12-1999

# Calibration of scanning laser range cameras with applications for machine vision

Brian Lee Chase

To the Graduate Council:

I am submitting herewith a thesis written by Brian Lee Chase entitled "Calibration of scanning laser range cameras with applications for machine vision." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

M. A. Abidi, Major Professor

We have read this thesis and recommend its acceptance:

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Brian Chase entitled "CALIBRATION OF SCANNING LASER RANGE CAMERAS WITH APPLICATIONS FOR MACHINE VISION." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

M. A. Abidi, Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of The Graduate School

# CALIBRATION OF SCANNING LASER RANGE CAMERAS WITH APPLICATIONS FOR MACHINE VISION

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Brian Chase

December 1999

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents, Dennis and Patricia Chase, for their support. I could not have done it without them.

Sincere thanks to Dr. M. A. Abidi for his guidance through the last few years and for his support through financial and technological resources which have moved this work from theory to implementation. Thanks also to Dr. W. L. Green and Dr. M. J. Roberts for being part of my thesis committee, and for their contributions to this work.

I would like to thank the U.S. Department of Energy for its support of this work through the DOE/OST/RTDP's University Research Program in Robotics (Universities of Florida, Michigan, New Mexico, Tennessee, and Texas) under grant DOE–DE–FG02–86NE37968.

Finally, I accept sole responsibility for any errors or weakness which might be discovered in this work hereafter.

# ABSTRACT

Range images differ from conventional reflectance images because they give direct 3–D information about a scene. The last five years have seen a substantial increase in the use of range imaging technology in the areas of robotics, hazardous materials handling, and manufacturing. This has been fostered by a cost reduction of reliable range scanning products, resulting primarily from advanced development of computing resources. In addition, the improved performance of modern range cameras has spurred an interest in new calibrations which take account of their unconventional design.

Calibration implies both modelling and a numerical technique for finding parameters within the model. Researchers often refer to spherical coordinates when modeling range cameras. Spherical coordinates, however, only approximate the behaviour of the cameras. We seek, therefore, a more analytical approach based on analysis of the internal scanning mechanisms of the cameras. This research demonstrates that the Householder matrix [14] is a better tool for modeling these devices.

We develop a general calibration technique which is both accurate and simple to implement. The method proposed here compares target points taken from range images to the known geometry of the target. The calibration is considered complete if the two point sets can be made to match closely in a least squares sense by iteratively modifying model parameters. The literature, fortunately, is replete with numerical algorithms suited to this task. We have selected the simplex algorithm because it is particularly well suited for solving systems with many unknown parameters.

In the course of this research, we implement the proposed calibration. We will

find that the error in the range image data can be reduced from more that 60 mm per point rms to less than 10 mm per point. We consider this result to be a success because analysis of the results shows the residual error of 10 mm is due solely to random noise in the range values, not from calibration. This implies that accuracy is limited only by the quality of the range measuring device inside the camera.

# Contents

# List of Tables

# List of Figures

# 1 INTRODUCTION

Modern autonomous robotics requires accurate environment mapping for navigation and object manipulation or avoidance. Range images are particularly well suited for this task because they give direct 3–D information about the scene. Each pixel in a range image contains information about the range from the camera to the corresponding point in the scene. The reported use of range images has significantly risen in the literature since the introduction of reliable commercial range cameras.

Much of the literature focuses on the interpretation of 3–D data. There is, however, little documentation about range camera calibration. Instead, the raw data are often treated as a surface in cartesian space and analyzed directly for edges and regions of continuity. Usually, if any calibration is done, the spherical coordinates model is used. No one has yet introduced a clear and general method for accurately modeling the complex interactions of the scanning mirrors inside range cameras. This work advocates the use of Householder matrices in the analytic camera model for relating points from the range image to corresponding points in three–space.

We intend to employ the Householder matrix to model a simple one mirror range scanner and next, a two mirror system. Specifically, we need the two mirror model for a commercially available range camera from Perceptron [29], which we will be using in an experiment designed to determine calibration parameters for this camera. We first need to devise a calibration strategy.

Most authors describe the use of targets to implement calibrations. Targets exhibit known geometry, which is exploited during calibration, to find the unknown parameters. We propose to design our calibration strategy around using a set of known target points, which may exist in the scene already. We wish to introduce

a rotation and translation inside the calibration loop which allows the target to be imaged from any position or angle. We hope these features will make calibration attractive for remote operations when human access is limited. Many authors employ a numerical algorithm to find the unknown calibration parameters once the target has been imaged. We favor the simplex algorithm because it works very well with many unknowns.

## 1.1 A Brief Description of Range Cameras

Each pixel in a range image contains information about the range (or distance) from the center of the camera to the corresponding point in the scene. Range, as in a range image, can be obtained by a variety of techniques. In general, systems are classified into one of the following types: triangulation (passive or active) [31], time of flight [16], focusing [32], holography [35, 34], diffraction [7], or Moire techniques [9].

Passive triangulation, or stereo, is perhaps the most familiar method, because humans perceive depth in this fashion. The correspondence problem in stereo matching is ill-defined and limits the number of application we may consider [3].

The correspondence problem of stereo vision can be eliminated by replacing one of the cameras with a moving light source. This is known as active triangulation. A prohibitive factor to the performance of this technique is the video frame rate limitation of the cameras. At NTSC rates, only 30 points can be measured per second.

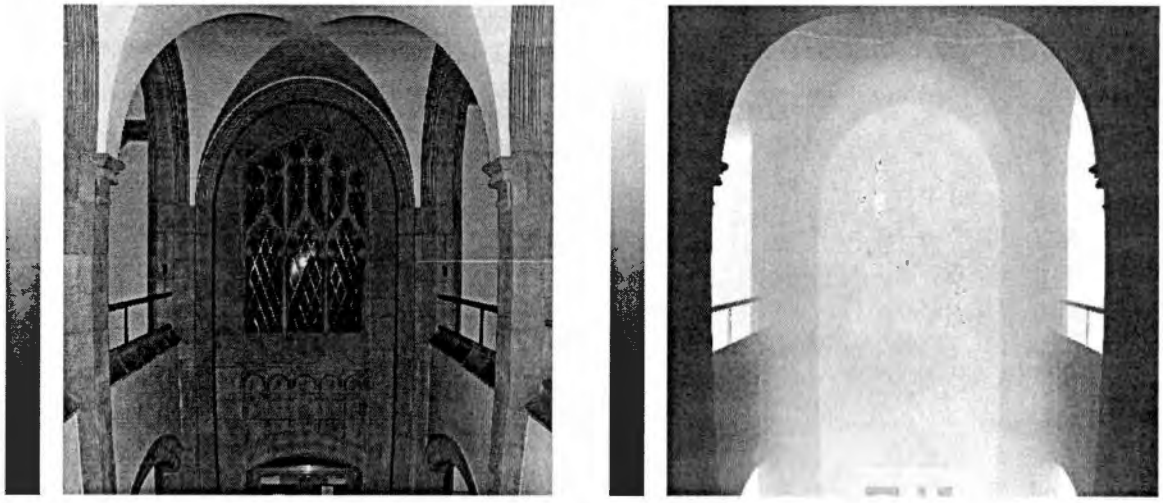
Focusing, holography, diffraction, and Moire techniques are seldom discussed in literature because of fundamental performance limitations. These systems are either too slow or do not have the required accuracy for autonomous robotics.

Most range imaging systems are in the time–of–flight category. The popularity of these systems stems from their high sample speeds and the elimination of the correspondence problem associated with passive triangulation. The time–of–flight concept includes ultrasonic pulse detection, light pulse [15] detection (pulsed laser), and continuous light wave (CW laser) techniques [6]. Time–of–flight systems determine range by measuring the time required for a signal to travel to an object, reflect, and return.

Depth resolution clearly is a function of signal velocity in this type of system. Ultrasonics usually provide very accurate range measurements because of the relatively low velocity of sound. Object resolution is low, however, since the sound cannot be concentrated into a very narrow beam. For this reason, ultrasonic range finding is better suited to autonomous navigation than to range image acquisition.

Laser light is coherent and has an extremely small wavelength in comparison to sound. These properties allow it to be collimated into a tight beam for measuring the range to just one point in the scene. Most laser range cameras use two mirrors mounted on orthogonal axes to scan the laser in a raster pattern. Range values are sampled from this pattern in a grid format to produce the well known range image see Figure 1.

The two laser range cameras used at the University of Tennessee are based on this principle. The Odetics [27] system has a $128 \times 128$ image size and 8 bits of range resolution. The field of view is approximately 60 degrees, both horizontal and vertical. Range images are acquired every 853 ms. The Perceptron [29] system (see Figure 2) is programmable. The view can vary from 15–60 degrees, up to 1024 pixels horizontally, and from 3–72 degrees, up to 2048 pixels vertically. Range is returned as 12 bit integers at the rate of 360,000 samples every second.



Intensity Component Range Component

Figure 1: An example range and intensity image pair obtained using the Perceptron laser range camera.

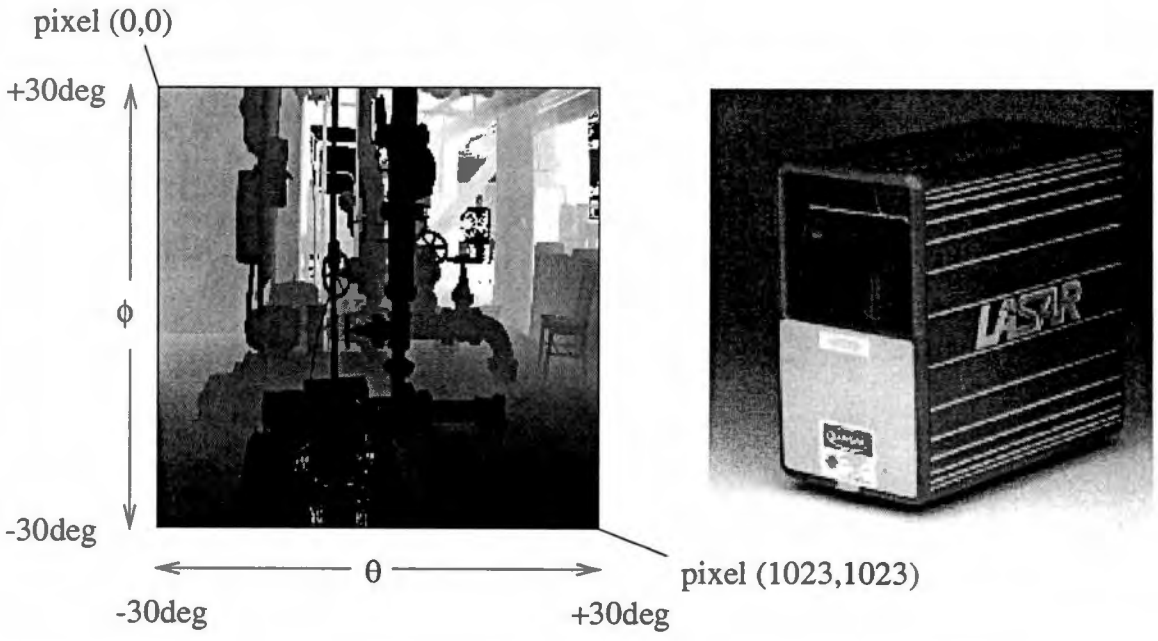



Figure 2: Field of view for a typical range image and the commercially available laser range camera from Perceptron [29] which captured this image.

## 1.2 Problem Definition: From Images to 3–D

The main function of a range camera is to return 3–D data, usually in the form of a range image. Single range images are useful in many machine vision applications. For the task of object recognition, grasping, etc., however, it may be necessary to view the object from many different positions. In fact, these multiple range images must be integrated to find the true 3–D shape of the object [10], [37]. Unprocessed range images, however, cannot be pasted together because they are from different viewpoints. We must therefore consider the topic of camera calibration in order to transform these images into a cartesian coordinate system where the data can be registered through homogeneous transformations.

Points from the range images are typically first transformed to Cartesian coordinates, then integrated to form a 3–D scene or object model. A calibrated camera model is therefore desired to move points from the range image to $(x, y, z)$ space. Sobel [36] recognized that calibration is divided into two main subproblems: modeling the camera, and calibration. In the literature, one often finds reference to the spherical coordinate model [4]. Although this model has known errors concerning the origin of its coordinate system, it has been used extensively without question. This perhaps is because of the complexity of modeling the internal scanning mechanism. Regardless, the introduction of appropriate mathematical tools for modeling and calibration are necessary to move beyond approximation.

## 1.3 Previous Work in the Field of Camera Calibration

The work of Jun Ni [26] addresses the calibration of a laser scanning system for use with computer numerical controlled (CNC) milling machines. Interfacing 3–D input devices such as this with milling machines allows quick digitization of parts for

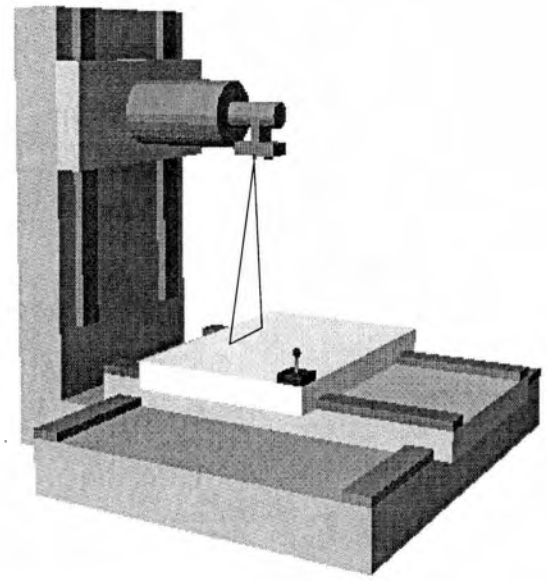


Figure 3: Structure of a multiple–axis CNC laser–stripe sensor. (Adapted from [26].)

inclusion into computer aided design (CAD) software (See Figure 3). The author uses a ball–target–based system for calibration of the 3–D multiple–axis system shown above.

The calibration utilizes four small metallic spheres mounted in the corner of the workcell as a target. The system is calibrated whenever a change in position or orientation occurs in setup. A matrix is developed to take measurements from the skewed coordinate system of the milling machine into Cartesian coordinates. Then, based on the results of the modeling, a constrained optimization calibration algorithm is developed. The four ball targets are measured and the distortion of their geometry from spherical is used to correct transformation parameters in the matrix. The radial distortion of the balls was reduced from an average of 1.44 mm to 0.04 mm over an 18.93 mm range. The significance of this work is that the calibration depends on the geometry of the target, not on careful placement.

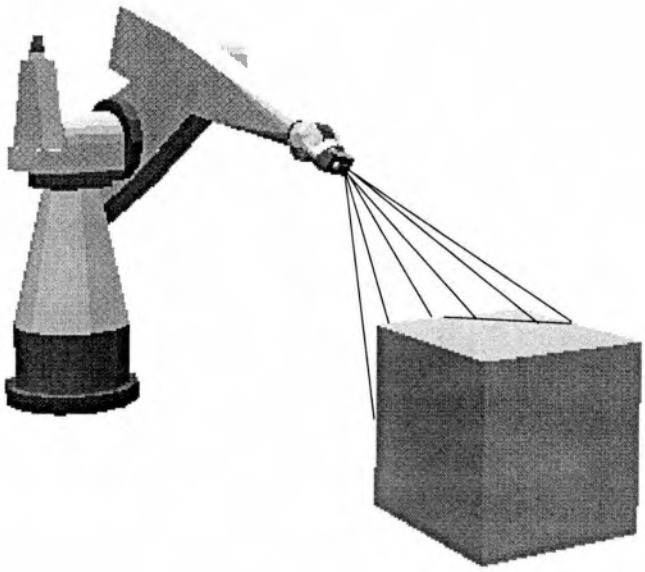


Figure 4: Robot engaged in scanning a scene with a detachable structured–light scanner. (Adapted from [8])

Chen and Kak [8] model and calibrate a robotic 3–D structured light scanner. The configuration is shown in Figure 4. The camera is based on the sheet of light principle and detects only one stripe or line of range values at a time. The illuminated scene points all lie in the same plane as the sheet of light. Scanning is accomplished by moving of the robot. The authors assume a pinhole camera model to simplify the mathematics. This assumption leads to a projection matrix that the authors use for recovering the 3–D location of detected points. However, several variables inside the matrix must be calibrated first. The authors state that four co–planar points provide enough equations to solve for the unknowns in the matrix, but, they claim this approach is impractical for the following reasons:

1. There are always some errors associated with the measurement of the four calibration points. Thus, the co–planarity condition is not completely guaranteed;

2. It is unrealistic to assume the camera can be modeled perfectly by a pin–hole. A pin–hole model is questionable, especially when zoom lenses are used. Additionally, when the pin–hole approximation breaks down, there may be no unique center of projection; and

3. Because of the non-zero thickness of the illumination stripe and other digitization aspects of camera imaging, there will always be some non–zero error associated with the location of the image point corresponding to an object point.

To avoid co–planarity problems, the authors devise a novel technique free from fixed coordinates. The authors propose using a flat trapezoidal object located permanently in the work area as shown in Figure 4. When the scanner projects a stripe intersecting the top edges of this object, the intersection points are guaranteed to be co–planar and to lie on one of the two fixed lines. To avoid the second and third problems, many points are sampled, and a linear least squares problem is solved to give the best fit answer for the matrix. The dimensions and relative orientation of the object are found as a part of the calibration procedure. In fact, the authors compare the recovered dimensions with actual size and find that they agree to within about one percent.

Storjohann and Saltzen [38] describe an imaging model derived for use with the Odetics unit. The Odetics camera transmits an amplitude modulated laser beam. The power of the returned signal is a function of the target reflectance, and the phase shift is a function of the travel time, corresponding to the range. A graphical description of the model is shown in Figure 5. While scanning the laser beam horizontally and vertically, the range and reflectance data are spatially registered in separate images. The scan mechanism consists of a planar mirror, rotating up and down for the vertical deflection, and a rotating polygonal mirror

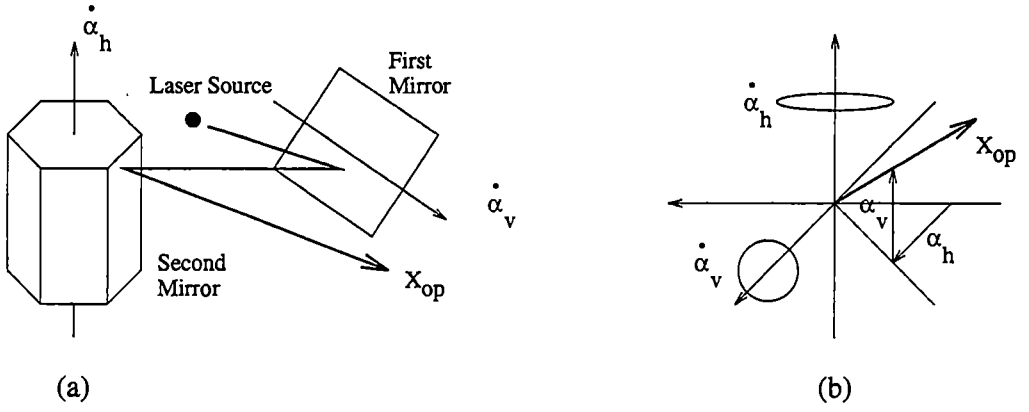



Figure 5: (a) Setup of the scan mirrors and (b) Coordinate system of the camera. (Adapted from [38])

for the horizontal deflection of the laser beam. A comparison is made to the raster scan of a television receiver.

It is appropriate at this point to introduce some terminology associated with the Odetics model shown in Figure 5. Commonly, elevation–azimuth (EA) is a term applied both to a particular class of scanners with equivalent kinematics, and to their associated spherical coordinate system. The notation elevation–azimuth refers to the order in which the vertical and horizontal deflections occur. In this scanner, the vertical deflection occurs first, and the horizontal deflection occurs second. If the order were reversed, it would be called an azimuth–elevation scanner. A more complete survey of scanner configurations was done by Besl and can be found in [28].

Storjohann and Saltzen explain how to experimentally measure the internal parameters (intrinsic calibration) associated with the elevation–azimuth model. They bring out one important and often neglected effect: the coupling of horizontal and vertical motions in the mirrors. During each scan line of the horizontal mirror, the vertical mirror remains in motion, producing a small offset from beginning to end

of the scan line. After all experimental procedures are developed, they produce a transformation is produced for mapping a reflectance image acquired in the range camera to a virtual image corresponding to the standard camera pinhole model. This is for the purpose of evaluating the range camera's external orientation (extrinsic parameters) by conventional techniques. Results indicate that the extrinsic parameters can be determined to within a mean distance of less than one pixel.

According to Baker [4], the transformation from $(r, i, j)$ to $(x, y, z)$ can be considered as a mapping function for the range camera. The form of the mapping can be determined from the design of the scan mechanism. Nominal parameters, such as scan rates can be found in the manufacturer's literature. Calibration of these parameters to fit the individual camera completes the mapping. Baker notes how range cameras are typically calibrated by imaging carefully designed targets and/or physically measuring absolute distances and angles, inferring the corresponding parameter settings. The complexity of such procedures often prohibits re–calibration, especially during remote operations.

As a solution, the author proposes a novel technique related to Chen and Kak's [8]. Calibration is considered automatic since it requires neither carefully designed targets, nor detailed analysis of the camera model. The premise is that the accuracy of a range camera's mapping can be determined by evaluating the planarity of its output when imaging a surface known to be planar. Importantly, the technique is orientation invariant which inspired the development of the calibration strategy in this thesis.

The plane surface provides a simple mathematical model for the acquired points. The algorithm which Baker refers to uses an eigenvalue decomposition (Reister and Morris [30]). When the eigenvalue iteratively converges to zero, the points are considered to be co–planar. If two eigenvalues converge to zero, the points are co–linear. A fast indicator of planarity is the determinate, which numerically equals the product of all eigenvalues.

Baker [4] states that four distinct forms of spherical mappings exist for laser range cameras, depending upon their internal design, and denotes them with a type number $T$.

Type $T = 1$:

$$x = r\sin(\alpha) \tag{1}$$

$$y = r\sin(\beta) \tag{2}$$

$$z = r\sqrt{\cos^2(\alpha) - \sin^2(\beta)} \tag{3}$$

Type $T = 2$:

$$x = r\sin(\alpha) \tag{4}$$

$$y = r\cos(\alpha)\sin(\beta) \tag{5}$$

$$z = r\cos(\alpha)\cos(\beta) \tag{6}$$

Type $T = 3$:

$$x = r\sin(\alpha)\cos(\beta) \tag{7}$$

$$y = r\sin(\beta) \tag{8}$$

$$z = r\cos(\alpha)\cos(\beta) \tag{9}$$

Type $T = 4$:

$$x = r\sin(\alpha)\cos(\beta) \tag{10}$$

$$y = r\cos(\alpha)\sin(\beta) \tag{11}$$

$$z = r\cos(\alpha)\cos(\beta) \tag{12}$$

Where,

$$\alpha = A/2 - (A+B)R/(\delta R - 1) - (A-C)C/(\delta C(\delta R - 1))$$
$$\beta = D/2 - (D+E)C/(\delta C - 1) - (D-F)R/(\delta R(\delta C - 1))$$
$$r = G + r_0.$$

Thus, between any two adjacent rows, the angle $\alpha$ is uniformly interpolated according to the value of the column, and similarly, between any two adjacent columns, the angle $\beta$ is uniformly interpolated according to the value of the row. This is necessary to model the coupling of the vertical and horizontal drives in the scanner so noted earlier by Storjohann and Saltzen [38]. By generalizing these calibration equations with parameter coefficients, they become applicable to any laser range camera which scans a parallelogram grid pattern.

Calibration of this system could be performed either manually, by guessing the parameters, or more effectively by an automatic numerical algorithm. If a single mapping takes only a few seconds, automatic calibration requiring a turnaround time of less than two hours permits just a few thousand guesses. Baker chooses to use a genetic search algorithm [13] because the search space has a significant dimensionality and a few thousand guesses covers less than one trillionth of a percent of the search space.

Despite using the simple spherical model for the camera, the genetic algorithm's search was largely successful; most of the error was removed. Baker postulated that

many of the remaining errors are a result of changing flight path length of the laser beam as the mirrors rotate. We will eliminate this problem in our model by using the Householder matrices. Additionally, the generality of such a technique naturally covers all four variants of scan types mentioned above.

# 2 NON–CONTACT DEPTH MEASUREMENT TECHNIQUES

Time–of–flight techniques determine range by measuring the time required for a signal to travel to an object, reflect, and return. Given the velocity of the signal, $v$, and the travel time of the signal, $\Delta t$, the range, $r$, is given by

$$r = \frac{v\Delta t}{2}. \tag{13}$$

This equation assumes that the signal travels to the object and back on the same path. The lower the velocity of the signal, the more accurate the range resolution will be, based on or because of finite precision in the measurement of time.

Time–of–flight ranging avoids the correspondence problem of passive triangulation and the range ambiguities inherent in active triangulation. The time–of–flight concept includes ultrasonic pulse detection, light pulse detection (laser), and continuous light wave (laser, LED) techniques.

## 2.1 Ultrasonic Pulse Detection

An ultrasonic range finder transmits a chirp over a short period of time and measures the time interval between the transmitted pulse and the returning echo, as shown in Figure 6. An ultrasonic ranging system uses a sound transducer for both the transmitter and the receiver. When transmitting, the transducer generates a short burst of sound, similar to a speaker. When receiving, the transducer acts as a microphone. According to Jarvis [17], several pulses at different frequencies are transmitted to reduce the probability of signal cancellation. Usually, ultrasonics provide very accurate range measurements because of the relatively low velocity of sound. Object resolution is low, however, since the sound cannot be concentrated

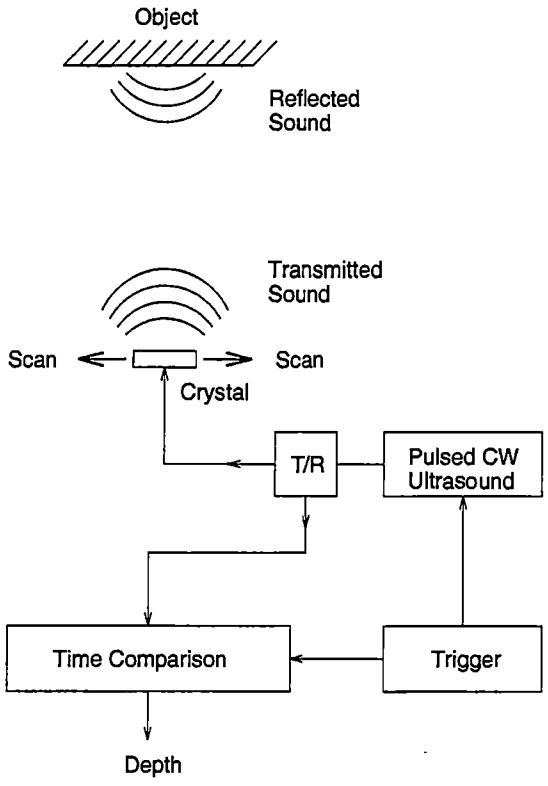



Figure 6: Ultrasonic ranging system. (Adapted from [18], ©John Wiley & Sons)

into a very narrow beam such as a laser. For this reason, ultrasonic range finding is better suited to low resolution tasks such as obstacle avoidance, whereas grasping requires the much higher resolution offered by laser range cameras.

## 2.2 Light Pulse Detection

A pulsed light system uses the same basic idea as the ultrasonic range finder. This technique measures the time it takes for an emitted light pulse to reflect from a surface then return along the same path. For this case, the range is given by

$$r = \frac{c\Delta t}{2}, \tag{14}$$

where $c$ is the speed of light and $\Delta t$ is the pulse travel time. In most cases today, the light originates from a laser.

Jarvis [16] developed a pulsed laser range finder which produces a $64 \times 64$ pixel range image with an accuracy of $\pm 0.25$ cm in a scanning time of 40 seconds. This includes averaging 100 samples per point. The depth–of–field for this system is approximately 1 to 4 meters. Heikkinen *et al.* [12] developed a laser range sensor with a resolution of approximately 20 mm at a range of 3.5 meters (at the system's maximum data rate of 10,000 points per second). Maatta *et al.* [19] implemented a pulsed laser system for profiling hot surface linings which require frequent inspection and repair. The system has a resolution of better than 1 mm at 20 meters, with a sample time of 0.5 seconds. The high resolution is obtained by averaging 4000 samples per point. Takeichi *et al.* [24] report on another pulsed system which was implemented using high speed streak cameras. The system operates with a precision of 1 cm at 100 kilometers. Sample times are limited by the time–of–flight at 100 kilometers. Finally, Golubev *et al.* [1] uses three color light for calculating distances traveled through the atmosphere. The three color method compensates for dispersive effects in the atmosphere which vary the pulse shape and time–of–flight. The result is an increased accuracy of 0.25 mm at 5 km. Sample times are limited to the time–of–flight.

The chirped light technique is an extension of the pulsed light concept. Since ambient noise is always present in the environment and in electronic circuitry, the accuracy of a range measurement heavily depends upon the signal– to– noise ratio or SNR. Therefore, the resolution of a pulsed laser system is limited by the peak power output of the laser. By chirping the signal, more power can be delivered over a longer time period as a chirped wave packet. This concept was first developed for radar. In this type of system, the distance is determined by an analog or digital correlation filter. Moreover, these filters have an averaging effect which reduces the

total amount of noise in the measurement. Collins *et al.* [33] experimented with a chirped laser system, obtaining a resolution of 10 cm at 20 meters range using only a 5 milliwatt laser source. The authors stated that further refinements in the signal processing electronics are expected to increase the resolution.

## 2.3 Continuous Light Wave

A continuous wave laser ranging system is an alternative to the pulsed laser system. Rather than using a short pulse, a continuous beam is emitted and the range is determined as a function of the phase shift between the outgoing and returning waves. This type of system can use either amplitude modulation or frequency modulation.

By amplitude modulating, a laser beam at a frequency $f_{AM} = c/\lambda_{AM}$ and measuring the phase difference $\Delta\phi$, defined as:

$$\Delta\phi = 2\pi f_{AM}\Delta t = 4\pi f_{AM}\frac{r}{c}, \tag{15}$$

between the outgoing signal and the returning signal, the range can be obtained from the following equation [6]:

$$r(\Delta\phi) = \frac{c}{4\pi f_{AM}}\Delta\phi = \frac{\lambda_{AM}}{4\pi}\Delta\phi. \tag{16}$$

This result is good for range values which are less than one–half of the wavelength of the modulated signal. Any phase measurement, $\Delta\phi$, also corresponds to $\Delta\phi + n2\pi$. Therefore, $\Delta\phi$ is actually determined to within an integer multiple of $2\pi$. This results in an ambiguity interval on $r$ of $c/2f_{AM}$, or $\lambda_{AM}/2$. If there are no ambiguity resolving mechanisms, the depth field must be constrained to less than one–half $\lambda_{AM}/2$ for Equation 16 to be valid. If higher modulating frequencies are used, better depth resolution is gained. If lower modulating frequencies are used, larger depth

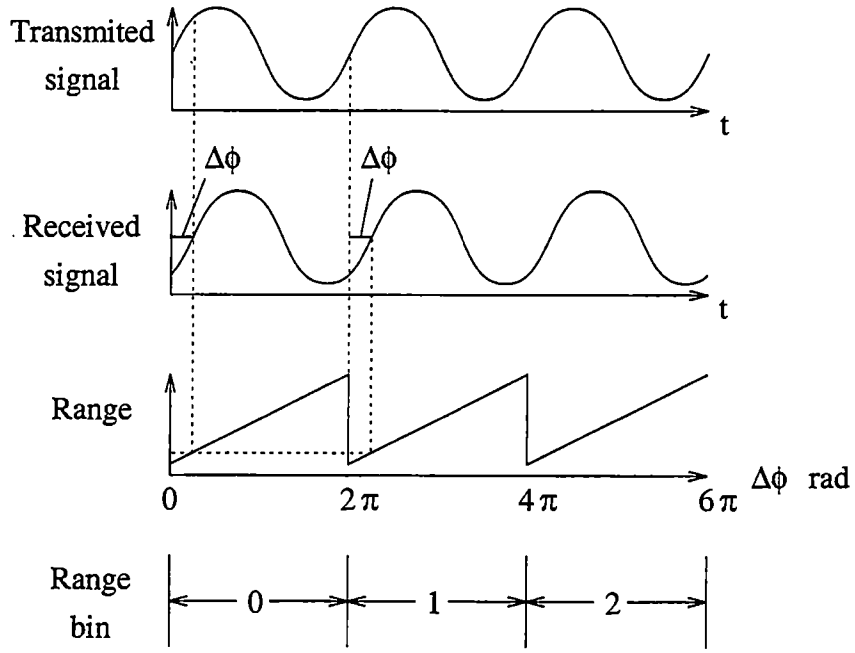



Figure 7: Range as a function of phase for an amplitude modulated continuous beam laser range finder. (Adapted from [27], ©Odetics)

field is obtained. Figure 7 illustrates the concept of amplitude modulation (AM) laser ranging. Seta and O'ishi [20] utilize the intermode beat of a He–Ne laser to obtain a continuous wave AM laser source for range measurements. This produces an AM modulation that is extremely stable, with nearly 100 percent depth of modulation and which occurs naturally in many laser diodes. At distances between 10 meters and 1300 meters, the resolution is within 0.1 mm. Sample time is not given. Deng and Wang [22] report nanometer scale resolution from an AM system they developed. Their system uses optical coherence domain reflectometry over a very short distance to make measurements at the 20 nanometer scale. Detection speed is limited by the peizoelectric crystal response of 1 kHz. Unfortunately, the method requires a mirror at the target, which precludes the possibility of non–cooperative target measurement.

In a frequency modulated (FM) system, if the optical frequency of the transmit-

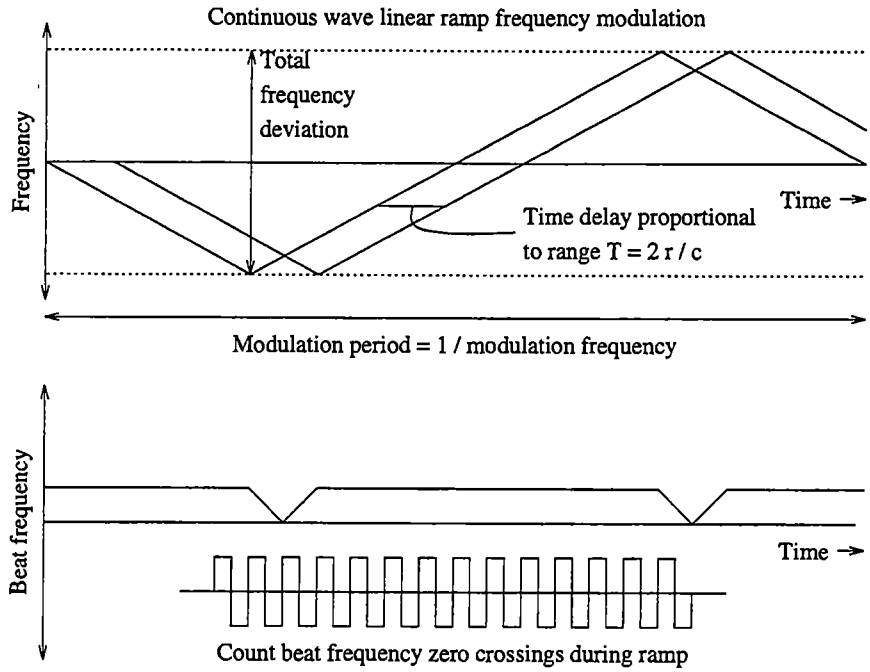



Figure 8: Frequency modulated continuous beam laser ranging principles. (Adapted from [6], ©Springer–Verlag)

ted signal is varied linearly between $(v - \Delta v/2)$ and $(v + \Delta v/2)$ during the period $1/f_m$, a reference signal can be coherently mixed with the returning signal to create a beat frequency, $f_b$, that is a function of range. Figure 8 illustrates this principle. The range is determined from the following equation [6]:

$$r(f_b) = \frac{cf_b}{4f_m\Delta v}. \tag{17}$$

The absence of phase measurement in this system precludes the possibility of an ambiguity interval. Beheim and Fritsch [5] developed an FM laser ranging system having a 29.3 points/second data acquisition rate. They gave experimental results showing the possibility of obtaining sub–centimeter resolution over a range of 1.5 meters.

One of the major noise sources in coherent heterodyne laser radar systems, such as pulse detection and FM, is a material property called speckle. Speckle effects derive from surface roughness at the microscopic level. When illuminated by a coherent wavefront of light, the reflections tend to be out of phase, giving rise to random constructive and destructive interference, which result in a specular–looking reflection. Leader [23] asserts that the effects of speckle can be minimized if an appropriate model is used. Green and Shapiro [11] actually implement such a system using a probabilistic speckle model. Results show good agreement between the theoretical predictions and the experimental data.

# 3 FROM THE RANGE IMAGE TO CARTESIAN COORDINATES: DEVELOPMENT OF MATHEMATICAL MODELS

An analytic camera model relating 3-D coordinates (x, y, z) to the range image coordinates (i, j, range) is desired. The usual assumption of spherical coordinates is not considered except as a special case. An alternate model is preferred, based directly on the internal scanning mechanism of the camera. Since most range cameras employ two rotating mirrors for scanning, the mirror reflection transform (Householder matrix) will form the basis of the optical kinematics study. Both pointing errors and static range measurement errors will be considered in the calibration of the camera model.

Classically, research in this field makes heavy use of geometry and triangular relationships to analyze the operation of scanner designs on a per–unit basis. The complexity of such an analysis often motivates specialized designs exhibiting particularly nice mathematics. Often, in the presence of small mechanical deviations in the device, the models can be shown to behave badly. A more direct method involving the popular Householder matrix is presented here. This type of operation enjoys familiarity in the mathematical community because of the extensive literature concerning the Householder [14] reflection matrix. The matrix was originally devised by Householder to speed machine inversion and decomposition of matrices. The objective was to replace the more costly rotation-matrix operations involving sine and cosine functions. In contrast, Householder matrix operations require only multiplication and addition.

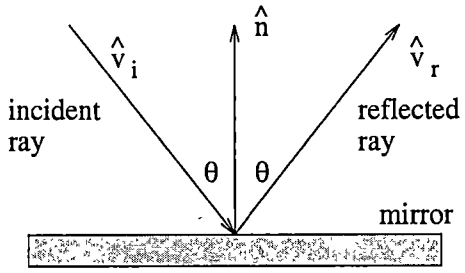



Figure 9: Snell's law. (Adapted from [21])

## 3.1 Reflection: Snell's Law

Let us begin by analyzing the simple reflection of rays with planar mirrors. Snell's law states that the reflected ray lies in the same plane as the incident ray and the normal to the mirror. Moreover, the angle between the incident and the normal is the same as the one between the reflection and the normal (illustrated in Figure 9). The physical interpretation is that photons are bouncing off the mirror surface. Conservation of momentum implies that normal component of motion is reversed, while the parallel component remains unchanged (Equation 18).

$$\vec{v_r} = \vec{v_i} - 2(\vec{v_i} \cdot \hat{n})\hat{n} \tag{18}$$

Consider an operator, ref($\hat{n}$), a $3 \times 3$ matrix that performs this reflection when it is multiplied by a vector.

$$\begin{aligned}
\vec{v_r} &= \vec{v_i} - 2\hat{n}(\hat{n} \cdot \vec{v_i}) \\
\vec{v_r} &= \vec{v_i} - 2\hat{n}(\hat{n}^T \vec{v_i}) \\
\vec{v_r} &= \vec{v_i} - 2\hat{n}\hat{n}^T \vec{v_i} \\
\vec{v_r} &= (I - 2\hat{n}\hat{n}^T)\vec{v_i} && (19) \\
\text{ref}(\hat{n}) &= I - 2\hat{n}\hat{n}^T && (20)
\end{aligned}$$

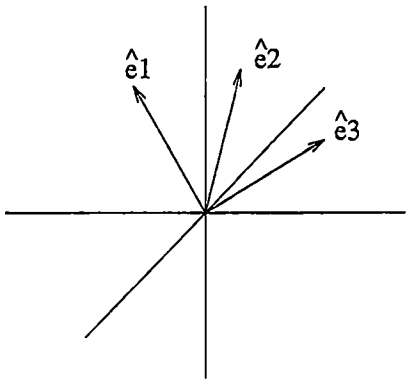



Figure 10: A three dimensional orthonormal basis

## 3.2 Formation of the Householder Matrix by Projection Matrices

The matrix $ref(\hat{n})$, also known as the Householder matrix, is formed by inverting one orthonormal basis in a vector space. Projection matrices can be used to show how this works. A projection matrix is an essential tool for projecting a higher dimensional space onto a lower dimensional space, such as a plane or line. The $3 \times 3$ projection matrix, $\hat{n}\hat{n}^T$ is the outer product of $\hat{n}$ with itself, as opposed to the more familiar inner product, written as $\hat{n}^T\hat{n}$. Notice that the inner product is a $1 \times 1$ matrix, or a scalar value, while the outer product is a $3 \times 3$ matrix.

Consider the product of $\hat{n}\hat{n}^T$ with a vector: $(\hat{n}\hat{n}^T)\vec{v}$. This can also be written as $\hat{n}(\hat{n}^T\vec{v})$, which is the projection of $\vec{v}$ onto the line spanned by $\hat{n}$. We next show how these matrices are related to the Householder matrix.

Take any three dimensional orthonormal basis as shown in Figure 10.

$$B = [\hat{e_1}, \hat{e_2}, \hat{e_3}]\,. \tag{21}$$

The outer product of $B$ within itself is the identity matrix, since $B$ spans the entire three dimensional space. Projecting a vector onto $B$ does not alter the original vector.

$$BB^T = \hat{e_1}\hat{e_1}^T + \hat{e_2}\hat{e_2}^T + \hat{e_3}\hat{e_3}^T = I_{3\times 3}. \tag{22}$$

Reflecting one of the bases gives:

$$\text{ref} = \hat{e_1}\hat{e_1}^T + \hat{e_2}\hat{e_2}^T - \hat{e_3}\hat{e_3}^T = I_{3\times 3} - 2\hat{e_3}\hat{e_3}^T. \tag{23}$$

Ref is a reflection operator that maps all vectors to a new space where one dimension has been reflected through the origin. In general, the special form $I_{3\times 3} - 2\hat{e_3}\hat{e_3}^T$ is known as a Householder matrix.

## 3.3 Modeling a One Mirror Scanner

Figure 11(a) illustrates a 2–D version of a laser range scanner. In this figure, the laser beam travels from the source to the mirror, is reflected through some angle, and hits the target. The total range measured is the sum of the lengths of path $a$ plus path $b$. Given a mirror angle and the measured range distance, we desire to calculate the position of the detected point in space.

The mirror seems, at first, to be an encumberance to calculating the detected point because of its interaction with the laser beam. In the absence of a scanning mirror, only points along the line of the laser beam can be detected. We calculate that the detected point equals the laser origin plus a vector representing a length of laser beam given by the measured range. Significantly, the presence of a mirror changes the geometry, but in an extremely predictable way. If we call this point, lying on the line of the laser beam, the virtual point, then the real point is related to the virtual point by a mirror transform. Therefore, calculating the real detected point consists of two simple procedures. First, calculate the virtual point. Second,

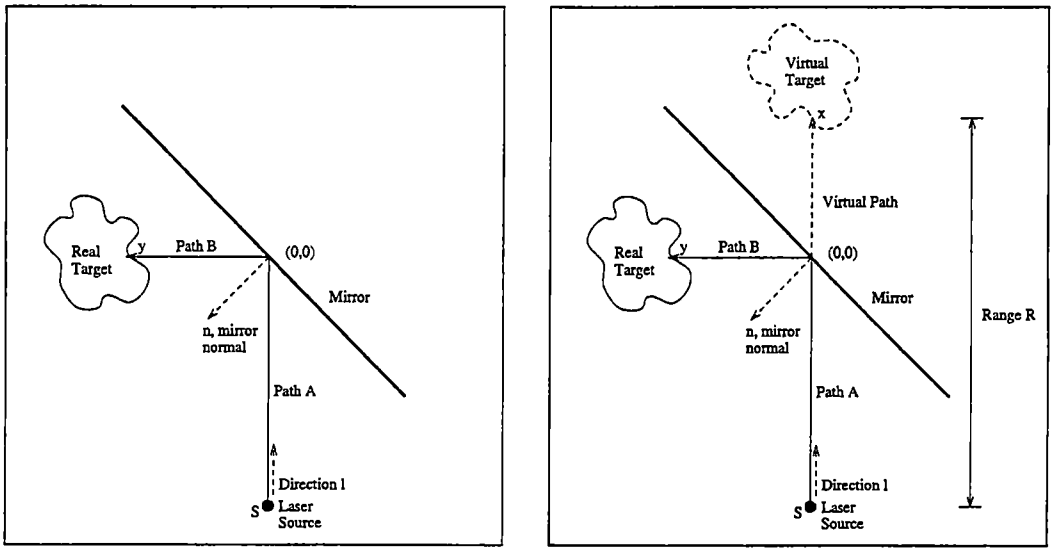



Figure 11: (a) Range from a simple laser scanner configuration. (b) Using the virtual image concept.

multiply the virtual point by the appropriate mirror transformation to find to real point.

Figure 11(b) shows the geometry associated with the virtual point concept The detected range point $x$ in the virtual scene is equal to $S + lR$, where $S$ is the laser origin, $l$ is a direction vector for the laser beam that is one range increment in length, and $R$ is the range value (scalar). The Householder matrix moves the virtual points to real points and *vice versa* as in Equation 25.

$$y = (I - 2\hat{n}\hat{n}^T)x \tag{24}$$

$$y = (I - 2\hat{n}\hat{n}^T)(S + lR) \tag{25}$$

These equations are valid for all cases where the mirror passes through the origin, since the Householder matrix produces reflection across the origin. If the mirror does not pass through the origin, the reflection will have to be done under a translated coordinate system. Some stationary point on the mirror (a point on

the mirror axis) should be subtracted from the coordinates before reflection, then added back afterwards. Reflection through two or more mirrors is done by chaining the Householder matrices end to end. The 2–D concept readily extends to 3–D.

## 3.4 Multiple Mirror Systems

Range cameras typically scan in a rasterized grid pattern. Generating the pattern requires one mirror for horizontal deflection and a second mirror for vertical deflection. The two mirrors can be arranged in a variety of ways, as noted by Besl [28]. For instance, the Perceptron [29] unit is an azimuth–elevation scanner, whereas the Odetics unit is an elevation–azimuth design. The azimuth–elevation variant is sometimes referred to simply as an azimuth scanner. Likewise, the elevation-azimuth design can be called an elevation scanner.

### 3.4.1 Kinematics of the Azimuth Scanner

The azimuth scanner is a generic name for a class of laser range cameras with equivalent kinematics. In this scanner, the laser beam first undergoes the azimuth reflection, and secondly, the elevation reflection. The Perceptron unit falls into this category. The Perceptron unit is a 2–D scanning laser range camera employing a polygonal azimuth mirror and a flat nodding elevation mirror. The mirrors move as shown in Figure 12. The spinning polygonal azimuth mirror creates horizontal deflection, while the flat elevation mirror rotates up and down for the vertical deflection.

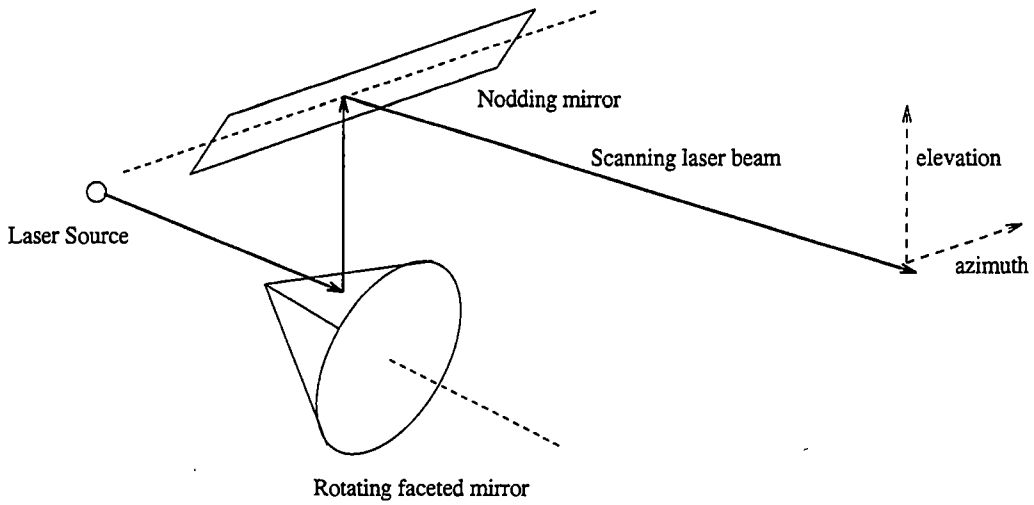



Figure 12: Azimuth Scanner

### 3.4.2 Kinematics of the Elevation Scanner

The Odetics unit is an elevation–azimuth scanner. In this configuration, the laser beam reflects off the elevation mirror first and the azimuth mirror second (see Figure 13). Thus, while AE and EA scanners appear to be functionally equivalent, their associated spherical coordinates are not as noted by Baker [4]. Furthermore, subtle differences exist in the design of the scanning mirrors, which introduce changes in the coordinate system. In addition to the deliberate design parameters, each scanner contains a unique set of calibration parameters, further modifying the coordinate system in a non–intuitive way. For instance, mirrors may be mounted slightly off axis. The focus of the next chapter is to develop a mathematical model general enough to cover all possible design variations in addition to calibration parameters.

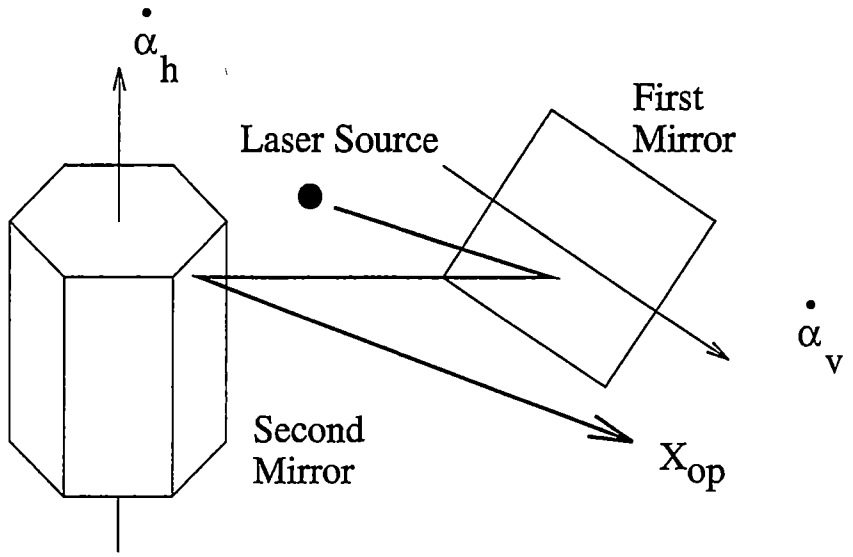



Figure 13: Elevation Scanner (Adapted from [38])

## 3.5 Modeling a Two Mirror Scanner

Modeling the two–mirror system involves chaining the Householder matrices. Previously we saw how to model a one–mirror system by using 2–D vectors. The following examples require 3–D vectors which can be manipulated in exactly the same way. The matrix notation is, in fact, exactly the same, no matter how many dimensions exist.

The first example of a 3–D case is the rotating faceted azimuth mirror in the Perceptron unit (Figure 14). A special frame of reference known as the body coordinates makes the calculations extremely convenient. By aligning the Z axis with the mirror's natural rotation axis, the normal vector to the mirror facet becomes a spherical coordinate with radius = 1, facet angle = $\alpha$, and mirror rotation = $\theta$ As the mirror rotates, the facet normal sweeps out lines of latitude. Each point on the latitude circle corresponds to a unique Householder matrix. Thus, by relating pixel coordinates in the range image to mirror rotation angles, it is possible to deduce the

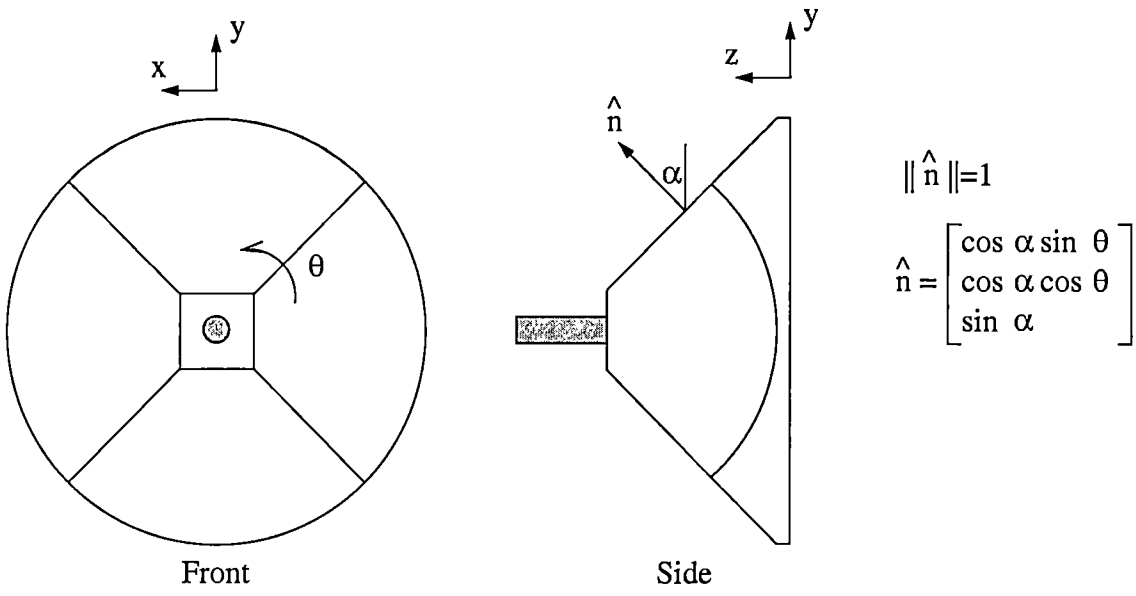



Figure 14: Characterizing the azimuth mirror normal.

appropriate Householder transformations for each point in the range image. Since the facets on the mirror are identical, every horizontal scan line has the same set of Householder matrices, which may lead to a speed up in computation time for complete images.

Consider the elevation mirror from the Perceptron unit. The axis of rotation is offset in the z and y coordinates to location $C = (0, y_o, z_0)$ (as illustrated in Figure 15).

It is now possible to complete the range scanner model. Derive the expressions for the two Householder matrices corresponding to the azimuth mirror and the elevation mirror.

$$H_1 = I - 2\hat{n}_1\hat{n}_1^T \tag{26}$$

$$H_2 = I - 2\hat{n}_2\hat{n}_2^T \tag{27}$$

By assembling these into a larger expression based on the same principal as Equa-

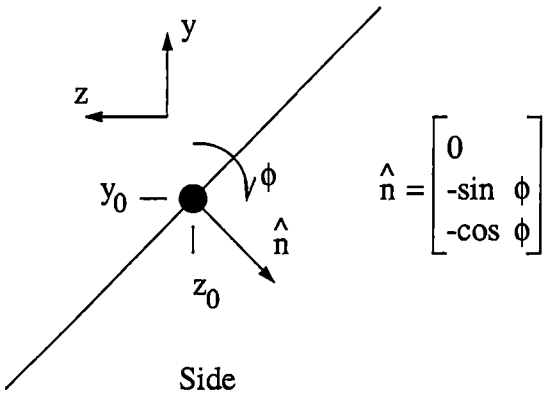



Figure 15: Characterizing the elevation mirror normal.

tion 25 we have the following:

$$P = H_2(H_1(S + lR)). \tag{28}$$

Furthermore, we may compensate for the offset between the two mirrors, $C$, by translating the coordinates before and after the second mirror reflection, yielding the more accurate equation:

$$P = H_2(H_1(S + lR) - C) + C. \tag{29}$$

In Equation 29, $P$ is the true 3–D coordinate of the point in the range image, $C$ is the elevation mirror offset, $H_1$ depends on the azimuth mirror angle, and $H_2$ depends on the elevation mirror angle.

## 3.6 From Range Images to Cartesian Coordinates: A Complete Model

Now that the theory of how to transform a single range measurement back into 3–D coordinates has been laid out, consider the more pressing question of how to transform the entire range image back into the 3–D geometry which it represents.

As each pixel coordinate $(i, j)$ corresponds to unique mirror orientations $(\theta, \phi)$ inside the camera, many Householder matrices must be constructed; in fact, two for each pixel coordinate. Therefore, knowing how to construct the two required Householder matrices from $(\theta, \phi)$, we must find the relationship between $(\theta, \phi)$ and $(i, j)$.

Because of the Householder matrices' dependence on the mirror normals, we must establish the relationship between the mirror normals and the range image coordinates $(r, i, j)$. The primary concern here is the sampling scheme by which range values are taken and the intervals at which they are taken. The following assumptions have been verified by consultation with the manufacturer and by experience.

1. Both mirrors rotate at a constant rate over the entire scan.
2. Samples are taken at regular intervals during each horizontal scan, beginning with the first pixel $(i = 0)$ and ending with the last pixel $(i = i_{max})$.
3. Horizontal line scans are also taken at regular intervals.

These three simple assumptions lead to the following conclusions:

1. The azimuth mirror angle is linearly linked to $i$, the column number of the pixel.
2. The elevation mirror angle at the beginning of each line scan is linearly linked to $j$, the row number of line.
3. The elevation mirror continues to rotate during each line scan so it is different by some angle at the end of each line compared to the beginning,

all of which may be expressed by the following relation:

$$\begin{bmatrix} \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \alpha & 0 & \theta_0 \\ \gamma & \beta & \phi_0 \end{bmatrix} \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} . \tag{30}$$

Where alpha is the angle increment per pixel in the horizontal direction, beta is the angle increment per pixel in the vertical direction, and gamma is a small value corresponding to the small change in the elevation mirror angle across the scan line. As indicated in the assumptions, both mirrors move continuously during a scan. If the elevation mirror angle is constant across a scan line, gamma is zero.

## 3.7 Resolving the Issue of Spherical Coordinates

The use of spherical coordinates dominates the literature concerning laser range camera modeling. In light of the new Householder model, we question the relationship of the Householder model to the spherical coordinates models. It is possible that spherical coordinates models can be derived from the Householder model under certain assumptions. In this case, we want to quantify the errors resulting from these assumptions.

Both models are based on two angles $(\theta, \phi)$ (mirror normals in the Householder model and laser angles in spherical model), which are calculated from the image coordinates $(i, j)$ using Equation 30, only with different factors in the matrix. We shall later see what kind of relationship exists between the two different sets of angles.

Consider the azimuth elevation spherical coordinates (type T=2) according to Baker [4] in Chapter 1. Refer to Figure 16 also.

$$x = r\sin(\theta) \tag{31}$$

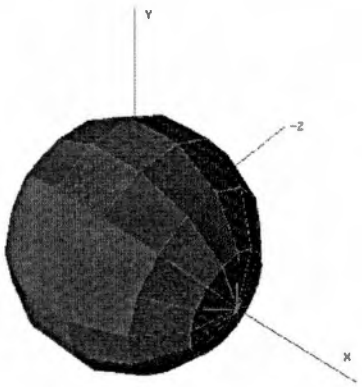



Figure 16: Azimuth elevation spherical coordinates are aligned to the x axis.

$$y = r\cos(\theta)\sin(\phi) \tag{32}$$

$$z = -r\cos(\theta)\cos(\phi), \tag{33}$$

$$\text{where } r = (\text{range})(\frac{\text{mm}}{\text{increment}}) + \text{offset}.$$

Although the spherical form shown above does not look very much like the Householder model Equation 29, an important link exists. Equation 29 is repeated below for convenience.

$$P = H_2(H_1(S + lR) - C) + C.$$

Where $P$ is the true 3–D coordinate of the point in the range image, $C$ is the elevation mirror offset, and $H_1$ and $H_2$ depend on the azimuth and elevation mirror angles $\hat{n}_1$ and $\hat{n}_2$.

$$H_1 = I - 2\hat{n}_1\hat{n}_1^T \tag{34}$$

$$H_2 = I - 2\hat{n}_2\hat{n}_2^T \tag{35}$$

From Figures 14 and 15, the mirror normals $\hat{n}_1$ and $\hat{n}_2$ are defined as follows:

$$\hat{n}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta) \\ \frac{1}{\sqrt{2}}\cos(\theta) \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{36}$$

$$\hat{n}_2 = \begin{bmatrix} 0 \\ -\sin(\phi) \\ -\cos(\phi) \end{bmatrix} \tag{37}$$

We can obtain the spherical form from the Householder model under certain assumptions. These are:

1. The scene is large in relation to the size of the scanner mechanism so the elevation mirror offset, $C$, can be ignored;
2. The laser source is located on the negative $z$ axis of the camera; and
3. The laser source points only in the negative $z$ direction of the camera.

In reality one cannot assume these conditions will hold. For instance, autonomous navigation tasks meet condition 1 while grasping tasks do not. This illustrates why the Householder model is more flexible. Expanding some of the terms in Equation 29 gives

$$(S + lR) = \begin{bmatrix} 0 \\ 0 \\ -\text{offset} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -\frac{\text{mm}}{\text{increment}} \end{bmatrix} \text{range} = \begin{bmatrix} 0 \\ 0 \\ -r \end{bmatrix}. \tag{38}$$

From Equations 34 and 36 we have

$$H_1 = I - 2\hat{n}_1\hat{n}_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta) \\ \frac{1}{\sqrt{2}}\cos(\theta) \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta) & \frac{1}{\sqrt{2}}\cos(\theta) & \frac{1}{\sqrt{2}} \end{bmatrix}. \tag{39}$$

or

$$H_1 = \begin{bmatrix} 1 - \sin^2(\theta) & -\sin(\theta)\cos(\theta) & -\sin(\theta) \\ -\sin(\theta)\cos(\theta) & 1 - \cos^2(\theta) & -\cos(\theta) \\ -\sin(\theta) & -\cos(\theta) & 0 \end{bmatrix}. \tag{40}$$

Because of assumption 1, $C = 0$ and the only remaining factor in 29 is $H_2$. From Equations 35 and 37 we have

$$H_2 = I - 2\hat{n}_2\hat{n}_2^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} 0 \\ -\sin(\phi) \\ -\cos(\phi) \end{bmatrix} \begin{bmatrix} 0 & -\sin(\phi) & -\cos(\phi) \end{bmatrix}, \quad (41)$$

or,

$$H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - 2\sin^2(\phi) & -2\sin(\phi)\cos(\phi) \\ 0 & -2\sin(\phi)\cos(\phi) & 1 - 2\cos^2(\phi) \end{bmatrix}. \quad (42)$$

The terms in 29 can be combined.

$$H_2(H_1(S + lR)) =$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - 2\sin^2(\phi) & -2\sin(\phi)\cos(\phi) \\ 0 & -2\sin(\phi)\cos(\phi) & 1 - 2\cos^2(\phi) \end{bmatrix} \begin{bmatrix} 1 - \sin^2(\theta) & -\sin(\theta)\cos(\theta) & -\sin(\theta) \\ -\sin(\theta)\cos(\theta) & 1 - \cos^2(\theta) & -\cos(\theta) \\ -\sin(\theta) & -\cos(\theta) & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -r \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - 2\sin^2(\phi) & -2\sin(\phi)\cos(\phi) \\ 0 & -2\sin(\phi)\cos(\phi) & 1 - 2\cos^2(\phi) \end{bmatrix} \begin{bmatrix} r\sin(\theta) \\ r\cos(\theta) \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} r\sin(\theta) \\ r(1 - 2\sin^2(\phi))\cos(\theta) \\ -r(2\sin(\phi)\cos(\phi))\cos(\theta) \end{bmatrix} = \begin{bmatrix} r\sin(\theta) \\ r\cos(2\phi)\cos(\theta) \\ -r\sin(2\phi)\cos(\theta) \end{bmatrix} \quad (43)$$

Notice the strong resemblance to Equations 31–33. If a substitution of $-\Phi/2$ is made for $\phi - \pi/4$, an equivalence occurs.

$$\cos(2\phi) = \cos(2(\phi - \pi/4) + \pi/2) = -\sin(2(\phi - \pi/4)) = -\sin(-\Phi) = \sin(\Phi)$$

$$\sin(2\phi) = \sin(2(\phi - \pi/4) + \pi/2) = \cos(2(\phi - \pi/4)) = \cos(-\Phi) = \cos(\Phi)$$

Equation 43 now becomes

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r\sin(\theta) \\ r\sin(\Phi)\cos(\theta) \\ -r\cos(\Phi)\cos(\theta) \end{bmatrix}. \tag{44}$$

The linear relation between $\phi$ and $\Phi$ is accounted for in matrix 30. Thus, $\phi$ is mirror angle while $\Phi$ is laser angle. The difference is because of the configuration of the mirror and laser, which causes the reflected beam to go through twice the angle of the mirror rotation.

This analysis demonstrates that the two models are reconciled by ignoring the mirror center offsets, which is appropriate for large scale scenery, and making other plausible assumptions. The spherical coordinates model is thus seen as special case of the Householder model.

# 4 CALIBRATION HARDWARE AND SOFTWARE WITH AUTOMATIC TARGET LOCATION

In the following sections, we design and implement hardware and software for calibration. In the first section, we design and build a target and mobile range workstation. In the second section, we develop software to locate the target and its geometric features, and to calibrate the camera parameters based on these features. Figure 17 shows a simple block diagram of the software flow during calibration.

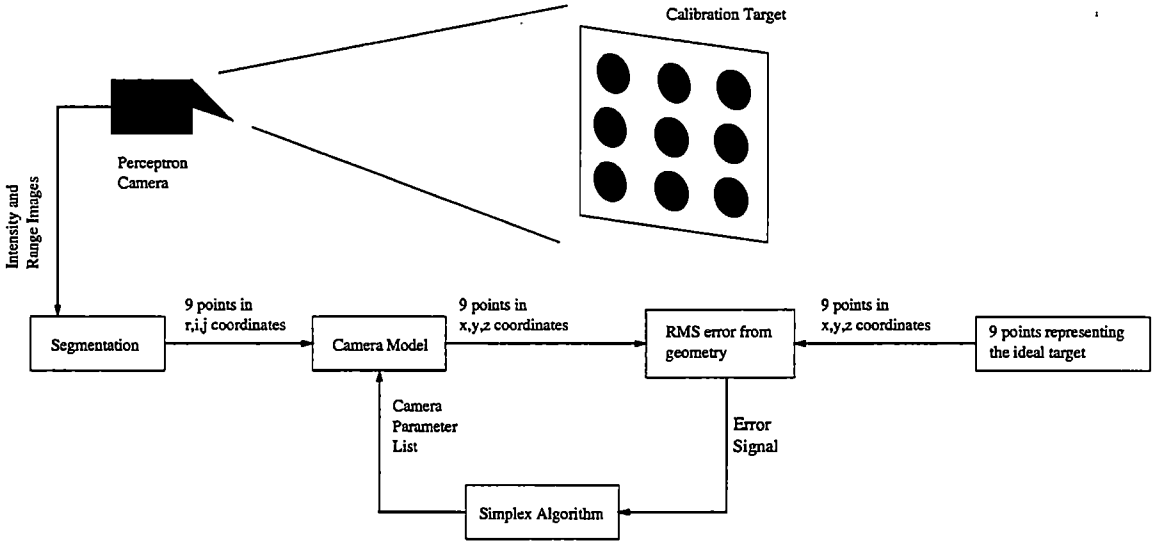



Figure 17: A simple block diagram of the software flow during calibration.

## 4.1 Developments in the Associated Hardware for Calibration

This section describes the practical considerations associated with the target and the range camera. The wide focus of this research is the deployment of a mobile type of mapping system. Portability is therefore considered paramount. This constraint

implies light construction for the target. Portability also prompts the design of a wheeled workstation containing all the associated electronics for the range camera, plus an X–windows workstation for analyzing images and calculating calibration.

### 4.1.1 Target Design and Mounting

In the literature, we find references to the use of targets for calibration [26], [4], [38], etc. Although targets vary greatly in design, we find that all exhibit a known geometry which is exploited in the calibration process. Sometimes the targets are spheres, rhomboids, or tetrahedrons. Other times, the geometric features are lines or spots printed on a surface.

Because our calibration technique does not rely on a specific target configuration, we are free to choose among many designs. Because of the wide availability of accurate, large scale printing equipment, we choose a printed target. Some disadvantages that must be considered include stretching, since the target is often mounted vertically, and expansion caused by humidity.

In order to keep the flexible printed target pressed as flat as possible and to reduce stretching of the target from uneven support or humidity, design of a suitable stand for the target must be at least as important as the design of the target itself. The stand must protect the target from physical damage. One approach is a frame containing two plexiglass sheets to press the target flat and protect the stand. Our design incorporates the plexiglass sheets in a modular stand which may be partially disassembled for portability. Figure 18 shows the original concept design and the final result.

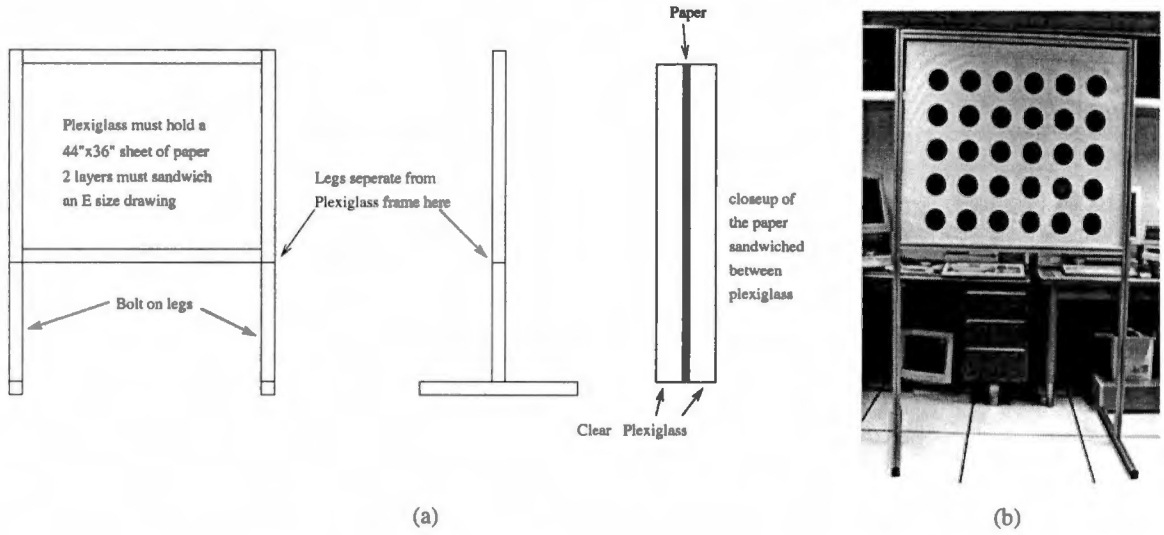


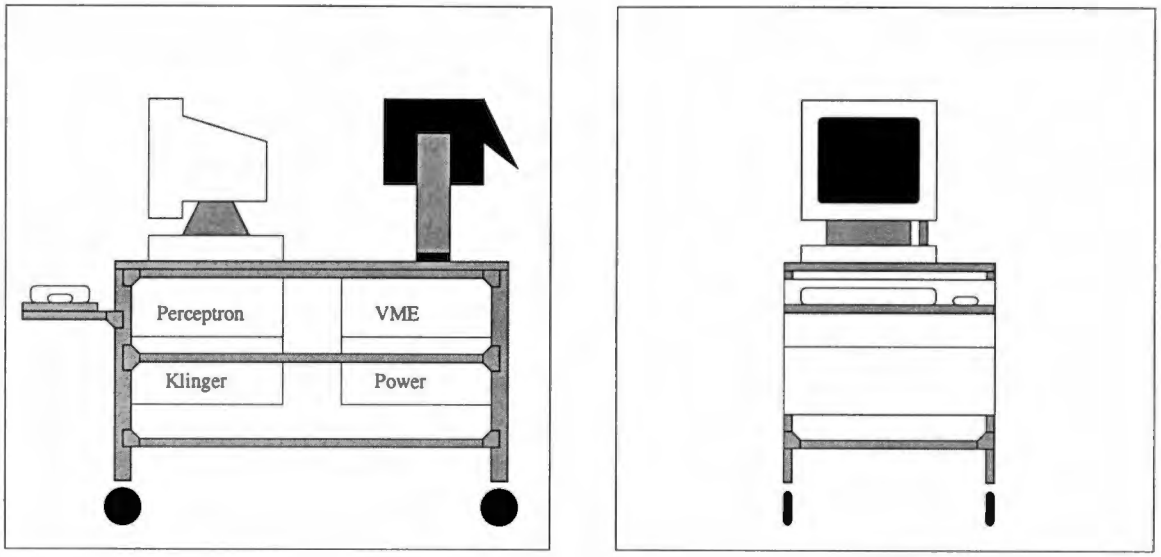

Figure 18: Target stand (a) original concept design and (b) final product.

### 4.1.2 Construction of a Mobile Platform for Camera and Support Hardware

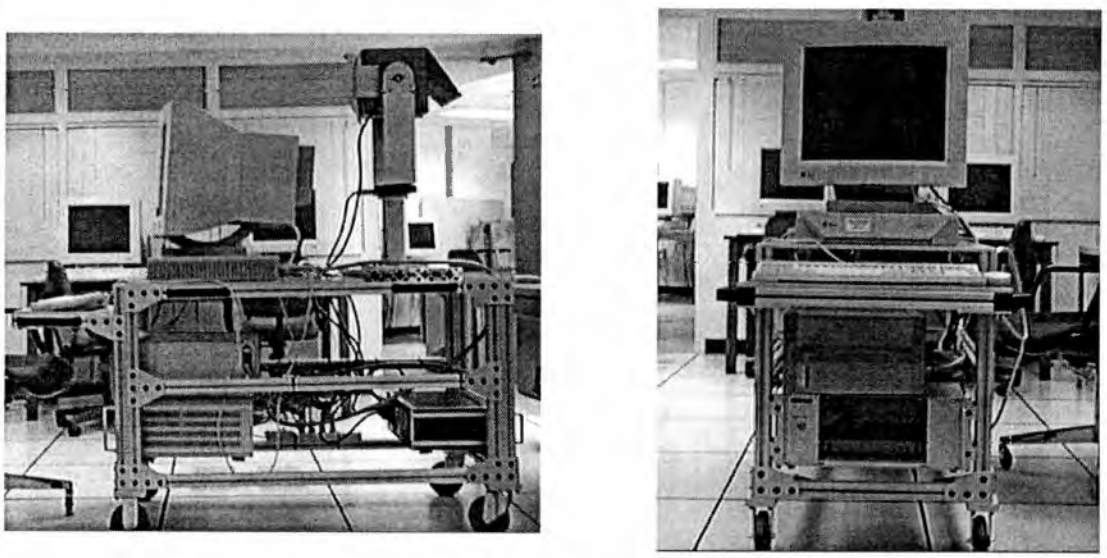
We design and build a mobile four wheeled cart for all the range camera equipment, as well as an X–Windows workstation for displaying, processing, and storing the images (see Figure 19). Several electronics boxes mount underneath the cart in a rack configuration, while the top surface is dedicated to the workstation and the camera.

### 4.1.3 First Target Images

Because the 500 milliwatt laser source in the Perceptron camera head is powerful enough to easily damage the retina at the close range of less than a meter, several measures are taken to assure safety during the image acquisition process. Laser safety goggles, opaque to the damaging radiation, must be used during operation of the camera at all times when the laser is energized. Additionally, several safety

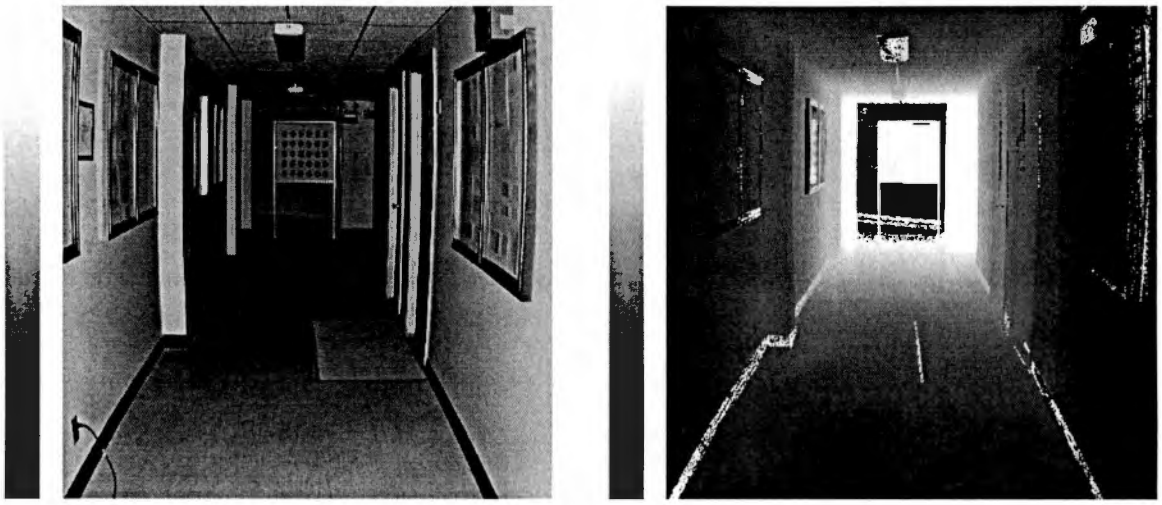



(a)



(b)

Figure 19: Design of the UTK mobile range scanner assembly (a) original concept design and (b) final product.



Intensity Component Range Component

Figure 20: Sample target image acquired from the Perceptron laser range camera. The high quality of these images exceed conventional photography where depth of field is limited by the lens.

interlocks, including an on-off key switch, can lock out the laser power supply.

A group of images are taken with the target placed at progressively greater distances. The images will be processed, then be used to test the calibration algorithm at various distances. These images are 1024x1024 resolution with 12 bits accuracy in both the intensity and the range images. A sample image is shown in Figure 20.

## 4.2 Algorithms for Calibration

In this section, we describe the theory and implementation of all algorithms necessary to calibrate the camera. The first set of algorithms must isolate the target from range images and extract the points within the target. We take the $i, j$ coordinates from the centroids of the target spots and the $r$ value from the average range value within the spots. The $(r, i, j)$ points are stored for use by the second

set of algorithms which are designed to check the calibration of the camera model against ground truth target geometry.

The camera parameters are checked by converting the $(r, i, j)$ coordinates to a local Cartesian system centered on the camera. Because these points (call them $P_{\text{sensed}} = \mathrm{T}(r, i, j)$) are in a local coordinate system and the target points (call them $P_{\text{ideal}}$) are in a world coordinate system, we must perform a translation and rotation before comparing the fit of the two point–sets. A modification of Horn's *et al.* [2] method for determining the least–square translation and rotation is employed. After correcting the translation and rotation, we calculate the RMS error between the two point sets; however, this alone is not enough to perform a calibration. We must find the particular calibration parameters which minimize the RMS error. For this task, we employ the well known Simplex algorithm [25].

### 4.2.1 Segmentation

Early in the calibration process, the target points and their centroids must be isolated from the range and intensity images. Many methods for target location and point extraction were investigated. The most successful technique relied on the combination of primitive functions documented below. After target extraction, the point sets are stored in a $3 \times n$ array on disk. These points are in $(r, i, j)$ coordinates, but may be transformed to $(x, y, z)$ coordinates using the camera model.

1. Isolate the target from the background in the range image. This function starts from a user–specified seed point located at the center of the target and runs North, South, East, and West looking for jump edges. Once detected, these boundaries form the target window. The actual window is a few pixels smaller for noise margin.

2. Threshold the target in the intensity image. This function locates the two distinct modes of the target histogram and thresholds exactly between them, forming a binary image.

3. Extract target spots using a four–connected algorithm. The threshold value is passed onto this function which in turn makes groups of four connected pixels. Background pixels are labeled zero, and the upper left dot on the target is labeled one. The rest of the dots are labeled in increasing order, left to right, top to bottom.

4. Calculate the centroids of the segmented spots to obtain the $(i, j)$ coordinates of each point.

5. Find the mean of the range values from each spot to obtain the $r$ coordinates of each point. This calculation can be either a simple average, a ranked mean, or a Gaussian weighted mean.

Refer to Figure 21.

### 4.2.2 Absolute Position Loop

In general, we wish for the detected points given by the calibrated camera model to be as close to the actual points as possible in a least–squares sense. That is, the distances from the ideal points ($P_{\text{ideal}}$) to the sensed points ($P_{\text{sensed}}$) should be minimized, where $P_{\text{ideal}}$ and $P_{\text{sensed}}$ exist in 3-D coordinates. Note, however, that $P_{\text{ideal}}$ can take the form of $(x_i, y_i, 0)$ so that they only occupy a plane within a 3–D space. The minimization can be done by iteratively changing the calibration parameters in the camera according to some optimization scheme such as the geometric simplex algorithm of Nelder and Meade [25]. The flaw with this scheme is

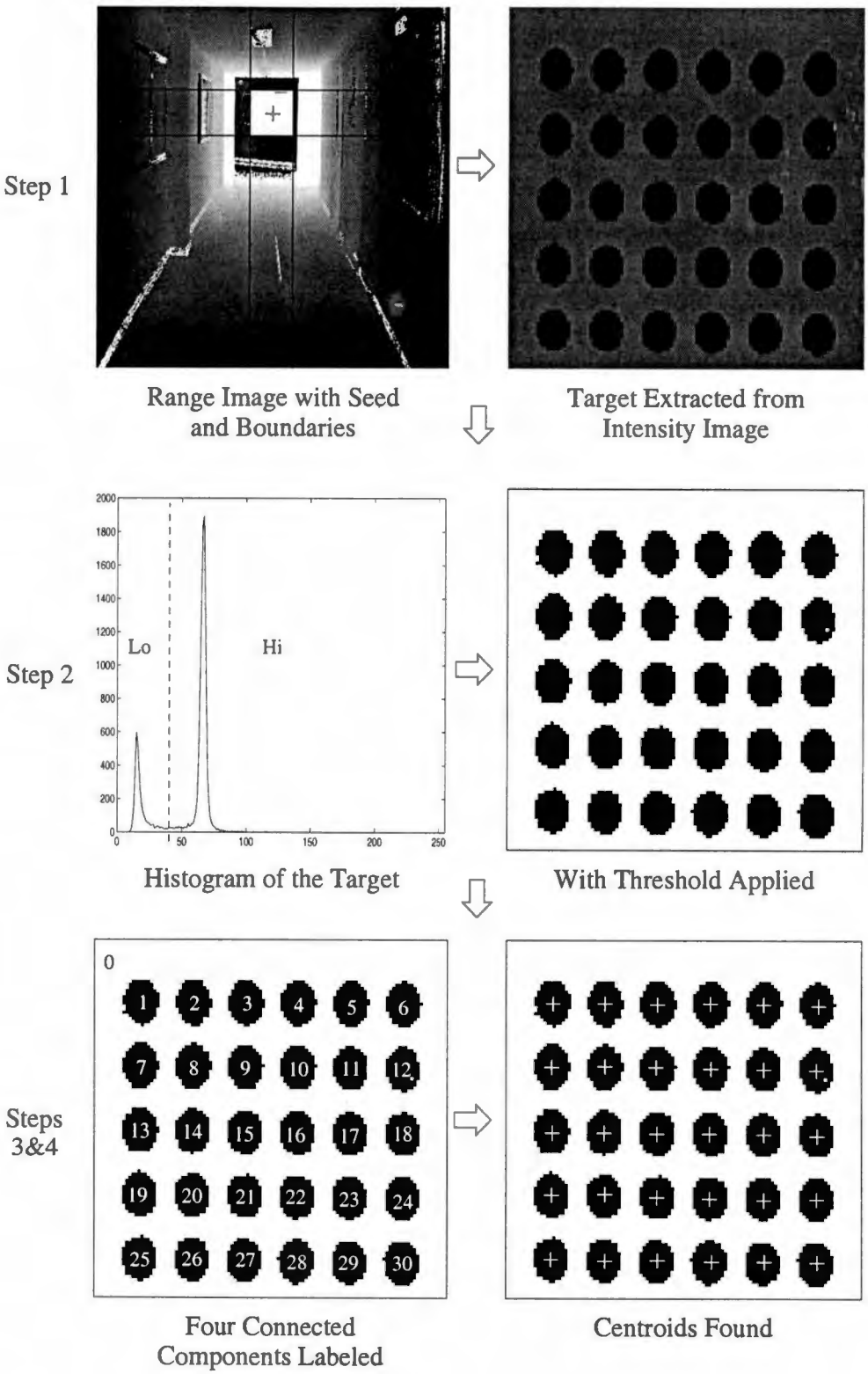



Figure 21: Results at significant stages in the automatic target segmentation process.

that the error function, $\Sigma \left\| P_{\text{ideal}} - P_{\text{sensed}} \right\|^2$, assumes that the $P_{\text{ideal}}$ is known to an arbitrary precision. In most cases, this is not true, nor is it desirable to attach the burden of absolute precise location to the calibration procedures. We therefore seek a slightly modified expression which drops the absolute position requirement in favor of a relative error measurement like $\Sigma \left\| P_{\text{ideal}} - P_{\text{sensed and aligned}} \right\|^2$. Where, $P_{\text{sensed and aligned}}$ are sensed points that have been projected into a three–dimensional coordinate system and aligned as best as possible with the set of ideal points through a translation and rotation.

We thus need the single rotation and single translation which brings $P_{\text{ideal}}$ into a best–fit orientation with the $P_{\text{sensed}}$ before calibration. Luckily, this type of problem has already been studied in other fields, especially photogrametry, which is concerned with the use of photographic images for measurement. Often, no single photograph can capture the entire scene to be surveyed. For many years, photogrameters have studied the mathematical problems associated with the registration of multiple images. Furthermore, many of the solutions give the necessary rotation and translation for registering two sets of corresponding points from two images. In 1988, Horn *et al.* [2] published a closed form solution for point registration in the Journal of the Optical Society of America. This method explicitly gives the necessary rotation-and translation to bring our $P_{\text{ideal}}$ into a best–fit orientation with the $P_{\text{sensed}}$. We refer to our implementation of this solution as the absolute position loop in future references.

The next step is the camera transform, discussed at length under scanner modeling. The associated matrix computations are carried out to transform the $(r, i, j)$ coordinates to $(x, y, z)$ coordinates. Afterwards, they are sent to the linear absolute position loop (Horn *et al.* [2]). This routine is the component which frees the calibration from tedious measurement of the camera location with respect to the target. The major component of the routine is the $US$ factorization loop which fac-

tors the cross correlation matrix into two parts: an orthogonal component $U$ and a symmetric component $S$. $U$ is the best–fit rotation from the camera coordinates to the target coordinates. Additionally, the trace of $S$ is the sum of least–squares residual error under $U$. A short proof follows:

Suppose two corresponding sets of points exist in two different coordinate systems. Call one coordinate system "camera" and designate points in these coordinates as $C_i$. Call the other system "target" and designate these points as $T_i$. We seek the translations and rotation bringing the $C_i$ points as close as possible to the $T_i$ points in a least squares sense. The translation between the two points is canceled out by removing their centroids. That is, subtract the centroids out of the two point sets to center them both on the origin. Consider the following. Call the centroid in the camera coordinates "$A$" and the centroid in the target coordinates "$B$".

$$\min \sum_{i=0}^{n} \|U(C_i - A) - (T_i - B)\|^2 \tag{45}$$

$$\frac{\partial}{\partial(A \text{ or } B)} \sum_{i=0}^{n} \|U(C_i - A) - (T_i - B)\|^2 = 0 \tag{46}$$

$$U \sum_{i=0}^{n} (C_i - A) - \sum_{i=0}^{n} (T_i - B) = 0 \tag{47}$$

$$U \left( \sum_{i=0}^{n} C_i - nA \right) - \sum_{i=0}^{n} T_i - nB = 0 \tag{48}$$

$$\text{Therefore, } A = \frac{\sum_{i=0}^{n} C_i}{n} \text{ and } B = \frac{\sum_{i=0}^{n} T_i}{n} \text{ as postulated.} \tag{49}$$

Call the new origin centered points sets $Co_i$ and $To_i$ respectively. Now we seek a rotation matrix $U$ to minimize,

$$\sum_{i=0}^{n} \|UCo_i - To_i\|^2 . \tag{50}$$

$$\min \sum_{i=0}^{n} \left[ (UCo_i)^T UCo_i - 2 (UCo_i)^T To_i + (To_i)^T To_i \right] \tag{51}$$

$$\min \sum_{i=0}^{n} \left[ (Co_i)^T U^T UCo_i - 2 \text{ Trace } \left( UCo_i (To_i)^T \right) + (To_i)^T To_i \right] \tag{52}$$

$$\min \sum_{i=0}^{n} \left[ (Co_i)^T Co_i - 2 \text{ Trace } \left( UCo_i (To_i)^T \right) + (To_i)^T To_i \right] \tag{53}$$

$$\max \sum_{i=0}^{n} \text{ Trace } \left( UCo_i (To_i)^T \right) \tag{54}$$

$$\max \text{ Trace } \left( U \sum_{i=0}^{n} Co_i (To_i)^T \right) \tag{55}$$

We therefore seek the matrix $U$ that can maximize the trace of $UM$ where $M = \sum_{i=0}^{n} Co_i (To_i)^T$. According to Horn *et al.*, we may factor any matrix $M$ into an orthogonal component, $U_0$ and a symmetric component, $S$ such that $M = U_0 S$. Furthermore, sections D and E of [2], Horn *et al.* demonstrate that $U_0$ is the maximizer of

$$\text{Trace } \left( U_0^T M \right) = \text{Trace } (S). \tag{56}$$

We infer from this result that the symmetric matrix has the highest trace possible and that $U$ is in fact $U_0^T$ obtained from the factorization. Horn *et al.* suggest using eigenvalues to factor $M$. When calibrating with planar patterns, however, $M$ is likely to be singular which ruins the approach.

We have developed an alternative method for factorization that works when $M$ is singular. We apply a series of simple orthogonal rotations, called Gibbs rotations, to $M$, until it becomes symmetric to machine precision. $U_i$ can be determined to make $M$ symmetric by two elements at a time. Unfortunately, successive rotations tend to disturb symmetries already established in the matrix. Each time through the loop, the matrix $M$ looks a little more symmetric until finally, after nine or ten complete iterations, the matrix is symmetric to machine precision. Therefore, $U = U_n U_{n-1} \ldots U_2 U_1$.

The residual error for the best–fit solution is very simply stated as

$$\sum_{i=0}^{n} \left[ (Co_i)^T Co_i + (To_i)^T To_i \right] - 2 \text{ Trace } (UM) \tag{57}$$

$$\text{or, } \sum_{i=0}^{n} \left[ (Co_i)^T Co_i + (To_i)^T To_i \right] - 2 \text{ Trace } (S) . \tag{58}$$

The code for this routine is quite compact and the only non–intuitive part is calculating the proper Gibbs rotations. Two rotation angles are always possible, but the smaller one, the rotation matrix that looks most like the identity matrix, (positives on the diagonal), is the best choice for good convergence. That choice minimizes the disturbance of the previously established symmetries.

### 4.2.3 Simplex Routine

Baker [4] uses a genetic algorithm for minimization. Other methods include conjugate gradients, simulated annealing, and the widely used simplex algorithm. Each method has advantages as well as disadvantages, and no method is guaranteed to find the best solution, only a good solution. Since the simplex algorithm is well suited for solving systems having many degrees of freedom, we chose this method for calibration. For instance, the conjugate gradients method requires "n" derivatives for a system with "n" degrees of freedom. In our case, we cannot even find one derivative because of the internal absolute–position loop.

The simplex algorithm (Nelder and Mead [25]) was invented in the mid–sixties to utilize computers in the optimization of multi–dimensional functions. The main advantage of the simplex is that only an expression for the functional needs to exist. This algorithm avoids using partial derivatives since they are often too hard to solve for, too many may exist, or some simply do not exist at points within the minimization region. For our purposes, where too many dimensions contribute

to the error functional and its derivatives depend on the complex behavior of the absolute position loop, the simplex approach is ideal.

A simplex is a geometrical object in n–dimensional space. The simplex comprises n+1 points or vertices and all interconnecting line segments, polygonal faces, etc. Furthermore, a simplex must contain volume; that is, it should fill some region of the n–dimensional space. For example, in three dimensions we may visualize the simplex as a tetrahedron, not necessarily a regular tetrahedron. Minimization of the objective functional is accomplished through reflections, expansions, and contractions of the simplex according to a well defined set of rules. The fundamental strategies developed by Nelder and Mead still hold. Figure 22 illustrates the basic operations of the downhill simplex algorithm.

Given an initial guess, the simplex proceeds to tumble downhill, stretching and squeezing through narrow valleys, flattening out and rolling across the plains. Finally, arriving at a local minimum, the simplex shrinks down to increase the precision of the solution. When plotted on a graph, the movement seems complex and deliberate. The movements are even less intuitive when moving downhill through the unimaginable complexity of some higher dimensional topography. This complex behavior derives from the repeated application of simple rules in an unknown environment.

The downhill simplex method must not be started with just a single point, but with n+1 points, defining an initial simplex. Thinking of one point as the initial starting point, $P_0$, we may take the remaining n points to be:

$$P_i = P_0 + \lambda e_i \tag{59}$$

where the $e_i$'s are n unit vectors, and where $\lambda$ is a constant determining the size of the initial simplex.

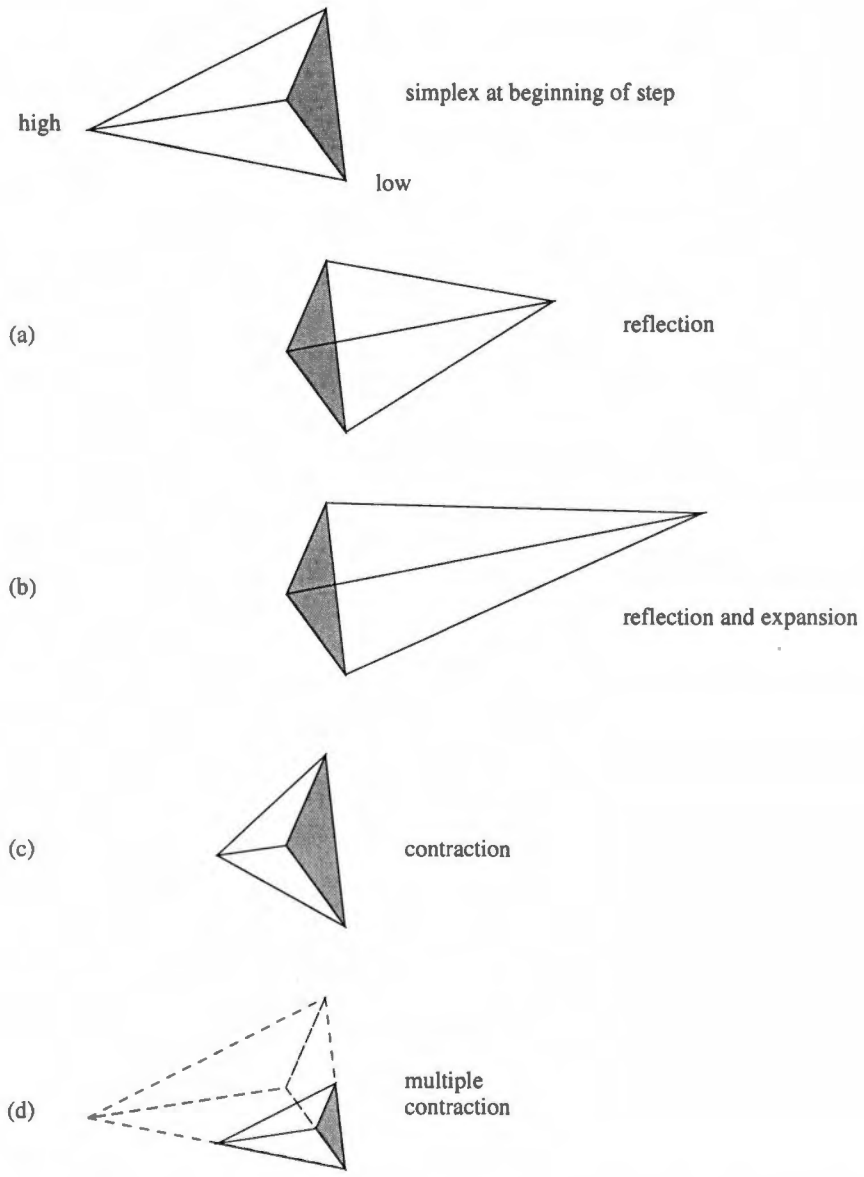



Figure 22: Possible outcomes for a step in the downhill simplex method. The simplex at the beginning of the step, here a tetrahedron, is shown, top. The simplex at the end of the step can be any one of (a) a reflection, (b) a reflection and expansion away from the high point, (c) a contraction along one dimension from the high point, or (d) a contraction along all dimensions toward the low point. An appropriate sequence of such steps always converges to a minimum of the function. (Adapted from Numerical Recipes in C [40], ©Cambridge University Press.)

The downhill simplex method takes a series of steps, most steps just moving the point of the simplex to where the function is largest ("highest point"), through the opposite face of the simplex to a lower point. These steps are called reflections, and are constructed to conserve the volume of the simplex (hence maintain its non–degeneracy). When it can do so, the method expands the simplex in one direction to take larger steps. When it reaches a "valley floor," the method contracts itself in the transverse direction and tries to ooze down the valley. If there is a situation where the simplex is trying to "pass through the eye of a needle," it contracts itself in all directions, pulling in around its lowest (best) point. The more detailed account can be found in Numerical Recipes in C [40].

## 4.3 Summary

A target and a mobile range cart were designed and fabricated to assist in carrying out calibrations. Their construction was carefully considered to fulfill the portability requirement. When the cart was completely assembled, the system was tested by acquiring some images of the target. Software was then developed to solve the calibration problem. This software is divided into two main components. The first component is a segmentation package for automatically finding the target and extracting the calibration points in range image coordinates. Trials on several of the test images proved that the segmentation package was robust. The second component is the main calibration routine. This program contains an unconventional double loop which automatically finds the orientation and location of the target in 3–D space. The problem of translation and rotation (absolute position) is solved in an inner loop at each step in the calibration of the camera model. The target position does not therefore have to be known prior to calibration. A block diagram of the double loop calibration routine can be found in Figure 23. The large outer

loop seen in Figure 23 represents operation of the simplex routine using a numerical technique to find the calibration parameters. Operation of the loop ceases when the parameters are found. The next chapter describes rigorous testing of the main calibration routine. The primary question is whether the routine will perform satisfactorily under real conditions which often include unusually high amounts of noise in the images.

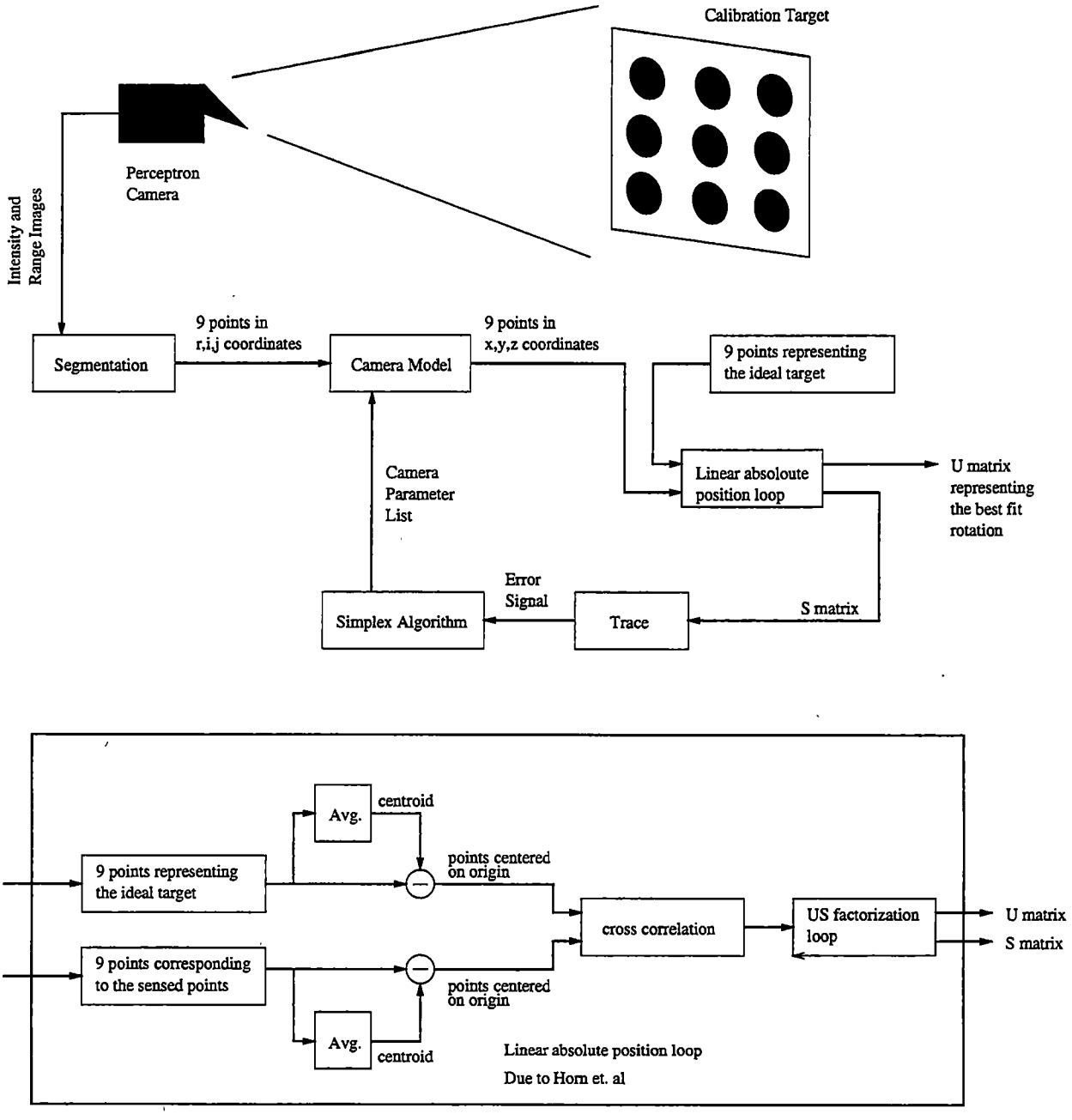



Figure 23: Unique double loop calibration scheme.

# 5 RESULTS: CALIBRATION PERFORMANCE IN A REAL EXPERIMENT

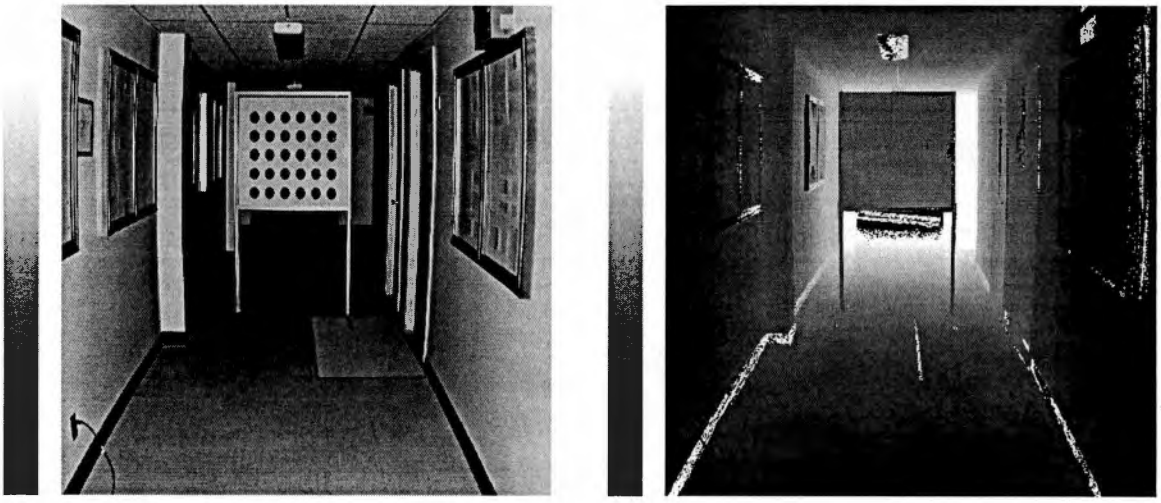
This chapter presents the results of an experiment designed to explore various aspects of performance of the calibration developed in Chapter 4. Of particular interest are speed of convergence, residual error after calibration, stability of the solution from different initial conditions, and the effect of different target placements within the scene.

## 5.1 Acquisition of a Series of Target Images

Range images of the target were acquired at several distances (34, 30, 26, 22, 18, 14, and 10 feet). Figure 24 is an example range and intensity image pair taken at 18 feet. Table 1 gives the (range, i, j) values extracted from the images. The entire sequence of images are stored on disk to form a database for further calibration work. Next, each image was segmented by the algorithm from Figure 21 in the last chapter, forming arrays of ordered $(r, i, j)$ coordinates. These coordinates are then saved in ASCII data files. This pre–processing speeds up execution of the calibration by more than 50 percent, since segmentation is historically a slow process.

## 5.2 Preliminary Evaluation of the Data

Since all the data were taken at different distances, a significant question arises about how distance affects the quality of the measurements. Good measurements can only be made when the signal to noise ratio exceeds some threshold. If the detected laser return is low, any noise contributed from the detector electronics



Intensity Component

Range Component

Figure 24: Intensity and range images of the target at 18 feet.

Table 1: Centroids of the target points found by the automatic segmentation.

| point | range | i | j | point | range | i | j |
|---|---|---|---|---|---|---|---|
| 1 | 1508.739 | 400.704 | 235.327 | 16 | 1481.859 | 496.641 | 314.569 |
| 2 | 1490.547 | 432.597 | 235.370 | 17 | 1485.210 | 529.007 | 314.792 |
| 3 | 1482.558 | 464.583 | 235.444 | 18 | 1477.587 | 561.436 | 314.983 |
| 4 | 1484.272 | 496.810 | 235.570 | 19 | 1510.284 | 400.428 | 353.368 |
| 5 | 1482.684 | 529.094 | 235.558 | 20 | 1488.020 | 432.341 | 353.729 |
| 6 | 1477.146 | 561.480 | 235.575 | 21 | 1477.650 | 464.350 | 353.985 |
| 7 | 1507.205 | 400.568 | 274.551 | 22 | 1482.378 | 496.575 | 354.227 |
| 8 | 1481.579 | 432.509 | 274.696 | 23 | 1485.899 | 528.834 | 354.515 |
| 9 | 1476.799 | 464.561 | 274.970 | 24 | 1478.371 | 561.241 | 354.816 |
| 10 | 1482.139 | 496.733 | 275.012 | 25 | 1513.489 | 400.554 | 392.822 |
| 11 | 1482.717 | 529.074 | 275.017 | 26 | 1502.182 | 432.361 | 393.189 |
| 12 | 1475.921 | 561.419 | 275.072 | 27 | 1490.098 | 464.322 | 393.533 |
| 13 | 1509.556 | 400.511 | 313.972 | 28 | 1486.908 | 496.475 | 393.833 |
| 14 | 1479.929 | 432.417 | 314.163 | 29 | 1486.657 | 528.632 | 394.095 |
| 15 | 1476.239 | 464.461 | 314.443 | 30 | 1477.858 | 561.037 | 394.517 |

could ruin the range measurements. Quality therefore depends on the amount of returned laser light, so one can successfully argue that the closer object returns more laser light to the camera. Thus, the closer object gives the better range values. This is an inverse $r^2$ effect. Additionally, an averaging of pixels takes place, improving the range values. When $n$ pixels contribute to the measurement, the variance improves by a factor $n$. Again, the close objects appear larger in each image. The number of pixels per dot, $n$, is also inversely proportional to $r^2$. We conclude that the variance of the measurement effectively worsens with $r^4$. This relationship places strict limits on the laser power, sample time curve for any fixed distance.

The data extracted from the seven test shots was graphed in $(r, i, j)$ space to allow a visual inspection to determine how the distance affected the data from the test targets. Figure 25 shows how the $r^4$ effect ruins the most distant data sets. Some other anomalies, such as unusual aberration from the correct orientation, are observed. As a compromise, only the best data sets, the 10, 14, 18, and 22 feet target data, were used in the calibration. Additionally, it is concluded that some distance restrictions may always apply to target placement within the scene.

## 5.3 Performance Analysis of the New Calibration Algorithm

A particular calibration run begins by selecting initial conditions for all model parameters. These essential camera model parameters are stored in a file. Many of these parameters can be determined from the manufacturer's specifications. The uncertain parameters must still be initialized to some value before calibration. These parameters can usually be guessed from the manufactures specifications or from looking at test data. In our case, four parameters are particularly variable

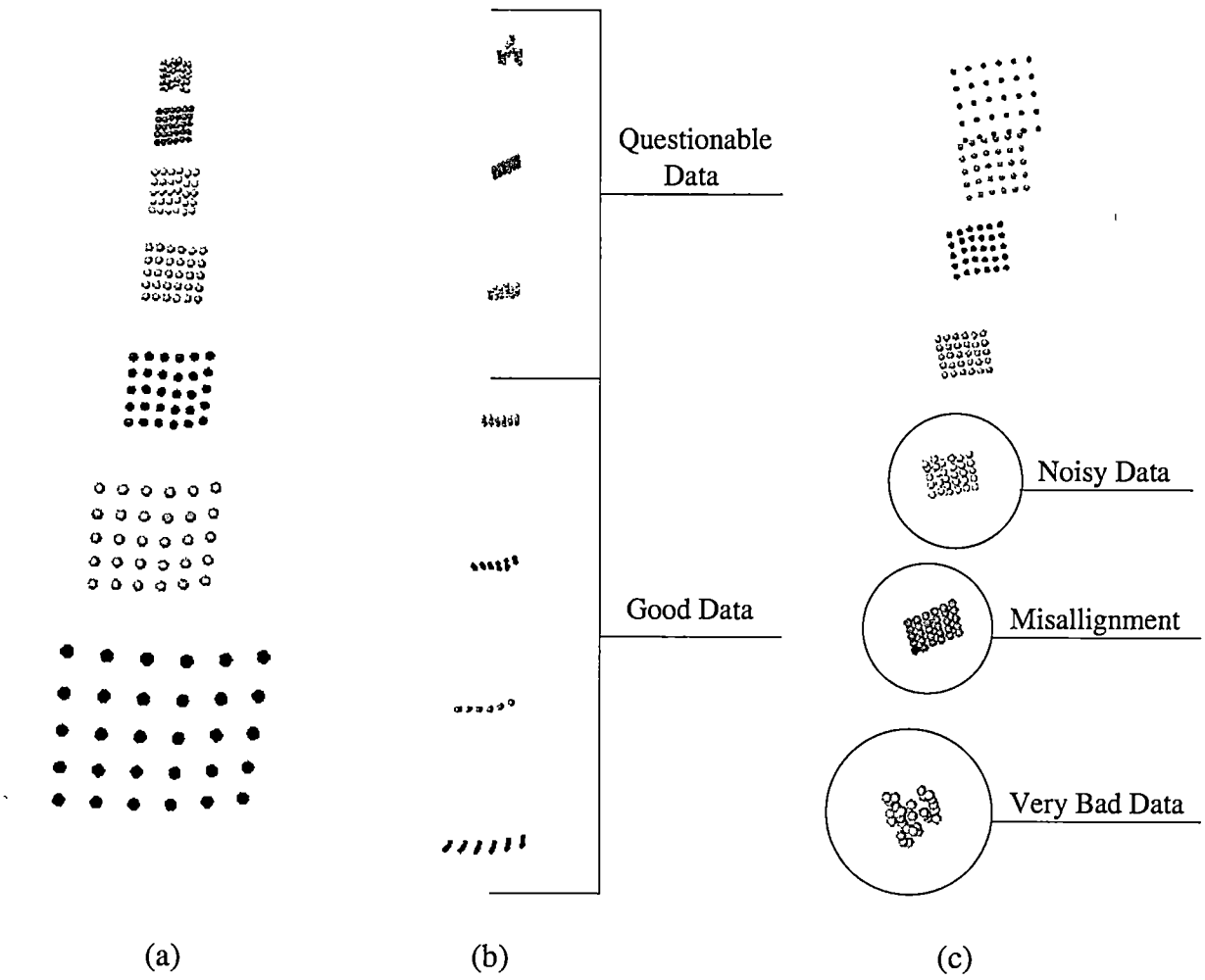



Figure 25: Test data viewed (a) from front, above (b) from above (c) from behind.

among the individual camera units. They are enumerated below:

1. Laser source location $S$;
2. Distance per range increment $l$;
3. Azimuth mirror angle increment per pixel $\alpha$; and
4. Elevation mirror angle increment per pixel $\beta$.

The following is a list of approximate values for these parameters, as determined from consulting with the manufacturer.

1. $S$ = -1700 mm $\hat{z}$;
2. $l$ = -2 mm $\hat{z}$;
3. $\alpha$ = 0.001022 rad per pixel; and
4. $\beta$ = 0.0005113 rad per pixel.

Before starting the calibration, we made a visual inspection of the error between the uncalibrated data and the ideal data by plotting the points in 3–D on a Silicon Graphics workstation. The plot is shown in Figure 26, where the uncalibrated data are in light gray and the ideal data are in dark gray. At this point, the calibration errors far exceed those resulting from random sources.

The behavior of a particular numerical algorithm during convergence is an important characteristic. Convergence curves are often studied to determine the efficiency and stability of a new algorithm. To characterize the performance of the algorithm presented here, we modified the program to record the state of each calibration parameter during every iteration. As the simplex walks through parameter

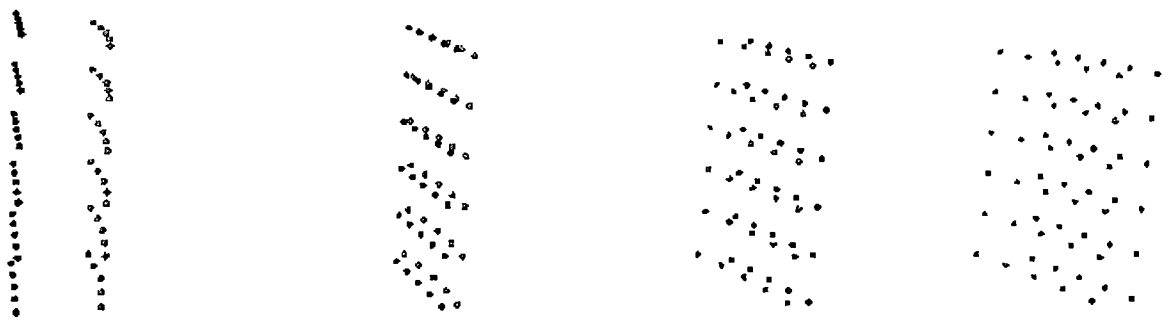


Figure 26: Initial bad fit of the data points to the ideal.

space seeking a minimum, the path can be quite complex, especially in four dimensional space. This leads to the question of how to view the four dimensional path of convergence in just three dimensions or less. Since no one method seems sufficient, we present a variety of graphs in Figures 27–33. These plots show the number of steps of convergence during calibration for each parameter, as well as the rms error before, during, and after calibration. Two final plots show 3–D projections of the convergence path in three of the four dimensions simultaneously.

From Figure 31 we see how the algorithm converges quickly, finding a solution in less than fifty iterations. Most importantly, the RMS error is reduced from more than 60 mm per data point to less than 10mm per data point. We shall compare a plot of the calibrated data to the ideal data later to determine whether the residual 10 mm error is because of a calibration error (ie. finding a local minimum) or whether it is due entirely to random sources of noise.

Upon completion, the program outputs new calibrated parameter values. For this example, the calibrated values are enumerated below.

1. $S$ = -2015.218750 mm $\hat{z}$;

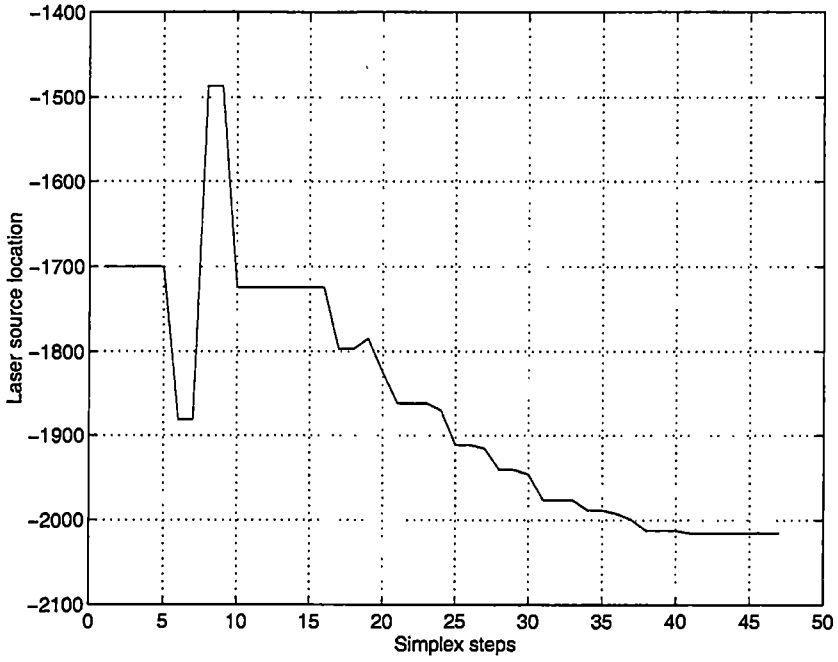



Figure 27: Convergence of the laser source location.

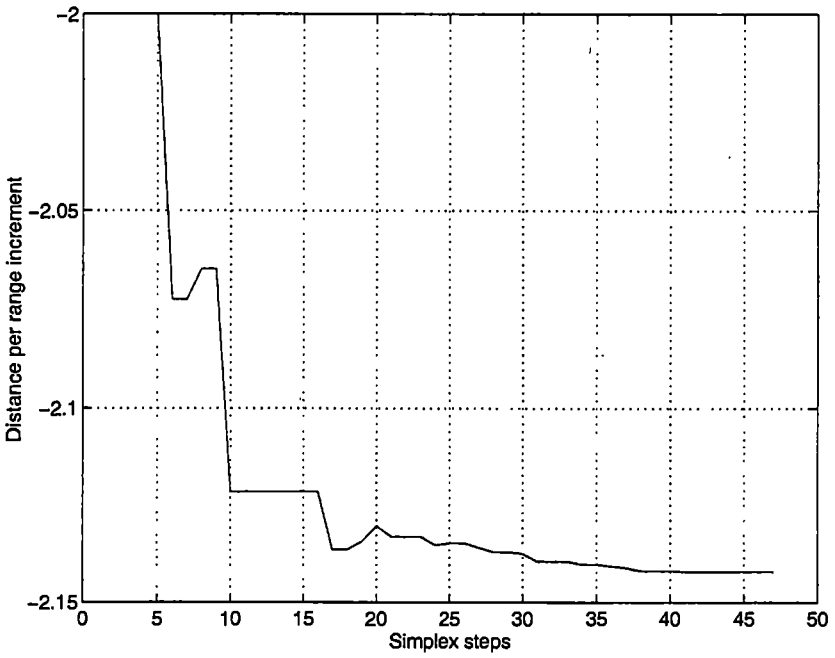



Figure 28: Convergence of the distance per range increment.

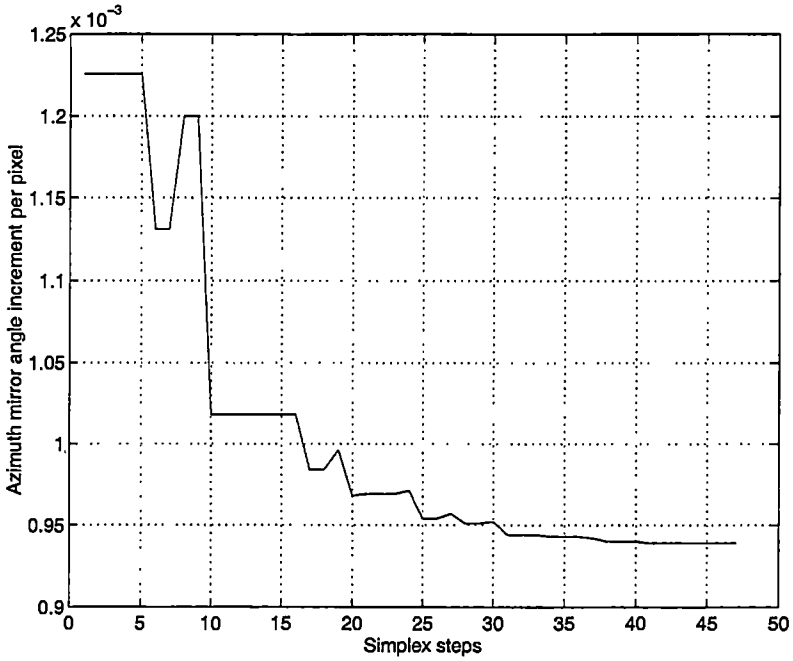



Figure 29: Convergence of the azimuth mirror angle increment per pixel.

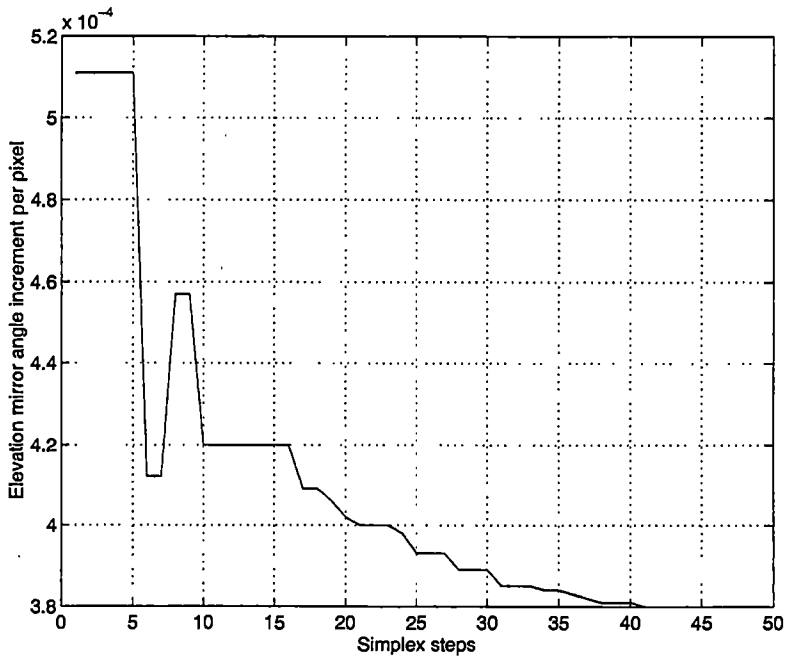



Figure 30: Convergence of the elevation mirror angle increment per pixel.

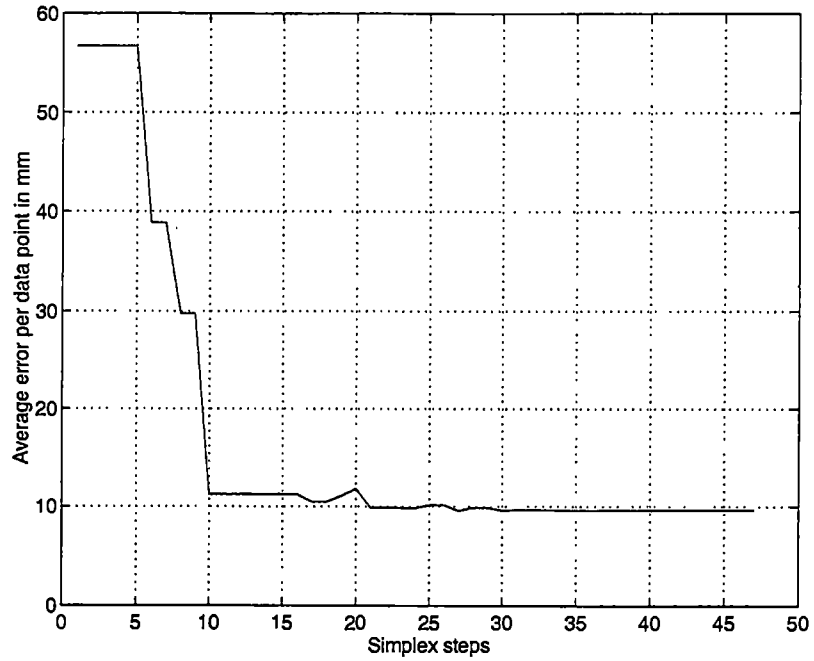



Figure 31: Minimization of the average distance error per data point.

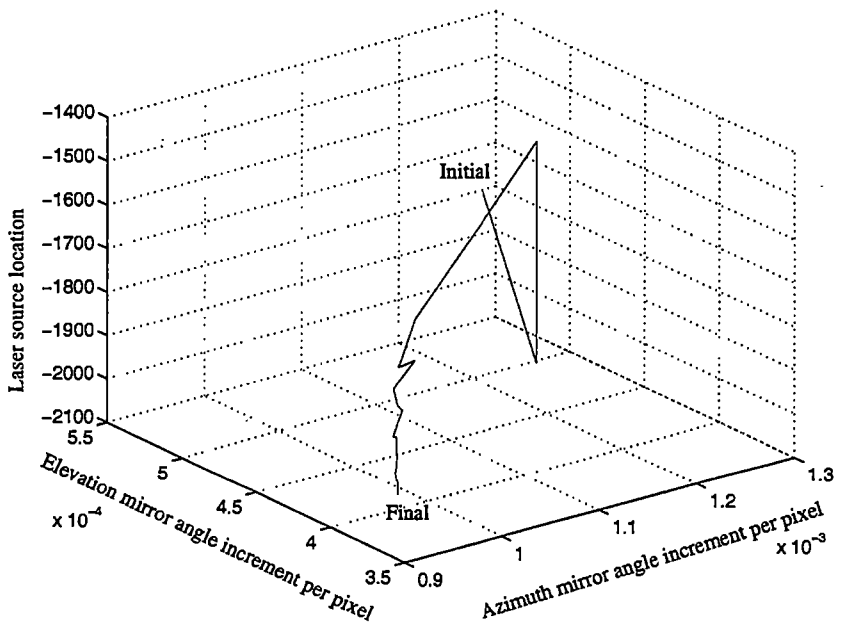



Figure 32: Movement of the simplex through three dimensions of the parameter space.

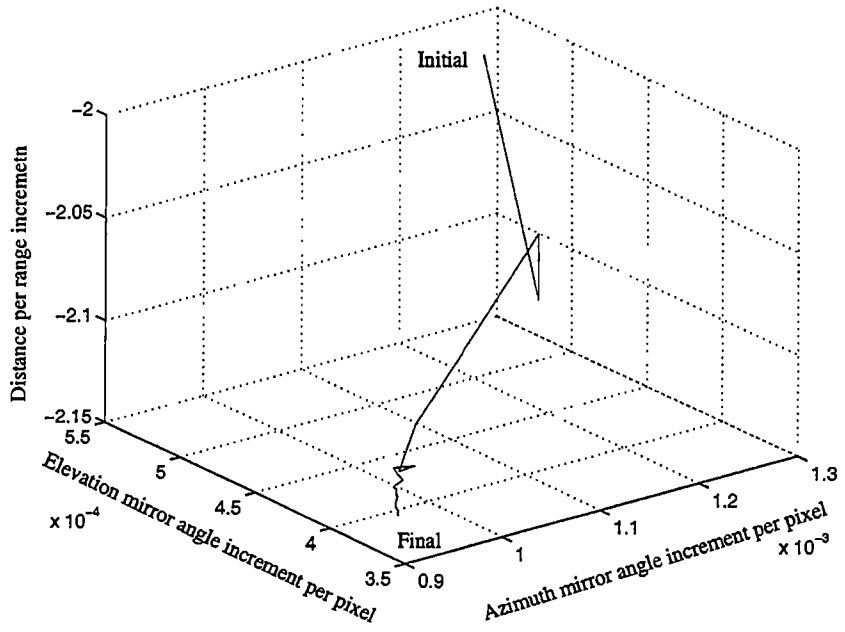



Figure 33: View of the simplex moving through another three dimensions of the parameter space.

2. $l$ = -2.142039 mm $\hat{z}$;

3. $\alpha$ = 0.000939 rad per pixel; and

4. $\beta$ = 0.000380 rad per pixel.

Figure 34 illustrates graphically the final good fit of the test data points (light gray) with the ideal data points (dark gray) after calibration. A closer view is shown in Figure 35. Primarily, there appears to be no consistent pattern to the errors, as we might expect if there were any remaining calibration error. Rather, the noise is random, with most of the contribution in the $\hat{z}$ direction. Furthermore, the fact most of the noise is in the $\hat{z}$ direction correlates to noise in the range values rather than pointing errors.

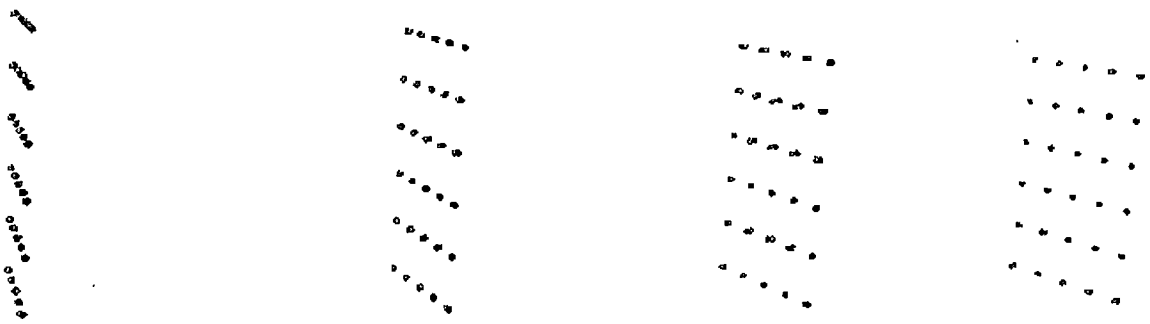


Figure 34: Final good fit of the data points to the ideal

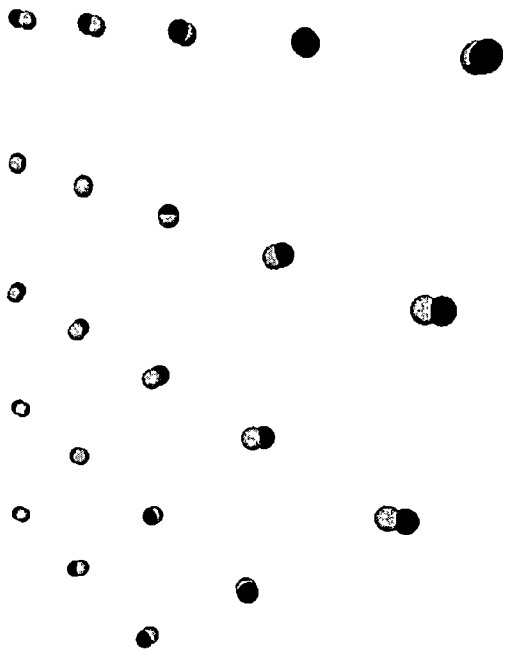


Figure 35: A close up of the final fit showing noise in the range values.

Numerical algorithms often exhibit sensitivity to initial conditions. That is, when the initial conditions are slightly changed, the results may also vary. When many local minima exist, it is entirely possible that the algorithm will only find the nearest local minimum, not the global solution. However, if the error function is smooth and contains few minima, the algorithm will exhibit very little sensitivity to initial conditions.

To determine the sensitivity to initial conditions, the calibration was initialized several times with different parameters. Fortunately, the routine always converged within 0.1 percent of the same solution. This kind of behavior clearly indicates a well defined global minimum in the error function. This result also reassures us of the method's stability.

## 5.4 Summary

An experiment was designed to explore the performance of a calibration method developed in Chapter 4. This experiment required test images from a pre–selected target. The necessary range and intensity images were acquired at several distances and stored in digital format. A segmentation algorithm was next applied to these image pairs to extract $(r, i, j)$ coordinates of specific target points. We then discussed the relationship between distance and range accuracy. Persuant to this discussion, we determined to plot our data and inspect for anomalies, especially at large distance. We indeed discovered some defects in the recovered target data. Furthermore, these all occurred at large range values. After isolating the four best (and closest) data sets, we initialized our new calibration routine with some approximate values. Upon completion, we evaluated the convergence efficiency by referring to several graphs created during the process. It is determined from this experiment that the new calibration is both efficient and robust. It is furthermore

concluded that most of the residual error results from random noise in the range measurment, not calibration error.

# 6 APPLICATIONS

This chapter introduces applications directly resulting from the calibrated camera model and examines its implications. Primarily, how does the new model relate to the older models based on spherical coordinates, and in what case is the spherical coordinates approximation valid? Secondly, we reconstruct a 3–D scene from a single range and intensity image pair as a first step toward the integration of multiple range and intensity images into a common 3–D model. Finally, we address the warping effects spherical projection causes. This effect is not an error but is rather because of the spherical type of sampling the camera uses. The curvature seen in the images is the result of using square pixels for mapping $(i, j)$ coordinates to the screen. A transformation is derived to re–map the range and intensity images so that straight lines in the scene compare to straight lines in the images. There are, of course, limitless applications for range data; however, the previous three items appear to be the most pertinent and immediate results linked to developing the calibrated camera model.

## 6.1 Addressing Warpage in the Images

A distinctive warping effect causes vertical lines in the scene to map to curved lines in the intensity and range images. This mapping results from the spherical coordinate projection used in the Perceptron camera. Compare this kind of imaging to perspective projection, where a lens or pinhole is employed in projecting the images. Perspective projection is one of the projections which always maps straight lines in 3–D to straight lines in the image. The perspective projection is preferable for at least two reasons:

1. It looks correct since it is the type of projection employed within the eye; and

2. Many computer algorithms assume perspective projection. For example, the Hough transform was designed to isolate straight lines under the assumption of perspective projection, which maps lines and planes in 3–D to lines and planes in the images.

For our purposes, we are interested in a projection preserving straight lines, but is also closely related to spherical coordinates. Equal–angle coordinates is an obscure coordinate system, but it does produce much less warped looking images. In fact, straight lines which are perpendicular to the $z$ axis are perfectly straight in (EA) projected images. Both the spherical and equal– angular systems possess a $\theta$ and a $\phi$, but in equal angular (EA) projection, their roles are more balanced. Recall the form of the spherical coordinates from Equation 33:

$$\begin{aligned} x &= \sin(\theta)\ (radius) \\ y &= \cos(\theta)\sin(\phi)\ (radius) \\ z &= -\cos(\theta)\cos(\phi)\ (radius) \end{aligned} \tag{60}$$

As denoted previously, the two angles, $\theta$ and $\phi$, share an equivalent role in EA projection. Figure 36 helps explain the geometry of EA projection. Looking at this figure, we infer the following equations from the two triangles:

$$\begin{aligned} \tan(\Theta) &= x/-z \\ \tan(\Phi) &= y/-z \end{aligned} \tag{61}$$

where, $\Theta$ and $\Phi$ are the EA projection coordinates on the horizontal and vertical axes respectively.

The goal is an analytic function for transforming the spherical coordinates to EA coordinates. Beginning with Equation 61:

$$\tan(\Phi) = y/(-z) = \frac{\cos(\theta)\sin(\phi)\ (radius)}{\cos(\theta)\cos(\phi)\ (radius)} = \frac{\sin(\phi)}{\cos(\phi)} = \tan(\phi)$$

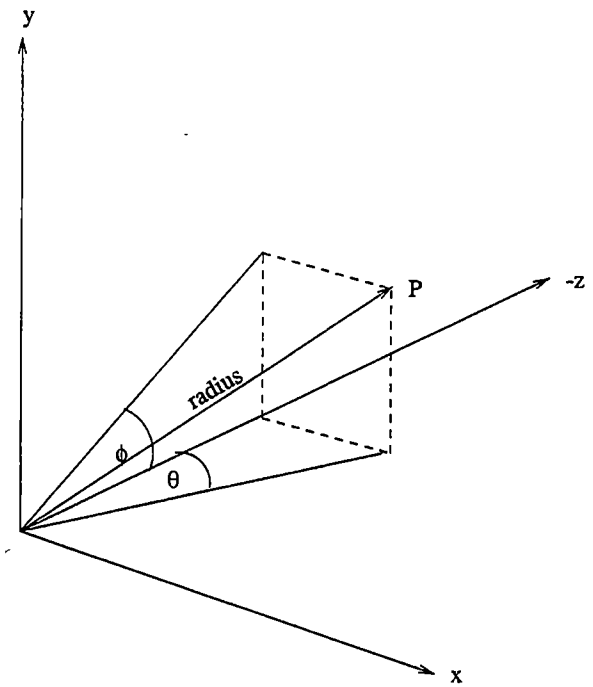



Figure 36: A diagram of the equal–angular projection system.

$$\tan(\Theta) = x/(-z) = \frac{\sin(\theta)\ (radius)}{\cos(\theta)\cos(\phi)\ (radius)} = \frac{\sin(\theta)}{\cos(\theta)\cos(\phi)} = \frac{\tan(\theta)}{\cos(\phi)}$$

$$\Phi = \phi \tag{62}$$

$$\theta = \arctan(\tan(\Theta)\cos(\phi)) \tag{63}$$

Since the vertical angles $\phi$ and $\Phi$ are the same in both systems, we conclude that a warping of $\theta$ is sufficient to correct the problem.

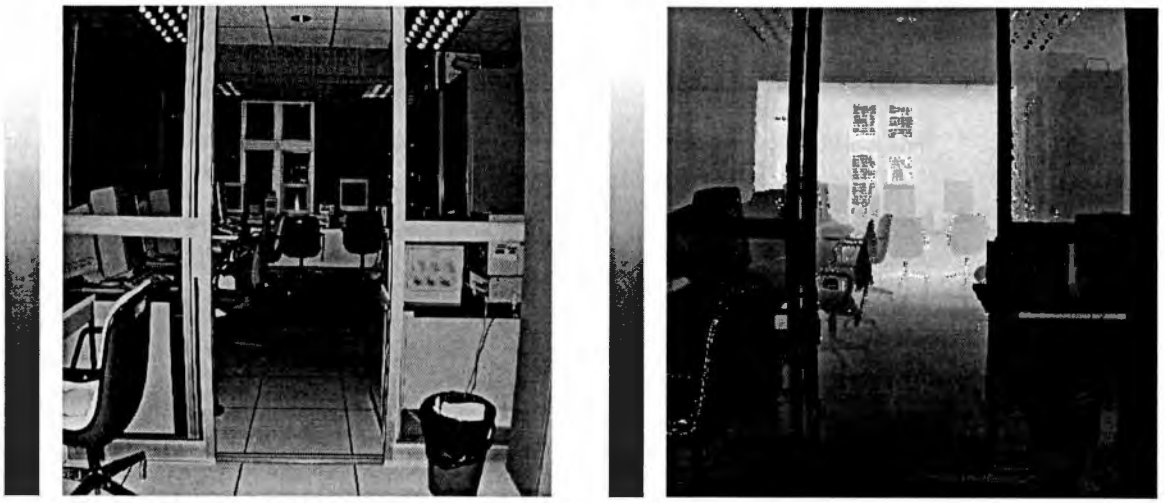
To de–warp the image, one must know the degrees per pixel for $\theta$ and $\phi$. Because $\phi$ is the same for both systems, no vertical warping of the images occurs. Each raster line is independent.

This analysis suggests an efficient algorithm for carrying out the de–warping. One scan line from the image is loaded into a buffer; it is re–sampled according to Equation 63 and stored in another buffer until it can be written out to the destination file. This operation is repeated on every scan line of the image. Therefore, only

enough memory to hold two scan lines from the image is needed. The efficiency of this technique will be appreciated once the images exceed a few million pixels in size.

The visual effect of equal–angular projection is to de–warp the apparent image stretching done by spherical projection. A dramatic example of warpage can be seen in Figure 37, where the spherical warping has an extreme effect on the left and right boundaries of the images (straight structural members appear curved). Compare these images with those seen in Figure 38 which have been transformed to the equal–angular coordinate system. The much improved appearance of the doorway and storage cabinet is once again because equal–angular projection can map straight lines in the scene to straight lines in the image. Figures 39 and 40 present another sample of the improvement coming from de–warping. Although spherical coordinates may represent the internal workings of the camera in addition to providing a method of recovering the 3–D geometry from the scene, we conclude that equal–angular transformation has useful applications for many edge sensitive algorithms.

As mentioned earlier, many computer algorithms assume perspective projection. The Hough transform, for instance, was specifically designed to isolate straight lines under the assumption of perspective projection. For example, consider the industrial application of automatically segmenting pipes from a range image. In other applications, objects are segmented from the scene and stored as closed polygons consisting of a minimum number of straight line segments. Sometimes these chains of linked line segments are used to sort out the objects of interest. Warpage obviously encumbers the search for straight edge segments as in Figure 41, by artificially breaking the edges into a multitude of nearly linear segments, wasting processor time.

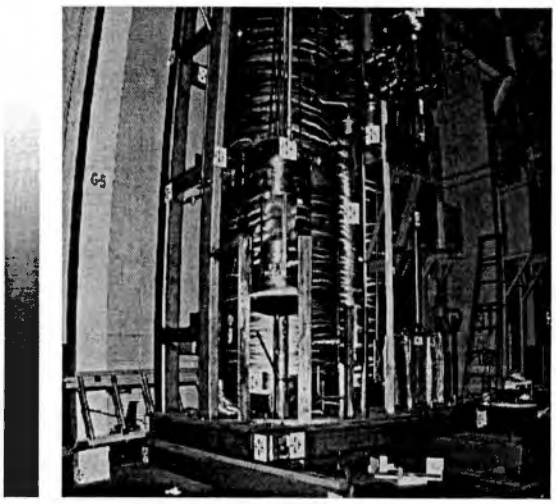


Intensity Component Range Component

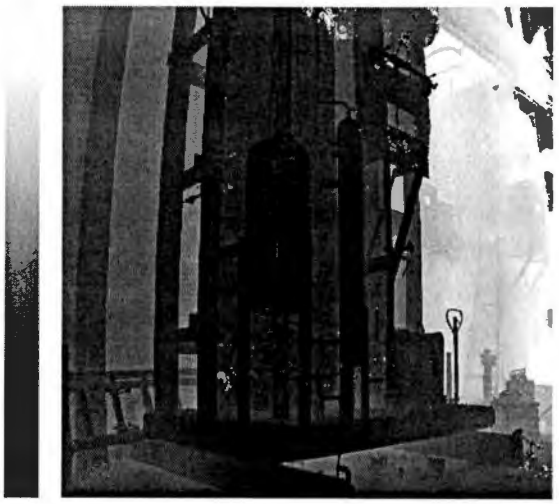
Figure 37: This image shows significant spherical warping near the left and right edges.

Intensity Component Range Component

Figure 38: The same image after de–warping

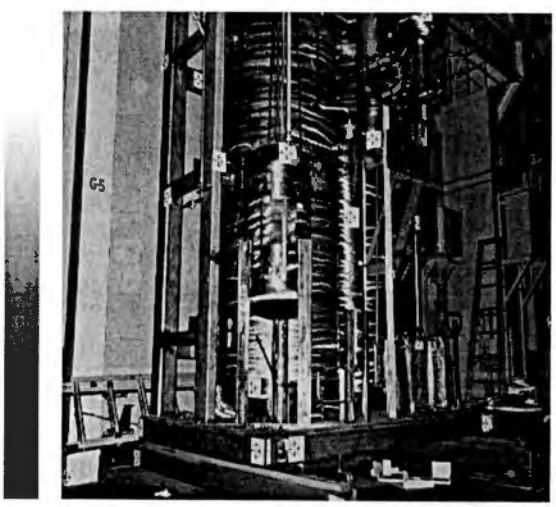


Intensity Component

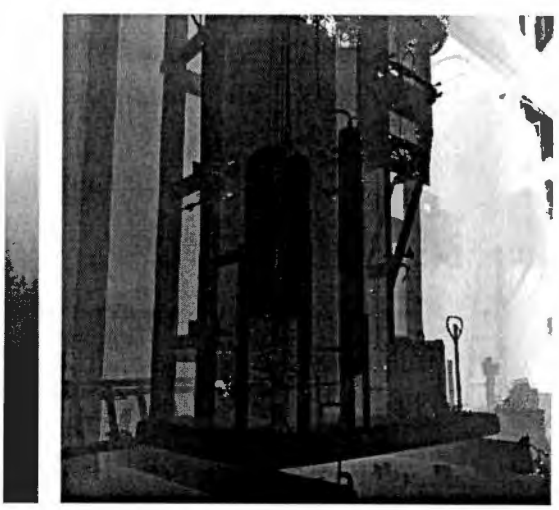


Range Component

Figure 39: Another image showing significant spherical warping near the left and right edges.

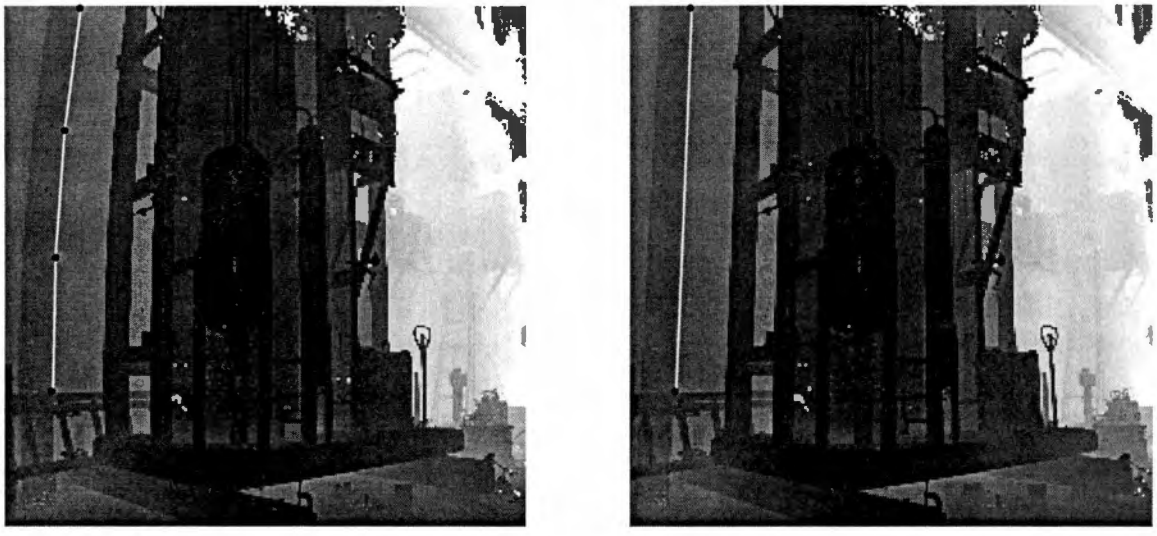


Intensity Component



Range Component

Figure 40: Again, the result of de–warping.



(a) (b)

Figure 41: The edge in (a) requires three line segments, while the same edge in (b) can be represented with only one line segment.

## 6.2 Full Angle Range Images

Another application of spherical coordinates is for storing a full angle image. By full angle, we mean every viewing direction that can be represented by a vector to the unit sphere surrounding the camera. Because the images from the Perceptron camera are in spherical coordinates, each image represents a small segment of the surface of a projection sphere around the camera. We conjecture that by taking multiple range images from the same location, a continuous map of the projection sphere can be made. The task, therefore, is to warp each image so that no overlap occurs with previously taken images. Similar forms of image mosaicing have been proposed in the past, using conventional cameras with perspective projection [39]. Szeliski describes a technique for mosaicing images of planar surfaces from different positions, and for mosaicing a panoramic image from several shots taken at the same

position. Rather than calibrating the camera, the author relies on local and global registration techniques to find the required transformations between images. In our work, however, camera calibration was viewed as the primary goal. Therefore, the method we propose relies on the calibration parameters we derived earlier. Our method results in a composite image spanning more than 360 degrees in azimuth, and nearly 160 degrees in elevation (nearly full angle). This large panoramic image is formed by mapping each small image from its natural $\theta$, $\phi$ coordinates to a common world $\theta$, $\phi$ coordinate system.

Mapping can be done in two steps using the algorithm developed for de–warping images. The first pass makes a de–warping transformation from the local $\theta$, $\phi$ system to an intermediate standard equal–angular coordinate system. The second pass makes a transformation from the equal– angular coordinates back to world $\theta$, $\phi$ coordinates. This mapping generally stretches either the top or the bottom of the images, depending on which hemisphere they were taken from. Near the equator of the sphere, little warping is apparent. Near the poles, however, the image becomes very stretched. The implication is that straight lines are always warped near the poles of the image (top and bottom). This problem is not correctable either, because there is no way to map a sphere to a plane without warpage. The new image essentially gives a full view around the camera in every direction. Figure 42 shows a nearly full angle intensity image obtained from the composite of 28 individual $1024 \times 1024$ images, resulting in a $7168 \times 4096$ pixel image. Likewise, Figure 43 shows a nearly full angle range image obtained the same way.

The registered range component forms a companion image, giving the direct spherical coordinates of every visible surface point. This single range image provides the most complete 3–D map of the environment for autonomous navigation and environment mapping.

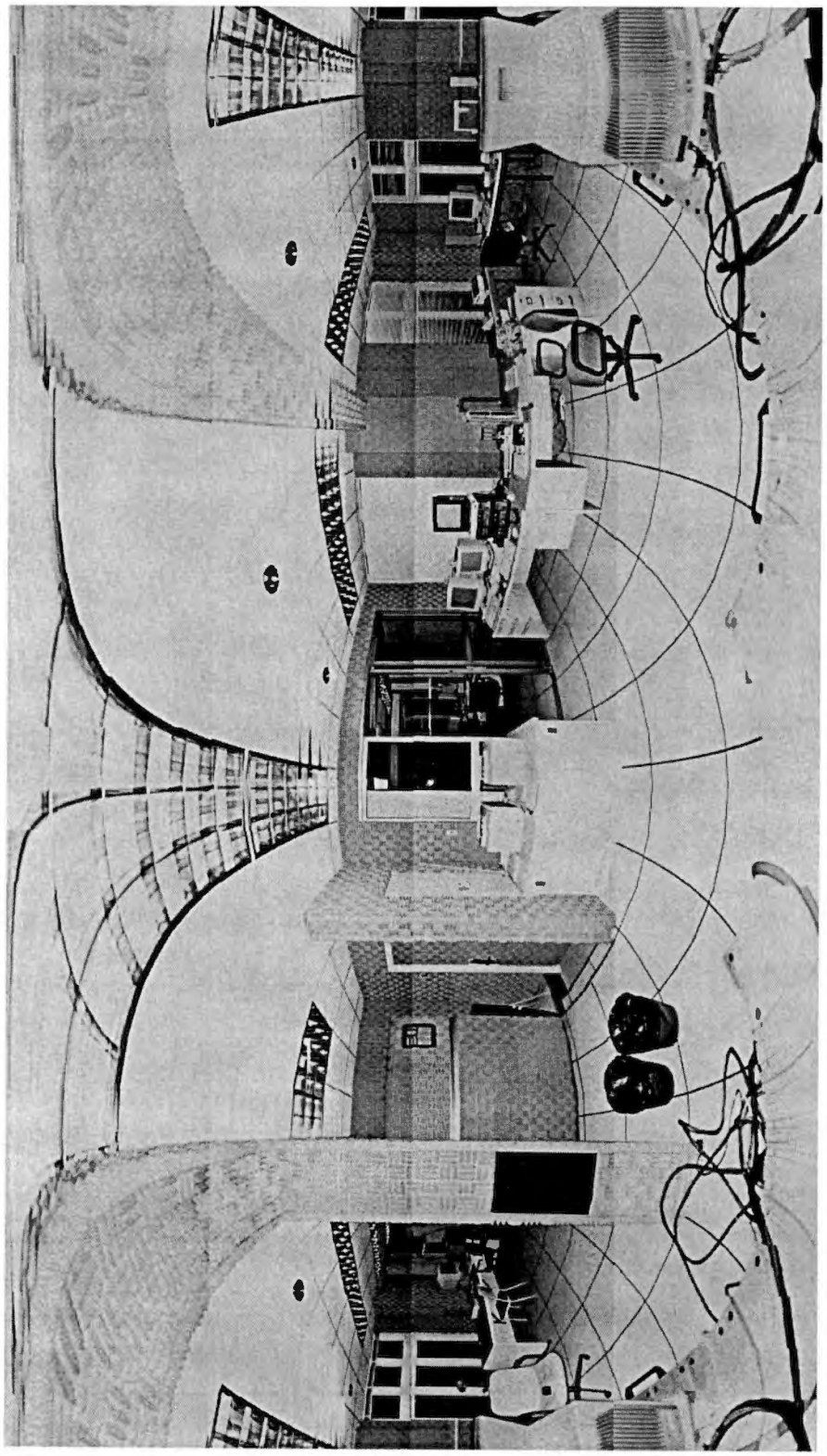


Figure 42: Example of a full angle image composed from 28 individual $1024 \times 1024$ images.

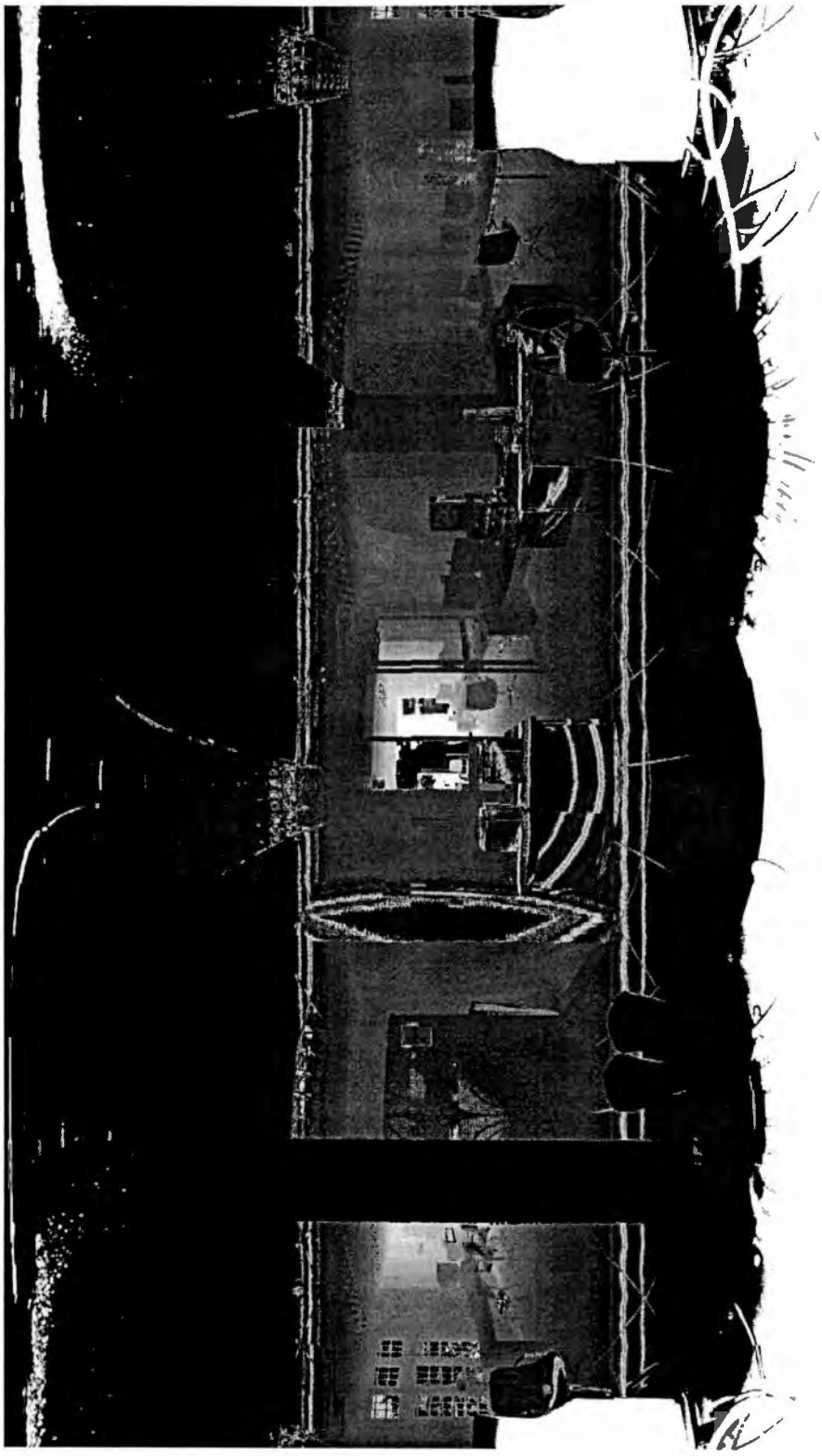


Figure 43: The range component of the same image composed from 28 individual $1024 \times 1024$ images.

Because these images contain all the scene around the camera, it should be possible to obtain any type of image projection (equal–angular, perspective, or wide angle lens) in any direction of view by properly re–sampling the image. We undertook to study a 180 degree lens projection of the image from any user– defined direction. The primary method is to calculate all the local $(x, y, z)$ unit vectors corresponding to the desired projection model, pass these vectors through the desired rotation matrix to obtain the correct look direction, and extract the corresponding $(\theta, \phi)$ world coordinates. The $(\theta, \phi)$ values are then used to select the appropriate set of pixels to compose the new image. Bi–linear interpolation is used to calculate the actual pixel values to avoid aliasing. Some results are shown in Figures 44–47 for different viewing directions. Note the missing data in Figure 47. Tilting the camera straight down is impossible because of the support post obstruction.

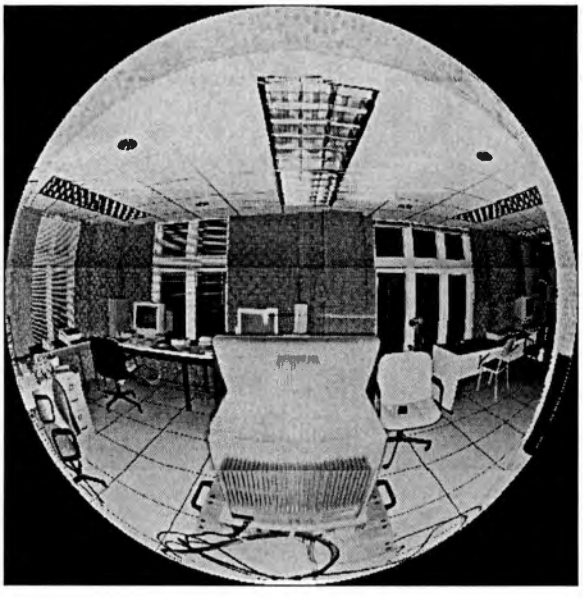


Figure 44: A fisheye lens projection of the scene behind the cart.

Figure 45: A fisheye lens projection of the scene in front of the cart.

Figure 46: Viewing the scene at a declined angle.

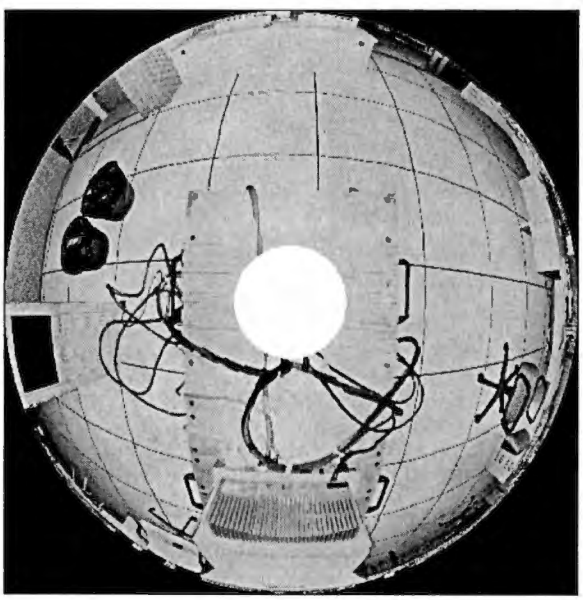
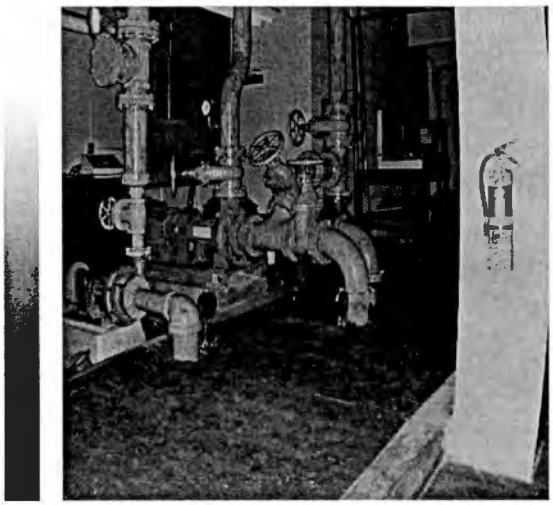


Figure 47: Looking down, we cannot see the area directly below the camera because of mounting hardware.

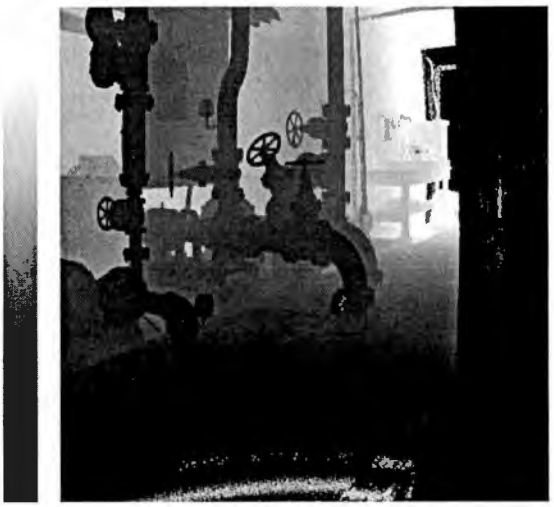
## 6.3 Reconstruction of a 3–D Scene

The partial reconstruction of a 3–D scene from a single range and intensity image pair is the first step in integrating multiple views to form complete 3–D scene models [10] and [37]. Registering and integrating multiple data sets into a common database (possibly in some compressed format) is the goal of more advanced research, which may incorporate the fusion of other types of data as well. This opens the possibility for creating accurate 3–D site models, to be used in managing remote operations within hazardous facilities.

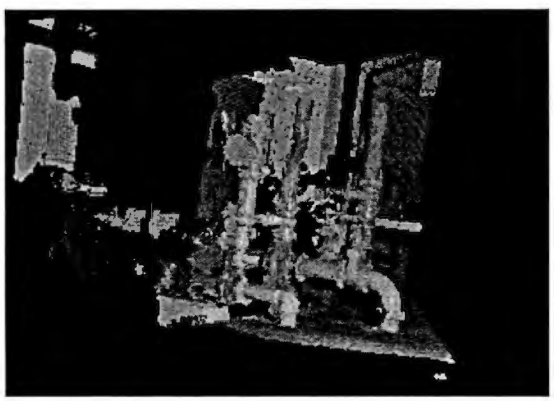
To demonstrate the 3–D nature of the data we recover from range images, we transformed an entire range image from range coordinates into $(x, y, z)$ coordinates. We then overlaid each corresponding $(x, y, z)$ sample point with the correct intensity value from the intensity image and rendered this data in 3–D on a Silicon Graphics workstation (see Figure 48). It must be emphasized that the data used in all three views was obtained from a single range, intensity image pair. Different viewing angles are shown to emphasize the depth feature of the image. The side views, furthermore, illustrate the occlusion concept, where objects in front block the sensor from detecting parts of objects in back. Thus, we conclude that many such images must be taken from different angles to complete an entire 3–D scene model.



Intensity Component

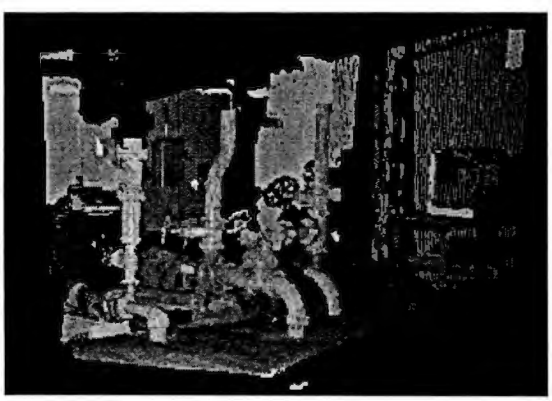


Range Component



(a)

(c)



(b)

Figure 48: 3–D information reconstructed from a range and intensity image pair. (a) Viewed from the left, (b) viewed from the right, and (c) viewed from above.

# 7 SUMMARY AND CONCLUSIONS

Because of a recent shift toward 3–D sensing techniques in machine vision, much of the literature now concerns on the use of range images. However, many applications require accurate $(x, y, z)$ information, especially where many images are used together. Therefore, just as photogrameters developed calibration techniques appropriate for lens based imaging, we propose to study the appropriate mathematics that leads to accurate calibrations for range cameras.

## 7.1 Literature Review

We reviewed the basic concept of what a range image is, how a range camera functions, and previous work in the field of calibration. Some of the literature inspired us to again think the traditional calibration strategy using a precisely placed target for calibration. This is particularly important because the cameras are likely to be placed in an environment so hazardous or radioactive that human intervention with a target is impossible. On this premise, we determined that algorithms which were invariant to target position were the most valuable. Many of the examples are based on geometric features like shape (ie. spheres [26] and planes [4]).

We next investigated the types of non–contact depth measurement techniques used in range cameras. Beginning with ultrasonic pulse detection, we derived equations for range based on signal velocity and travel time. Proceeding, we discussed the need to use smaller wavelengths to increase imaging resolution. The next range measurement system discussed was the laser pulse detection system, used in early range cameras. We then discussed developments in continuous light wave systems such as the amplitude modulation and frequency modulation systems used in the

latest range cameras.

## 7.2 Development of Mathematical Models

The effort of all previous research in calibration of laser scanners seems to be flawed by deficiencies in modeling the internal scanning mechanism. It appears that no one has ever proposed a clear method for modeling the complex interactions of multiple rotating mirrors. Most previous work involving camera calibration pertained to television type cameras. Where there was some work done with range cameras, the predominant mathematical model used was the spherical coordinates system. It was shown, however, that this model broke in near scenery, the exact environment in which most robots function.

A new model using better mathematical tools was called for. We recalled Snell's law from physics, an old law that used 2–D geometry to describe how light rays reflected from a mirror. Unfortunately, the 2–D nature of Snell's law offered little help in solving the general 3–D problem. This specifically led to the introduction of the Householder reflection matrix. This, in addition to other tools developed here provided an analytical approach to modeling not seen before.

Based on the new mathematical tools introduced in Chapter 3, we derived a model for the Perceptron camera to understand the type of calibration errors that could be present. By properly adjusting certain parameters inside the model, we intended to produce an accurate analytic function for transforming range image data into the more useful $(x, y, z)$ data. The task of automatically selecting proper parameter values, however, invoked the development of a complex theory of calibration.

## 7.3 Calibration Hardware and Software with Automatic Target Location

The target and mobile scanner platform required six weeks to construct and will continue to contribute to research at UTK and ORNL. The cart has applications toward the design of an autonomous mapping system. The autonomous mapper is envisioned as a lighter and more sophisticated platform with on–board power, motility, an advanced range camera, plus other sensors for color vision and radiation detection.

The combination of the simplex algorithm and Horn's auto–registration algorithm make the implementation of calibration simple to set up and perform. The need for precise target location is obviated so long as the target itself is precisely manufactured. Furthermore, the quick convergence of the simplex algorithm (less than 100 steps) is also satisfying and indicates the possibility of calibrating the camera on–line from known structures in a facility (an important virtue in nuclear environments where the camera might become contaminated and inaccessible to humans).

## 7.4 Results: Calibration Performance in a Real Experiment

The relatively large errors experienced before camera calibration were reduced to less than 10 mm per target point after calibration. After comparing plots of the detected points against the ideal target, we determined that most residual error results from noisy range measurement, not pointing error.

The above results imply that calibration is limited only by the quality of the

range measuring device inside the camera. The Perceptron unit used in this particular set of experiments has been superseded by a new model. An FM based laser range finder which is both quantitatively and qualitatively superior is available from Coleman.

## 7.5 Applications

The advent of a fully calibrated camera model cast a few of the traditional machine vision problems in a new light. Briefly, a new chapter was established to explore the implications of this new model and its relationship to several problems which had murky underpinning in the past. We primarily questioned the relationship between this new model and the traditional, but incorrect, spherical coordinates model. Astoundingly, our mathematical analysis shows how to derive spherical coordinates as an approximation of our new model under strict assumptions. This reconciliation of the two models was pleasing and lead to a set of conditions for invoking the spherical approximation.

In turn, the newly resurrected spherical coordinates model explains why a curious warping effect occurs in all the range images, resulting in curved vertical lines. This new understanding leads to a dewarping algorithm for machine vision applications which are particularly edge sensitive. One of the examples given is the Hough transform, which detects long, straight line segments in the image space. Performance of this algorithm can be drastically improved by dewarping.

Another topic concerning the spherical coordinates issue is full angle range images. If the range images are indeed in spherical coordinates, each image represents a small segment of the surface of a projection sphere around the camera. We realize that by taking multiple range images from the same location, much of the projec-

tion sphere can be filled. We propose a technique that results in a composite image spanning more than 360 degrees in azimuth, and nearly 160 degrees in elevation. This single range image thus provides a very complete 3–D map of the environment.

Finally, in order to demonstrate the 3–D nature of the data we recover from range images, we transform an entire range image from range coordinates into $(x, y, z)$ coordinates. We then overlay each corresponding $(x, y, z)$ sample point with the correct intensity value from the intensity image and rendered this data in 3–D on a Silicon Graphics workstation. We show different viewing angles to emphasize the depth feature of the image. Furthermore, the side views illustrate the concept of occlusion, where objects in front block the sensor from detecting parts of objects in back. We thus conclude that many such images must be taken from different angles to complete an entire 3–D scene model.

## 7.6 Conclusions

Several problems exist with the amplitude modulated range finders. Such systems are susceptible to many noise sources and become unreliable when the laser return is weak. Particularly, shiny and dark surfaces pose a real problem. Reportedly, the recently developed frequency modulation solves many problems associated with the more primitive amplitude modulation system. The first commercially available frequency modulated system has been developed by Coleman Research Corporation for Oak Ridge National Laboratories and is now available commercially.

The current integration of multiple views into a unified 3–D model is one paradigm for reducing noise since noise mainly affects the range dimension. Thus, when an object is viewed from many angles, errors from previous views can be mitigated. This is a large area of research and is beyond the scope of camera modeling

and calibration.

The experiment performed in Chapter 6 clearly demonstrates the feasibility of recovering reliable, calibrated $(x, y, z)$ surface information from the environment using range images. Future work will focus on using this $(x, y, z)$ data for autonomous navigation, facility mapping, and remote manipulation of hazardous materials.

# References

# References


[1] Anatol N. Golubev and Alexander M. Chekhovshy. "Three–color optical range finder," *Optical Society of America*, 33(31):7511–7517, November 1994.

[2] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour. "Closed–form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America*, 5(7):1127–1135, July 1988.

[3] H. H. Baker and T. O. Binford. Depth From Edge and Intensity Based Stereo. In *Int'l Joint Conf. on Artificial Intelligence*, volume 6, pages 631–636, Vancouver, B.C., Canada, August 1981.

[4] James E. Baker. Automatic calibration of laser range cameras using arbitrary planar surfaces. Technical Report DE–AC05–84OR21400, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, April 1994.

[5] Glenn Beheim and Klaus Fritsch. "Range Finding Using Frequency–Modulated Laser Diode," *J. Applied Optics*, 25, no. 9:1439–1442, May 1986.

[6] Paul J. Besl. Active Optical Range Imaging. In J. L. C. Sanz, editor, *Advances in Machine Vision*, pages 1–53. New York, NY: Springer–Verlag, 1989.

[7] P. Chavel and T. C. Strand. "Range Measurement Using Talbot Diffraction Imaging of Gratings," *J. Applied Optics*, 23, no. 6:862–871, March 1984.

[8] C. H. Chen and A. C. Kak. Modeling and Calibration of a Structured Light Scanner for 3–D Robot Vision. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 3, pages 807–815, 1987.

[9] H. E. Cline, W. E. Lorenson, and A. S. Holik. "Automatic Moire Contouring," *J. Applied Optics*, 23, no. 10:1454–1459, May 1984.

[10] B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. In *Proceedings, SIGGRAPH96.* ACM, 1996.

[11] T. J. Green and J. H. Shapiro. "Maximum–Likelihood Laser Radar Range Profiling with the Expectation–Maximization Algorithm," *Optical Engineering*, 31(11):2343–2354, November 1992.

[12] T. Heikkinen, R. Ahola, M. Manninen, and R. Myllyla. Recent Results of the Performance Analysis of a 3D Sensor Based on Time of Flight Measurements. In *Optical Techniques for Industrial Inspection, Proc. of the SPIE*, volume 665, pages 168–173, June 1986.

[13] J. H. Holland. Genetic Search with Approximate Function Evaluations. In *Proceedings of the International Conference on Genetic Algorithms and Their Applications*, pages 112–120, Pittsburgh, PA, 1985.

[14] A. S. Householder. *Lectures on Numerical Algebra.* Mathematical Association of America, Williamstown, Massechusetts, 1972. Notes on lectures given at the 1972 MAA Summer Seminar William College.

[15] R. A. Jarvis. "A Laser Time–of–Flight Range Scanner for Robotic Vision," *IEEE Trans. Patt. Anal. and Mach. Intell.*, PAMI–5, no. 5:505–512, September 1983.

[16] R. A. Jarvis. "A Perspective on Range Finding Techniques for Computer Vision," *IEEE Trans. Patt. Anal. and Mach. Intell.*, PAMI–5, no. 2:122–139, March 1983.

[17] R. A. Jarvis. "A Perspective on Range Finding Techniques for Computer Vision," *IEEE Trans. Patt. Anal. and Mach. Intell.*, PAMI–5, no. 2:122–139, March 1983.

[18] A. C. Kak. Depth Perception for Robots. In Shimon Y. Nof, editor, *Handbook of Industrial Robotics*, pages 272–319. John Wiley & Sons, Inc., New York, NY, 1985.

[19] Kari Maatta, Juha Kostamovaara, and Risto Myllyla. "Profiling of hot surfaces by pulsed time–of–flight laser range finder techniques," *Applied Optics*, 32(27):5334–5347, November 1993.

[20] Katuo Seta and Tadanao O'ishi. "Distance meter utilizing the intermode beat of a He–Ne laser," *Applied Optics*, 29(3):354–359, January 1990.

[21] Alonzo Kelly. Concept Design of a Scanning Laser Rangefinder for Autonomous Vehicles. Technical Report CMU–RI–TR–94–21, The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, May 1994.

[22] Kung–Li Deng and Jyhpyng Wang. "Nanometer–resolution distance measurement with a noninterferometric method," *Applied Optics*, 33(1):113–116, January 1994.

[23] J. C. Leader. "Speckle Effects on Coherent Laser Radar Detection Efficiency," *Optical Engineering*, 25(5):644–650, May 1986.

[24] M. Takeichi, Y. Warashina, A. Takeshima, I. Ogawa, K. Ichie, and Y. Mizushima. "Streak–camera–based long–distance range finder with $10^{-7}$ resolution," *Applied Optics*, 33(13):2502–2510, May 1994.

[25] J. A. Nelder and R. Mead *Computer Journal*, 7:308–313, 1965.

[26] Jun Ni. Ball–target–based Extrinsic Calibration of a Multiple–axis Laser–stripe Scanning System. 1994.

[27] Odetics, INC. *3–D Laser Imaging System User's Guide.* Anaheim, CA, 1990.

[28] P. J. Besl. *Surfaces in Range Image Understanding*, pages 311–316. Springer–Verlag, New York, NY, 1986.

[29] Perception INC. *Product Information.* Farmington Hills, MI, 1987.

[30] D. B. Reister and M. D. Morris. A Method for Obtaining the Least Squares Fit of a Hyperplane to Uncertain Data. Technical Report ORNL/TM–12686, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, 1994.

[31] M. Rioux. "Laser Range Finder Based Upon Synchronized Scanners," *J. Applied Optics*, 23, no. 21:3837–3844, November 1984.

[32] M. Rioux and F. Blais. "Compact Three–Dimensional Camera for Robotic Applications," *J. of the Optical Soc. of America*, A3, no. 9:1518–1521, September 1986.

[33] S. F. Collins, W. X. Huang, M. M. Murphy, K. T. V. Grattan, and A. W. Palmer. "Ranging measurements over a 20 metre path using an intensity–chirped laser diode," *Measurement Science and Technology*, 5(6):753–755, June 1994.

[34] O. Sasaki and H. Okazaki. "Analysis of Measurement Accuracy in Sinusoidal Phase Modulating Interferometry," *J. Applied Optics*, 25, no. 18:3152–3158, September 1986.

[35] O. Sasaki and H. Okazaki. "Sinusoidal Phase Modulating Interferometry for Surface Profile Measurements," *J. Applied Optics*, 25, no. 18:3137–3140, September 1986.

[36] Irwin Sobel. "On Calibrating Computer Controlled Cameras for Perceiving 3–D Scenes," *Artificial Intelligence*, 5:185–198, May 1974.

[37] M. Soucy and D. Laurendeau. "A General Surface Approach to the Integration of a Set of Range Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):344–358, April 1995.

[38] Kai Storjohann and Eric Saltzen. Laser Range Camera Calibration. Technical Report DE91–014939, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, July 1991.

[39] Richard Szeliski. Image Mosaicing for Tele–Reality. Technical Report CRL 94/2, Digital Equiptment Corporation, Cambridge Research Laboratory, Cambridge, Massachusetts 02139, May 1994.

[40] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*, pages 408–412. Cambridge University Press, New York, NY, 1988.

# VITA

Brian Lee Chase was born in Kingsport, Tennessee on December 30, 1969. His family lived just outside of Kingsport, in the suburb of Colonial Heights, where Mr. Chase attended Rock Springs Elementary School, Colonial Heights Middle School, and Sullivan South High School in Sullivan County, Tennessee. He graduated 12th in his high school class in 1988.

In Autumn of 1988, Mr. Chase moved to Cookeville, Tennessee to attend Tennessee Technological University. He was awarded a B.S. degree in Electrical Engineering in December 1993, graduating with honors. During his undergraduate studies, he participated in the cooperative engineering program, gaining valuable work experience at DuPont, in Nashville, Tennessee.

Mr. Chase entered the graduate program in Electrical Engineering at the University of Tennessee in January, 1994. As a graduate research assistant with the Imaging, Robotics, and Intelligent Systems Laboratory, he focused his research efforts on range camera calibration and range image processing for 3–D data extraction. He was actively involved with the DOE/OST/RTDP's University Research Program in Robotics (Universities of Florida, Michigan, New Mexico, Tennessee, and Texas) under grant DOE–DE–FG02–86NE37968. After graduation, Mr. Chase was employed by Perceptics corporation, in Knoxville, Tennessee. Perceptics specializes in optical character recognition and machine vision applications. In September 1999, Mr. Chase accepted a position at ASE, in Knoxville, Tennessee. ASE produces set top boxes for satelite and digital video applications.

This thesis was typed on both Sun and Silicon Graphics workstations. The text and graphics were formatted by the Latex document preparation system with the BibTex bibliography program, and the dvips PostScript driver.