Transactions on Graphics, Vol. 1, No. 2, April 1982, pgs. 85-108.

Pratt, W. K., "Digital Image Processing." John Wiley & Sons, Inc., 1978.

Pratt, W. K., Kane, J. and Andrews, H. C., "Hadamard Transform Image Coding." Proceedings of the IEEE, Vol. 57, No. 1, Jan. 1969, pgs. 58-68.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T., "Numerical Recipes - The Art of Scientific Computing." Cambridge University Press, Cambridge, UK, 1986.

Schade, Sr., O. H., "Image Quality - A Comparison of Photographic and Television Systems." RCA Laboratories, Princeton, New Jersey, 1975.

Schade, Sr., O. H., "Image Quality - A Comparison of Photographic and Television Systems." Society of Motion Picture and Television Engineers-SMPTE Journal, Vol. 96, No. 6, June 1987.

Schlag, J.F., Sanderson, A.C., Neuman, C.P. and Wimberly, F.C., "Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control." Dept. of Electrical Engineering, Carnegie-Mellon University, CMU-RI-TR-83-14, 1983.

Selker, T., "Image-based Focusing." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol. 360, 1983, pgs. 96-99.

Shafer, S. A., "Geometric Camera Calibration for Machine Vision Systems." Manufacturing Engineering, Vol. 102, No. 3, March 1989, pgs. 85-88.

Shazzer, D. and Harris, M., "Digital Autofocus using Scene Content." Proceedings of the Society of Photo-Optical Instrumentation Engineers, Vol.534, 1985, pgs. 150-158.

Sistler, F. E., "Robotics and Intelligent Machines in Agriculture." IEEE Journal of Robotics and Automation, Vol. RA-3, No. 1, Feb. 1987, pgs. 3-6.

Sobel, I., "Camera Models and Machine Perception." Stanford Artificial Intelligence Project, Memo AIM-121, May 1970.

Strand, T.C., "Optical Three Dimensional Sensing for Machine Vision." Optical Engineering, Vol. 24, No. 1, January 1985.

Sugai, M., Kanuma, A., Suzuki, K. and Kubo, M., "VLSI Processor for Image Processing." Proceedings of the IEEE, Vol. 75, No. 9, Sept. 1987, pgs. 1160-1166.

Tsai, R. Y., "A Versatile Camera Calibration Technique For High-Accuracy 3D-Machine Vision Metrology using Off-The-Shelf TV Cameras And Lenses." IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, August 1987, pgs. 323-344.

Tseng, H. F., Ambrose, J. R. and Fattahi, M., "Evolution of the Solid-State Image Sensor." Journal of Imaging Science, Vol. 29, No. 1, Jan-Feb. 1985, pgs. 1-7.

Wolpert, H. D., "Autoranging/Autofocus: A Survey of Systems." Photonics Spectra Magazine, June 1987, pgs. 165-168.

Yarlagadda, R. and Hershey, J. E., "Architecture of the Fast Walsh-Hadamard and Fast Fourier Transforms using Charge-Transfer Devices." International Journal of Electronics, Vol. 51, No. 5, 1981, pgs. 669-681.

VITA α

Paul Reese Weckler

Candidate for the Degree of

Doctor of Philosophy

Thesis: OPTICAL RANGE-FINDING FROM IMAGE FOCUS

Major Field: Agricultural Engineering

Biographical:

Personal Data: Born in Brigham City, Utah, April 22, 1958, the son of Gene P. and Helen R. Weckler.

Education: Received Bachelor of Science Degree with Honors in Agricultural Engineering from California Polytechnic State University at San Luis Obispo in June, 1982; received Master of Science Degree in Agricultural and Irrigation Engineering from Utah State University at Logan in August, 1984; completed requirements for the Doctor of Philosophy Degree at Oklahoma State University in December 1989.

Professional Experience: Staff Engineer, Hydratec Inc., Delano, CA, June 1982, to September 1982. Research Assistant, Agricultural and Irrigation Engineering Department, Utah State University, September 1982 to June 1984. Graduate Assistant, Electrical and Computer Engineering Department, University of Arizona, August 1984 to June 1985. USDA National Needs Ph.D. Fellow, Agricultural Engineering Department, Oklahoma State University, June 1985 to June 1988. Teaching Assistant, Agricultural Engineering Department, Oklahoma State University, August 1988 to December 1988.

Professional Organizations: Tau Beta Pi; Institute of Electrical and Electronic Engineers; Society of Photo-Optical Instrumentation Engineers; American Society of Agricultural Engineers; Sigma Xi; Alpha Epsilon.