Development of a Versatile Wide-angle Lens Characterization Strategy for Use in the OMNIster Stereo Vision System

Keith B. Johnson

University of Tennessee - Knoxville

Recommended Citation

This Thesis is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Masters Theses by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.
To the Graduate Council:

I am submitting herewith a thesis written by Keith B. Johnson entitled "Development of a Versatile Wide-angle Lens Characterization Strategy for Use in the OMNIster Stereo Vision System." 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:

P. W. Smith, R. T. Whitaker, W. L. Green

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 Keith B. Johnson entitled “Development of a Versatile Wide-angle Lens Characterization Strategy for Use in the OMNister Stereo Vision System.” 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:

[Signatures]

Accepted for the Council:

[Signature]
Associate Vice Chancellor
and Dean of The Graduate School
DEVELOPMENT OF A VERSATILE WIDE-ANGLE LENS
CHARACTERIZATION STRATEGY FOR USE IN THE
OMNISTER STEREO VISION SYSTEM

A Thesis
Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Keith B. Johnson
December 1997
This thesis details the development of an accurate and efficient wide-angle stereo vision system. Wide-angle or fisheye stereo is desired because it provides the capability to recover depth information for a large scene from a single stereo image pair. However, nonlinear image distortions caused by the camera optics complicate the necessary stereo processes of camera modeling and disparity analysis. The characterization and removal of these lens distortions therefore is considered vital to stereo evaluation of fisheye images. Existing wide-angle stereo systems have maintained the use of pinhole projections to model the respective camera systems. This ideal projection model does not parametrize lens distortion, and as a result, distortions must be described using a highly nonlinear error function. Systems which incorporate high-order polynomial point mappings, however, have failed to provide accurate distortion description and correction throughout the system's field-of-view. Thus, the field-of-view advantage of the wide-angle vision system is reduced. This work initially investigates the characterization of nonlinear wide-angle distortions using the spherical lens projection model which inherently describes the existence of radial distortions within its perspective transformations. Although this physical distortion characterization of the spherical lens model is computationally efficient, it proves inaccurate when removing typical lens distortions. As a result, a more general lens characterization based conceptually on the framework of the spherical lens model, is developed to more accurately describe wide-angle lens distortions. More importantly,
this lens characterization strategy provides the framework which is used to develop the OMNIster wide-angle stereo vision system. This novel system avoids the customary methods of wide-angle stereo which require complete correction of the image pair prior to stereo analysis. Instead, a correlation search strategy is developed that defines the nonlinear epipolar search constraints between the distorted image pairs. Further, the algorithm is tested in a controlled stereo setup using both nonlinear lens characterization models, and the accuracy of the depth measurements of each are compared.
ACKNOWLEDGMENTS

I would first like to thank my parents, Peggy and Carey Johnson, for their continuous and avid encouragement and support throughout the years while I worked towards my educational goals. I also would like to extend my sincere gratitude to my new wife, Heather, for her untiring devotion, love, and patience during the long hours spent in research. A further acknowledgment is due to my grandfather, Randolph Johnson, who recently passed away, for his ever-present love and support. And finally, I wish to thank my advisors, Dr. P. W. Smith, Dr. W. L. Green, and Dr. M. A. Abidi for their guidance and assistance throughout my program. Appreciation also to the members of my committee, Dr. R. T. Whitaker, Dr. P. W. Smith, Dr. W. L. Green, and Dr. M. A. Abidi, for their help and constructive criticism.

The work in this thesis was supported by the DOE's University Research Program in Robotics (Universities of Florida, Michigan, New Mexico, Tennessee, and Texas) under grant DOE–DE–FG02–86NE37968. Additional support was provided by Mechanical Technology Incorporated and the U.S. Department of Energy Federal Energy Technology Center under grant DE–AR21–95MC32093.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview of Chapter Contents</td>
<td>10</td>
</tr>
<tr>
<td>2. Lens Model Descriptions</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Conventional Pinhole Camera Model</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Stereo Depth Estimation Vector Geometry</td>
<td>14</td>
</tr>
<tr>
<td>2.3 The Spherical Lens Model</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Spherical Lens Projection Mathematics</td>
<td>20</td>
</tr>
<tr>
<td>3. Stereo Evaluation of the Spherical Lens Model</td>
<td>25</td>
</tr>
<tr>
<td>3.1 Dewarping Evaluation</td>
<td>25</td>
</tr>
<tr>
<td>3.1.1 Choice of Optics</td>
<td>26</td>
</tr>
<tr>
<td>3.1.2 Test Procedure</td>
<td>27</td>
</tr>
<tr>
<td>3.1.3 Accuracy</td>
<td>27</td>
</tr>
<tr>
<td>3.2 A Stereo Test Analysis</td>
<td>31</td>
</tr>
<tr>
<td>3.2.1 Stereo Results</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Conclusions</td>
<td>36</td>
</tr>
<tr>
<td>4. An Enhanced Distortion characterization Model</td>
<td>37</td>
</tr>
<tr>
<td>4.1 Simplification of the Spherical Lens Model</td>
<td>38</td>
</tr>
<tr>
<td>4.1.1 Choosing the Spherical Model Correction Factor</td>
<td>43</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.2 Description of a General Wide-angle Lens Model</td>
<td>47</td>
</tr>
<tr>
<td>4.3 Lens Surface Model Characterization</td>
<td>52</td>
</tr>
<tr>
<td>4.4 Dewarping Implementation</td>
<td>57</td>
</tr>
<tr>
<td>4.5 Statistical Error Analysis</td>
<td>62</td>
</tr>
<tr>
<td>4.6 Conclusions</td>
<td>65</td>
</tr>
<tr>
<td>5. OMNister: an Omnidirectional Stereo Vision System</td>
<td>67</td>
</tr>
<tr>
<td>5.1 Distorted Epipolar Correlation Strategy</td>
<td>68</td>
</tr>
<tr>
<td>5.2 Stereo Test Procedure and Results</td>
<td>77</td>
</tr>
<tr>
<td>5.3 Conclusions</td>
<td>82</td>
</tr>
<tr>
<td>6. Concluding Remarks</td>
<td>83</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>86</td>
</tr>
<tr>
<td>VITA</td>
<td>89</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>6</td>
</tr>
<tr>
<td>2.1</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>18</td>
</tr>
<tr>
<td>2.4</td>
<td>19</td>
</tr>
<tr>
<td>2.5</td>
<td>21</td>
</tr>
<tr>
<td>2.6</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>28</td>
</tr>
<tr>
<td>3.2</td>
<td>29</td>
</tr>
<tr>
<td>3.3</td>
<td>31</td>
</tr>
<tr>
<td>3.4</td>
<td>32</td>
</tr>
<tr>
<td>3.5</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>40</td>
</tr>
<tr>
<td>4.2</td>
<td>42</td>
</tr>
<tr>
<td>4.3</td>
<td>45</td>
</tr>
<tr>
<td>4.4</td>
<td>46</td>
</tr>
<tr>
<td>4.5</td>
<td>48</td>
</tr>
<tr>
<td>4.6</td>
<td>49</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>4.7 A cross section of the new lens model</td>
<td>52</td>
</tr>
<tr>
<td>4.8 Lens surface calibration board</td>
<td>54</td>
</tr>
<tr>
<td>4.9 Plots of the axial change in lens radius</td>
<td>56</td>
</tr>
<tr>
<td>4.10 Plots of the reconstructed lens surface along respective axes</td>
<td>57</td>
</tr>
<tr>
<td>4.11 Inventor reconstructions of the quadric and spherical lens models</td>
<td>58</td>
</tr>
<tr>
<td>4.12 Forward dewarping using the quadric surface lens model</td>
<td>60</td>
</tr>
<tr>
<td>4.13 Lens surface model in undistorted image coordinates</td>
<td>61</td>
</tr>
<tr>
<td>4.14 Final dewarping results using back-projection method</td>
<td>63</td>
</tr>
<tr>
<td>4.15 Error plots of correction results</td>
<td>64</td>
</tr>
<tr>
<td>5.1 Many-to-one mapping of undistorted to distorted projection</td>
<td>71</td>
</tr>
<tr>
<td>5.2 The formation of the curved epipolar search path</td>
<td>72</td>
</tr>
<tr>
<td>5.3 Exemplary correlation windows from right/left image pairs</td>
<td>74</td>
</tr>
<tr>
<td>5.4 Diagram of a method of distorted correlation</td>
<td>75</td>
</tr>
<tr>
<td>5.5 Diagram of the <em>warping window</em> stereo correlation strategy</td>
<td>76</td>
</tr>
<tr>
<td>5.6 Stereo reconstruction of the planar surface using spherical lens model</td>
<td>78</td>
</tr>
<tr>
<td>5.7 Stereo reconstruction of test board using the quadric lens model</td>
<td>78</td>
</tr>
<tr>
<td>5.8 Plot of the x-axis depth measurement errors</td>
<td>79</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

Computer vision techniques play a significant role in many applications such as robotics, automation, and remote sensing for automatic vehicle guidance. They enable automated systems to understand their environments using visual information. For many applications the primary goal of the computer vision system is the acquisition of three-dimensional scene information. One of the most widely used methods for gathering depth information from a scene is stereo vision, since stereo vision can provide accurate, efficient distance measurements over a large range of depths using off-the-shelf camera systems. Intuitively, stereo is the simplest three-dimensional vision method to understand [1], since it is regarded as the most important way in which humans capture depth information [2]. As a result, researchers have attempted to imitate this visual process using cameras for the purpose of enabling computers to "see."

Because of its passive nature, this stereo-based depth estimation method using triangulation has a unique advantage over active sensing techniques in many applications where intrusive ranging methods cannot be applied. In stereo vision systems, depth to a world point is calculated by measuring the disparity between the two dimensional imaged positions of the point in a stereo pair of images taken from disparate locations. Since a single 3-D point will project differently onto a camera's
sensor when imaged from different locations, the 3-D world position of the point is reconstructed using the geometric technique called triangulation.

The effectiveness of a stereosystem is often measured by the system’s performance in a wide variety of situations. Conventional stereo systems, however, have been limited by field-of-measurement and scene modeling efficiency, since camera systems traditionally utilized for stereo imaging possess relatively narrow viewing angles. This small field-of-view is necessary to maintain an approximately rectilinear, or pinhole, perspective of the imaged scene and thus simplify the stereo correspondence and range calculation algorithms. A narrow field-of-view, however, reduces the area in the scene that can be measured for depth from a single stereo position. As a result, the distance between cameras must remain relatively small, limiting the amount of useful depth information gathered from a single stereo pair of images. Therefore, in order to reconstruct large scenes or model close-up objects, multiple stereo sensors are required or repositioning of the entire stereo setup must be performed to obtain the needed depth information. In figure 1.1, only a small portion of the scene (a model of an industrial setting) is imaged using an ordinary rectilinear camera. Thus, several stereo image pairs would be required to capture the entire model demanding painstaking repositioning of the camera into successive positions. Alternatively, the same camera can capture the entire model if a wide-angle lens is employed, as is shown in figure 1.2.

Another approach might be to use an expensive, highly accurate orientation mechanism to redirect the pose of the camera system. For instance, in research by Ishiguro et al. [3], a 360° omni-directional stereo system was described that uses a single cam-
Figure 1.1: Field of view limitation of typical stereo camera systems. This image taken with a regular 35mm lens demonstrates the reduced field-of-view exhibited in common camera systems that are commonly used for stereo vision.
Figure 1.2: An image taken using a fisheye lens. Images such as this one can provide up to three to four times the field-of-view of ordinary rectilinear cameras. However, the inherent lens distortions make processing generally difficult.
era mounted with offset to a rotating axis. Stereo images are generated using a single camera system with two vertical slits. Each slit, one pixel in width, forms a single panoramic view as the camera swivels by piecing together each of the individual imaged slits. Therefore, the images are created with a disparate baseline equivalent to the distance between slits. For accurate results, this technique requires a rotary device with very high precision; Ishiguro claims a need for an angular resolution of 0.005 degrees which is very difficult to achieve. Another omnidirectional stereo system by Benosman et al., [4] uses a method very similar to one described by Ishiguro. However, in this approach, high resolution line sensors are used to create the cylindrical panoramic views. However, since camera scanning is used in these systems, a simultaneous viewing capability is not provided. These methods, therefore, may not be applicable in situations where immediate and simultaneous stereo viewing of the environment is required such as the monitoring of hazardous materials.

Due to a need for simultaneous “whole world” viewing [5], the use of wide-angle/fisheye optics for stereo has been investigated by several researchers. Images obtained using wide-angle optics provide a simple method of recording a near $2\pi$ steradian scene without camera scanning. Figure 1.2 shows a typical image taken with fisheye optics. “Omnivision,” as this ability to view in very wide fields has been termed, yields significant advantages for both robot navigation and three-dimensional scene reconstruction [5]. Difficult positional calibrations and setup procedures are reduced by the elimination of mechanical orientation devices for repositioning the stereo system. Furthermore, complete depth measurement recovery of a large scene is afforded from the wide-angle perspective which is available from a
single stereo pair of fisheye images.

Although significant advantages seemingly result from the use of wide-angle optics in a stereo system, such benefits cannot be realized without considering additional problems. The distortion evidenced in fisheye images is a serious hindrance to the general application of an omnivision stereo system. For example, figure 1.3 helps demonstrate the difficulties which result in stereo analysis using fisheye stereo image pairs. First, a linear epipolar relationship between horizontal image pairs is non-existent. That is, "no simple relationship exists in the left-right stereo pair" [6] with regards to expected image feature locations. And second, corresponding features in the two images are no longer similar in shape or intensity. This complicates the automatic point or feature matching task of the stereo system. Processing of fisheye images for stereo applications, therefore, requires accurate characterization of the
lens distortions to retain the linear epipolar geometry and feature similarity traditionally required between stereo pairs of images, making image distortion correction an important task for wide-angle stereo systems [7].

Several researchers have described methods for wide-angle or fisheye stereo vision. Interestingly, only a few have devised fisheye computer vision methods which avoid the difficult task of image restoration. In research by Cao et al. [5], a simple technique is devised that uses the imaged locations of three reference beacons to describe the characteristic distortions. The known horizontal relationship between the three beacons is used as the basic input data in the positioning computation. Another novel method described by Morita et al. [8], uses a spherical mapping method to fit a great circle to a projected linear feature; this method similar to the Hough Transform. A line made up of several points is transformed and concentrated at a single point, or pole. The vector extending from the center of the modeled sphere (fisheye lens) to the pole is parallel to the linear imaged feature in three-dimensional space; the direction of the line can then be inferred. Thus, a method of finding the three-dimensional location of lines in a scene from a stereo pair of fisheye images is obtained without distortion correction.

In general, however, removal of the nonlinear distortions is deemed necessary for traditional stereo analysis of wide-angle images. Restoration of stereo fisheye images, for example, is accomplished by Onoe [6] using a priori information of a stereo imaged scene of buildings. The process described is based on a geometrical transformation of points on the same half radius in the fisheye lens image. By knowing the approximate depth from the camera to an imaged roofline and the half-
radius representation of that roofline in the image, a transformation is constructed to provide a reasonable restoration for stereo analysis.

For a more accurate correction of wide-angle distortion, other researchers have developed highly nonlinear point to point mapping strategies which describe a direct mapping of image coordinates to their undistorted locations. This mapping attempts to characterize radial distortion as error from an ideal projection characteristic of the pinhole camera, a model traditionally used to describe the point projections of cameras in a stereo vision system. Several methods have evolved to calibrate the high order polynomials needed to describe this point mapping. A line straightness method discussed by Prescott and McLean [9] uses a routine which iteratively tests the distortion model coefficients to evaluate the straightness of imaged linear features after correction. Nomura, et al[10], define the correction mapping as separate coordinate functions by utilizing a point symmetry trait of the image distortion to decompose an ordinary 2-D model fitting into two 1-D fittings on the columns and rows of an image. Shah and Aggarwal [11] demonstrate two high-order polynomial transforms to describe both the radial mapping and angular correction of a point to an undistorted location. None of these techniques have attempted to give the lens surface a physical characterization, and thus, rely solely on a point-to-point calibrated polynomial mapping. High order polynomials are very sensitive to over-fitting near data limits and the ability of this mapping to properly correct in the image extremes is not clear and has not been well-evaluated by previous researchers. Therefore, avoiding this difficult and sensitive polynomial distortion correction is necessary for accurate and efficient wide-angle stereo reconstruction throughout the entire field-of-view.
Lens distortions are not described in the pinhole camera model except by means of high order mapping functions which require calibration of the distortion parameters of the system. However, a nonlinear projection model such as the spherical lens model will describe the distortive behavior of the lens as a natural consequence of the projection. For instance, Zimmermann [12] develops an efficient dewarping algorithm based upon the spherical lens model in his development of the OMNiview motionless camera system. In this real-time video monitoring system, the properties of the spherical lens model are employed to describe the perspective transformations necessary for correcting fisheye lens distortions as a function of a single parameter - the lens radius. As a result, the point-to-point polynomial mapping used to describe lens distortion is replaced by a simple correction equation based on one constant parameter. This provides a simple means of correcting for distortions in a pair of fisheye images, and thus, the acquisition of stereo depth measurement is afforded utilizing traditional pinhole stereo geometry. For instance, in early research by Walsh et al. [13], the use of the spherical lens model based OMNiview system for stereoscopic triangulation control of a robot is investigated. Range measurement results using this teleoperated OMNiview stereo system are not provided. However, in the report he outlines the limitations on the system’s accuracy as (1) the quality fisheye lenses, (2) the accuracy of the lens radius value for the model which controls the correction of nonlinear distortions, and (3) the knowledge of camera’s setup. Although findings on the accuracy and usefulness of this OMNiview stereo system are inconclusive, the development of the system demonstrates the applicability of the spherical lens model for describing nonlinear wide-angle lens distortions. This
knowledge will provide a basis for the wide-angle stereo research developed in this research.

However, as a result of this limitations to the system accuracy described by Walsh et al, this work will employ some of the lens projection used by the ideal spherical lens model to develop a more general description of the typical distortions characteristic of actual wide-angle lenses. By allowing for non-spherical lenses, this enhanced projection model provides for descriptions of general surfaces of projection for a particular fisheye lens. The physical characterization of the nonlinear projections will allow for an accurate and efficient stereo implementation that eliminates the previous need to correct image distortions prior to stereo analysis, enhancing system accuracy and processing efficiency for high resolution wide-angle stereo scene reconstruction.

1.1 Overview of Chapter Contents

An omnidirectional stereo vision system is developed, implemented, and evaluated in the following chapters. In Chapter 2, the necessary models for describing an ideal wide-angle stereo vision system are described. More specifically, this chapter provides the basic pinhole model stereo geometry used for general depth estimation, and it details the development of the distortion correction transformation equations that are described by the spherical lens model. Chapter 3 assesses the projection accuracy of the spherical lens model by evaluating the OMNIview camera system. First, an analysis of the system’s dewarping algorithm is performed to characterize the errors associated with the correction of wide-angle lens distortions. Finally, the
section provides the results, with an accuracy evaluation, of a simple stereo test using
the OMNIview system. The next chapter develops the transformations for describing
the distortions due to general nonlinear surface projections. This includes a develop­
ment of transformation equations, a description of a structured lens characterization
routine, and an exhibition of the final distortion correction results. Chapter 5 then
develops the final stereo vision system, “OMNIster”, and demonstrates the system’s
scene reconstruction results with comparison to the spherical model. A main feature
in this chapter and the stereo development is the implementation of a novel epipo­
lar search path characterization strategy and point matching algorithm. The final
chapter summarizes the development of OMNIster from a simple routine based on
the principles of the spherical lens model to the final omnidirectional stereo vision
system.
CHAPTER 2

Lens Model Descriptions

When establishing a camera-based 3-D measuring system, an essential task is to describe a simple and accurate projection model of the imaging system. The development of an accurate perspective transformation is necessary for describing the projection of a world point onto the camera’s sensor plane. Knowledge of this transformation forms the foundation for inversely relating an image pixel to a three-dimensional world location. Although an image point cannot uniquely determine the location of a corresponding world point, the missing depth information can be obtained using stereoscopic techniques, or stereo vision as described in chapter 1. For an ideal camera system, the pinhole camera model provides a very simple relationship for obtaining stereo depth measurements. However, when wide-angle optics are used in the stereo system, this geometry is complicated due to the nonlinear projections characteristic of the lens system. As a result, in order to maintain the pinhole stereo projection geometry, these nonlinear distortions must be characterized and removed. Since the pinhole model has no intrinsic parameterization of this nonlinear projection, a different lens model will be investigated in this research to describe wide-angle distortions...the spherical lens model. From this nonlinear projection model, an algorithm will be developed that naturally characterizes distortion in fisheye images and provides, in turn, the pinhole image from which the simple stereo geometrical
relationships can be obtained.

2.1 Conventional Pinhole Camera Model

Optical systems with disparate locations will image an object differently depending on the distance of that object from the lens. By relating that image disparity from two known camera locations through an appropriate projection model, one can ascertain depth to the point. As a result, the first step in developing the projection mathematics for a stereo system is to build a camera/lens model. The simplest model is undoubtedly the pinhole camera model. In this camera model, all world coordinate projections are linear and pass through the lens center. Figure 2.1 depicts this projection of a point in an object plane onto the sensor. Therefore, reconstruction of a world point’s direction vector is easily performed once the point’s image location and the camera’s intrinsic parameters such as pixel scale and focal length are known.

A simple mathematical means exists for calculating the depth to an object when the two cameras are positioned such that they are separated by a known distance and their sensor planes are coplanar. Since all projections are linear and the three-dimensional locations of the lens centers are known or can be determined, one can employ simple trigonometry from two camera positions to acquire the 3-D location of the point of interest in this case. The general stereo mathematics used for depth estimation are shown below in equation 2.1,

$$z = \frac{f \times b}{(x_2 - x_1)}$$  \hspace{2cm} (2.1)

where \(f\) is the focal length of the camera, \(b\) is the measured distance between the
Figure 2.1: Diagram of the pinhole camera model. This camera model is generally used for stereo scene reconstruction due to its simple geometrical relationships. Characterizing the perspective transformations of the stereo camera(s) using this model is a major goal in calibrating the system.

centers of the camera lenses or baseline, and $x_2 - x_1$ is the unit length disparity between the point locations in the two sensor planes. A more detailed development of the general stereo mathematics is given by Gonzalez and Woods [14].

2.2 Stereo Depth Estimation Vector Geometry

If a strict linear positional constraint is not maintained, as may be described by wide-angle distorted image pairs, intersection of the respective camera projection vectors cannot be guaranteed. Therefore, a more versatile means of calculating an objects three-dimensional coordinate location is desired to account for inaccuracies in the vector intersection [13]. Vector calculus provides the techniques necessary for solving for the nearest points of intersection when no true intersection exists. Figure
Figure 2.2: Stereo triangulation geometry. The projection of two direction vectors are shown in the diagram above. Using vector analysis techniques the position vectors \( P(a) \) and \( Q(b) \) exemplifying the world points of nearest intersection can be found.

2.2 depicts the vector relationships.

Using the respective camera model (the pinhole model is demonstrated) two direction vectors are known: \( d\vec{P} \) and \( d\vec{Q} \). The position vectors formed by their intersection are the unknowns.

\[
P(a) = P_0 + a \cdot d\vec{P} \quad (2.2)
\]

\[
Q(b) = Q_0 + b \cdot d\vec{Q} \quad (2.3)
\]

Defining \( \vec{S} \) to be orthogonal to both \( d\vec{P} \) and \( d\vec{Q} \), the dot product relationship of the two vectors to \( \vec{S} \) is then zero. Therefore,

\[
d\vec{P} \cdot (Q(b) - P(a)) = 0 \quad \text{and} \quad (2.4)
\]

\[
d\vec{Q} \cdot (Q(b) - P(a)) = 0 \quad (2.5)
\]
Expanding and using Cramer's Rule, we get

\[
\begin{vmatrix}
\bar{d}P \cdot \bar{d}P & -\bar{d}P \cdot \bar{d}Q \\
\bar{d}Q \cdot \bar{d}P & -\bar{d}Q \cdot \bar{d}Q
\end{vmatrix}
\begin{pmatrix}
a \\
b
\end{pmatrix}
= \begin{vmatrix}
\bar{d}P \cdot (Q_0 - P_0) \\
\bar{d}Q \cdot (Q_0 - P_0)
\end{vmatrix}
\tag{2.6}
\]

Solving for \(a\) and \(b\),

\[
a = \frac{\begin{vmatrix}
\bar{d}P \cdot (Q_0 - P_0) & -\bar{d}P \cdot \bar{d}Q \\
\bar{d}Q \cdot (Q_0 - P_0) & -\bar{d}Q \cdot \bar{d}Q
\end{vmatrix}}{A} \tag{2.7}
\]

\[
b = \frac{\begin{vmatrix}
\bar{d}P \cdot \bar{d}P & \bar{d}P \cdot (Q_0 - P_0) \\
\bar{d}Q \cdot \bar{d}P & \bar{d}Q \cdot (Q_0 - P_0)
\end{vmatrix}}{A} \tag{2.8}
\]

where

\[
A = (\bar{d}Q \cdot \bar{d}P)(\bar{d}P \cdot \bar{d}Q) - \|\bar{d}P\|^2\|\bar{d}Q\|^2 \tag{2.9}
\]

Thus, equations 2.2 and 2.3 are used to calculate the points of nearest intersection. If perfect intersection is not acquired, the location of the target world point becomes the average of the two position vectors. This method of defining the 3-D point of intersection will be used in the stereo analysis conducted throughout this research.

### 2.3 The Spherical Lens Model

For camera systems which have traditionally been used for stereo, the pinhole camera model has been sufficient for modeling the perspective transformations. However, as the viewing angle of a lens increases, the projection of a point deviates from the linear type that is characteristic of the pinhole model, with nonlinear distortions
becoming more evident. Therefore, in order to maintain the use of the pinhole camera model representation for describing the camera imaging transformations, these non-linear distortions must be characterized and removed. Once the projection properties of a spherical lens are modeled, a transformation from a fisheye view to a pinhole characteristic representation is defined.

The fisheye lens, with a field-of-view of 180°, provides a circular view of a hemispherical region. Within this viewing area, a "barrel-warped" distortion exists in that horizontal and vertical lines tend to be mapped into circles as the direction of view extends to angles far off the optical axis. Figure 1.2 shows an image taken using such a lens. This image does not provide a full 2π steradian view because of the limited size of the camera's CCD sensor. However, for stereo research this is acceptable and somewhat desirable due to the significant loss of resolution at viewing angles near 180°. Figure 2.3 shows how the image of figure 1.2 was formed.

The perfect fisheye lens can be modeled as a sphere through which scene projections are described by two basic properties. First, the field-of-view encompasses 2π steradians and produces a circular image so that distortions are symmetrical about the image center. Second, the fisheye lens possesses an infinite depth-of-field in that all objects in the image are in focus. Furthermore, the formation of nonlinear image distortion is governed by two postulates, the azimuth angle invariability and the equidistant projection rule. These postulates describe the projection of object points onto the sensor and will directly affect the dewarping algorithm that will be subsequently developed.

The first postulate, the azimuth angle invariability, governs the projection of
Figure 2.3: Sensor and fisheye image projection relationship. The fisheye image shown previously in Figure 1.2 is actually a limited view of the circular image typically characteristic of the fisheye image. This limited view of a fisheye image is used to maximize the resolution capabilities of the viewing system. Fisheye images which contain the full hemispherical view leave much of the image sensor unused.
Figure 2.4: Diagram of the Azimuth Angle Invariability Postulate. Here, all points contained in the content plane are projected onto the same line formed by the intersection of the content plane and the sensor plane.

points lying in the plane that passes through the optical axis, perpendicular to the sensor plane, as illustrated in figure 2.4. This surface is termed the content plane. The postulate states that all such object points are mapped along the radial line created by the intersection of the sensor plane and the content plane. In figure 2.4, object points P1, P2, and P3 are contained in the same plane and are separated by only height and distance. The azimuth angle, delta ($\delta$), of the projection of each of these points is always the same. Therefore, the azimuth angle of the object points and their projections onto the sensor remain unchanged due to differences in the object distance or elevation within the content plane.
The equidistant projection rule, the second postulate, describes the relationship between the radial distance of an image point in the sensor plane to the zenith angle created by the vector from the image center to the world object point as defined in figure 2.5. This rule states that for a spherical lens a linear relationship exists between the center to image point radial distance, $r$, and the zenith angle, $\beta$.

This relationship is as follows:

$$r = k \beta$$

(2.10)

where $k$ is a constant. As the zenith angle varies from 0 to 90 degrees, the radial distance of the corresponding image point varies linearly from 0 to a maximum value $R$, determined by the modeled sphere's radius. The mathematics related to these governing postulates and the fisheye perspective transformations will be detailed in the following section.

### 2.4 Spherical Lens Projection Mathematics

Using the properties and postulates presented previously, the development of the mathematical transformations that describe the fisheye distortions can be easily obtained. Although not re-investigated here, these mathematical transformations will be restated for convenience. Additional background can be found in [12]. The transformations, in general, describe a planar rotation about each directional axis (the z-axis being along the optical center) and a normalized projection of an object plane through a hemispherical surface. The coordinate reference frame representing the mathematical transformations is shown in figure 2.6 and should be referred to as
Figure 2.5: Diagram of the Equidistant Projection Rule. This rule maintains that a linear relationship exists between the angle of incidence $\beta$ and the radial distance of its projection onto the sensor.
Figure 2.6: Diagram of the coordinate reference frame depicting the projection of data in an object plane through a fisheye lens and onto the sensor plane.

the equations are presented. In this reference frame, the image plane is represented by the \((x,y)\) coordinate system. The \textit{Image Object Plane} \((u,v)\) contains the undistorted pinhole image data prior to projection through the fisheye lens. The important relationships are given in the following:

\[
x = \frac{R \ast [uA - vB + mR \sin \beta \sin \delta]}{\sqrt{u^2 + v^2 + m^2R^2}} \tag{2.11}
\]

\[
y = \frac{R \ast [uC - vD - mR \sin \beta \cos \delta]}{\sqrt{u^2 + v^2 + m^2R^2}} \tag{2.12}
\]

where
\[ u, v = \text{object plane coordinates} \]
\[ x, y = \text{image sensor plane coordinates} \]
\[ R = \text{radius of the image circle} \]
\[ \beta = \text{zenith angle} \]
\[ \delta = \text{Azimuth angle in the image sensor plane} \]
\[ \theta = \text{Object plane rotation angle} \]
\[ m = \text{Magnification factor} \]

and

\[ A = (\cos \theta \cos \delta - \sin \theta \sin \delta \cos \beta) \]
\[ B = (\sin \theta \cos \delta + \cos \theta \sin \delta \cos \beta) \]
\[ C = (\cos \theta \sin \delta + \sin \theta \cos \delta \cos \beta) \]
\[ D = (\sin \theta \sin \delta - \cos \theta \cos \delta \cos \beta) \]

These equations describe the projection of data from an object plane through a fisheye lens and onto the camera sensor. Not shown on the diagram is the distance from the center of the sensor plane along the direction of view (DOV) to the object plane origin. This distance is the effective lens radius of the spherical model multiplied by the magnification or “zoom” factor \( m \).

\[ ||DOV|| = mR \quad (2.13) \]

This scaled radius parameter of the modeled sphere controls the amount of distortion described by the system. Therefore, by accurately choosing this radius factor, the
inherent distortion of the wide-angle lens can be properly modeled and subsequently removed. This *dewarping* process will be detailed in Chapter 4.
CHAPTER 3

Stereo Evaluation of the Spherical Lens Model

A method of wide-angle lens distortion characterization is described in the previous chapter which defines a nonlinear projection through a spherical surface. To assess the accuracy of this lens model's characterization of a particular wide-angle lens system, the OMNIview system previously described in Chapter 1 will be used. The OMNIview motionless camera system employs the projection transformations described by the spherical lens model to seemingly provide for distortionless viewing throughout a hemispherical region in realtime without physical motion of the camera system. As a result, this system offers an efficient method of testing the correction accuracy of the spherical lens model for a stereo vision implementation. This chapter will first evaluate the distortion correction performance of the lens model and the accuracy of the stereo reconstruction results obtained when eliminating distortions using the OMNIview system and spherical lens projection model.

3.1 Dewarping Evaluation

Correction of image distortions is considered essential for the development of an accurate wide-angle stereo system. As a result, before implementation of a spherical lens model based stereo system, the accuracy of the model's dewarping characteristics will be tested using the OMNIview video monitoring system. The goal of this
evaluation is to perform a complete error analysis and provide a best dewarping pa-
rameter $R$ for a particular camera and lens. In previous experiments by Walsh et al.,
mention was made as to the significant emphasis placed on quality lens choice and
the selection of a proper dewarping factor constant $R$ in maintaining a high level of
accuracy in their vision system. That is, lenses that cannot be accurately modeled by
a spherical lens model, which is the case with all wide-angle lenses, cannot be cor-
rected completely and accurately by the OMNiview system. Tests demonstrated in
this research will show how the overall accuracy and effectiveness of a stereo vision
system is reduced by limitations of the spherical lens model.

3.1.1 Choice of Optics

When deciding on the wide-angle optics for the OMNiview system, it is best to choose
a quality lens which most closely approximates the ideal fisheye lens. However, as
with all optics, fabrication of a perfect lens is impossible. As a result, since OMNiview
assumes a spherical surface of projection, error will exist in the dewarping of the
distorted input. These errors will then propagate into the stereo range calculation
by means of the disparity measurements and point matching results and will be
evidenced in the scene's reconstruction. The following test procedure will determine
the error existent in the OMNiview dewarping results for a particular camera and
lens. In this experiment and all others in this research using the OMNiview system,
a Toshiba IK-M41A color CCD camera with a 3mm wide-angle lens will be used. The
field-of-view of the lens and camera is $115^\circ \times 88^\circ$. The wide-angle lens possesses non-
symmetric distortions which cannot be completely compensated for by a spherical
lens model. This choice of camera and lens exemplifies the error associated with the use of the OMNIview (spherical projection) dewarping algorithm. The errors are expected to be substantial.

3.1.2 Test Procedure

This test will evaluate OMNIview's ability to correct for distortions in the camera and lens system described above using several different values of the dewarp factor. The goal is to characterize the errors in dewarping and find an optimal correction factor for later experiments. The procedure followed is relatively simple. A calibrated test pattern, shown in figure 3.1, is aligned perpendicular to the camera and located a measured distance away. The warped input is then fed through the OMNIview system and corrected. Once distortions are removed, the output should approach a perspective characteristic of the pinhole model. The centers of each circle are then selected as the points of interest. Then, by means of the pinhole projection transformations, the point's three-dimensional location can be determined. By comparing the planar coordinates of the measured points to the locations of the corresponding projections an evaluation as to the accuracy of the system can be made. The following section presents some of the results of this test.

3.1.3 Accuracy

As mentioned, this accuracy evaluation was performed at various dewarp factors. Test results for each dewarp factor setting were plotted against the original planar coordinate measurements. The best results are shown below in figure 3.2. These
Figure 3.1: Image of calibration board. The test pattern is used for evaluating the accuracy of the OMNiview dewarping algorithm. Correction of this distorted view will be used to approximate a pinhole modeled image. Inverse projection of featured image points are then compared to the actual coordinate values.
Figure 3.2: Correction accuracy of dewarping algorithm. Comparison of the actual locations of the calibration points to their respective dewarped projections. The top image (a) shows the best results obtained in the y direction, whereas (b) demonstrates the most accurate results along the x axis.
plots demonstrate that for this particular wide-angle lens the lens distortions vary between axes. The best selection of a dewarp factor, therefore, is different for the $x$ and $y$ directions of the image. That is, the top image provides the most accurate dewarping for horizontally oriented linear features. However, vertical features are over-corrected by the use of this particular factor. On the other hand, the right image provides the best results in the $x$ direction so that vertical features are most corrected. Notice that the dewarp factor $R$ is different for each set of results. This discrepancy occurs due to the use of a wide-angle lens which does not ensure a radially symmetric distortion. For future stereo tests using this camera and lens, a best choice of the dewarping parameter will have to be made according to some objective criteria.

Figure 3.2 provides the best correction results in the $x$ and $y$ direction, respectively. However, only a single dewarp parameter value can be set for the entire image at a given time. Choosing either factor value independently, therefore, results in a poor correction of the distortion in one of the axis directions. Inaccurate vertical correction, for instance, causes an erroneous epipolar relationship between stereo images, and thus, complicates the matching of corresponding features. Correction errors in the horizontal direction, however, create false disparity measurements. Therefore, minimizing the combined average error in both axial directions leads to the best choice of a dewarp factor for the particular camera and lens. The following graph, figure 3.3, charts the progression of the average error in the dewarping results in each direction using various dewarp factor values. The combined average error deviation in both axes directions is minimal for a dewarp factor value of 470. This dewarp factor value will be chosen for stereo tests involving the Toshiba camera and
Figure 3.3: Plot of dewarp factor selection vs. average correction error. The progression of the average error in both the \( x \) and \( y \) directions when warping the distorted wide-angle image using various values for the dewarp factor. The minimal error for combined image axes is at a lens radius setting of 470.

wide-angle lens system.

### 3.2 A Stereo Test Analysis

Here will be described a simple stereo vision system which utilizes the dewarping capability of the OMNiview system and the spherical lens projections. The goal of this test is to determine the maximum stereo depth measurement accuracy for the camera and lens system described previously. Measures are taken to reduce sources of error and simplify the physical calibration of the system. The stereo pairs of images are acquired by a single camera mounted to a linear translation stage. Horizontal movement of the camera is performed to create left and right stereo images (shown in figure 3.4) and preserve a horizontal epipolar geometry. By using a single camera,
Figure 3.4: Corrected stereo image pair. Above are the dewarped stereo pair of images used in this experiment. The highly randomized pattern is useful in reducing the likelihood of matching errors in the correspondence algorithm.
the difficult physical calibration procedures needed to accurately align a dual camera head stereo system are minimized. The target for the stereo evaluation test is a highly randomized pattern mounted to form a plane perpendicular to the orientation of the camera. Therefore, stereo reconstruction should again model closely a planar surface. Deviation from a planar geometry will quantify the system error.

Stereo images of the pattern are obtained using the OMNIView system to correct for the wide-angle lens distortions. The dewarp factor value found previously, controls the image restoration. Various points representative of the entire field-of-view will be selected from one image, the corresponding point in the image pair found by means of a simple correlation [15] method, and the three-dimensional world coordinates of the projection calculated, as described in Chapter 2. The correlation equation is shown below.

\[
C(a, b) = \frac{a_1b_1 + \ldots + a_mb_m}{\left\{ (a_1^2 + \ldots + a_m^2)(b_1^2 + \ldots + b_m^2) \right\}^{1/2}} \tag{3.1}
\]

where \((a_n, b_n)\) represent the gray scale value of the respective left/right correlation mask. The randomized pattern ensures a high degree of accuracy within the point matching algorithm and thus minimizes this facet of the stereo process as a source of error. However, correspondence cannot be eliminated completely from consideration as an error factor. This is due to the inaccuracies in the dewarped image obtained from OMNIView which will necessitate an increase in the correspondence search area. From the previous section, the dewarp factor of 470 was chosen for these stereo tests. Although selection of this value minimizes the correction error in the "x" direction, thus maximizing the accuracy in disparity calculation, it does not ensure an exact horizontal epipolar relationship between images. As a result, for many of
these tests it was necessary to increase the search domain to multiple lines in order to increase the likelihood of a correct match, especially for outlying features. Once corresponding points have been matched, the pinhole stereo geometry of equation (2.1) is used to define the depth to the world point. The range measurement results for the test should ideally provide a planar formation with all test point projections having the same $z$ value. Deviations from this vertical plane will be the measure of error.

### 3.2.1 Stereo Results

The results of the stereo test using the OMNIview correction model are shown in figure 3.5. This particular surface depicts the reconstructed plane formed from point-wise stereo analysis of the random pattern board. The test was performed at a camera to surface perpendicular distance of 7.9 inches, with a stereo baseline of two inches. The relatively short depth maintained in this experiment is due to the small focal length of the test camera's lens (3mm). The decreased distance ensures proper imaging of the pattern, but is sufficient to demonstrate the errors in range measurement.

The overall field-of-measurement of the stereo system using the two inch baseline is $93^\circ \times 79^\circ$. Throughout this region the average error is approximately 2.9%, with a maximum error of near 8.0% near its limits. Similar error results were attained for varying test depths from 4 to 10 inches. Furthermore, several tests were made using different dewarp factors ranging in values from 440 to 500. Comparable, yet less accurate results were obtained. As a result, this demonstrates that the
Figure 3.5: Stereo reconstruction of test board. This is a display of the reconstructed planar surface formed from the stereo experiment described. The three views are used to demonstrate the curvature of the surface. The gray coded map depicts the change in depth. The curved surface is a direct result of the error exhibited in OMNIview's dewarping of the input image distortions.
system possesses a fairly wide range of values from which the dewarp factor can be selected and comparable accuracies acquired. Errors in range measurement are a result of both inaccuracies in the distortion correction and the loss of resolution at wider angles of view. That is, when correcting significant distortions, OMNIview interpolates multiple data points to represent a single feature point from the input image. Therefore, pixel-accurate point matching is not possible in highly corrected regions of the image.

3.3 Conclusions

This chapter presents an evaluation of the distortion characterization accuracy of the spherical lens model. Accurate description is essential in creating a successful wide-angle stereo process using fisheye optics. As evidenced, errors in the correction of lens distortion result in significant depth estimation error in a stereo vision system.

The first area of investigation is OMNIview's correction of fisheye distortions. Theoretical derivation of the device's dewarping mathematics shows that the correction algorithm is based on projection properties of the spherical lens. However, as demonstrated here actual lenses cannot be accurately characterized using this model. The correction factor $R$ defined in the projection model is not constant when describing true wide-angle lenses. Since the spherical model demands a constant radius, it fails to correct typical distortion accurately throughout the fisheye image. The errors characteristic of this limitation are shown in both the dewarping and stereo evaluations presented in this chapter.
CHAPTER 4

An Enhanced Distortion Characterization Model

The previous chapter investigates a wide-angle, OMNIview-based stereo vision system to demonstrate the insufficient lens distortion characterizations described by the spherical lens model transformations. This chapter presents: (a) an alternative lens model which can more accurately parametrize the nonlinear warping characteristics of a given fisheye/wide-angle lens, (b) a method for determining this lens parameterization, and (c) a comparative error analysis of the distortion correction results.

The development of this enhanced lens characterization strategy remains consistent with the spherical lens properties presented in Chapter 2. For instance, all distortions inherent in the lens are assumed radially symmetric as described by the Azimuth Angle Invariability postulate. Therefore, all projections and corrections again exist along the same radial direction. The difference between the models, however, is in how the lens model describes the formation of radial distortions. The spherical projection equations approximate the radial distortion by the constant radius relationship exhibited in equation (2.11). Therefore, in the mathematical development of the perspective object-plane transformations, the ideal fisheye lens is modeled by a hemisphere with constant radius, \( R \) (figures 2.4, 2.5, 2.6). Equations (2.12) and (2.13) show the mathematical incorporation of this constant dewarp factor.
However, in chapter 3, errors in the correction of wide-angle images proved that this constant radius assumption is not valid for typical fisheye and wide-angle lenses. That is, wide-angle lens distortions cannot be accurately characterized by a normalized projection through a hemispherical surface. A more general surface must be defined.

The goal of this chapter is to redefine the lens surface projection relationship. First, however, a brief revisitation of the spherical lens projection transformations will be performed to give an overview of how the model is used for correcting radial lens distortions. From this model, a more general projection approach can be derived which characterizes distortions that are representative of actual wide-angle lenses. A detailed mathematical development of these general projection transformations will also be given. A simple calibration procedure for defining a surface that properly characterizes the lens will then be provided. Finally, a distortion correction algorithm which implements the surface characterization is presented with results and comparisons.

4.1 Simplification of the Spherical Lens Model

In order to maintain an accurate correction of lens distortions throughout the field-of-view, a physical characterization of the optical system’s perspective transformations is needed. This avoids the high order polynomial mapping problems outlined previously. Zimmermann accomplishes this by defining a normalized projection through a spherical surface to describe the deformation of image features.
This method gives the lens model a physical description. Even though this projection strategy has been shown to be insufficient in characterizing actual lens distortions, the development of the distortion and correction transformations will prove beneficial when deriving the general lens model.

The general mathematics developed by Zimmermann for projecting a point from an arbitrary undistorted world object plane \((u,v)\) through a fisheye lens and onto an image sensor \((x,y)\) are stated in Chapter 2. The OMNIview system he developed, however, is an elaborate device providing for distortionless pan, tilt, rotation, and magnification throughout a hemispherical field-of-view. If the orientation functions are not of concern, equations (2.11) and (2.12) can be greatly simplified. By letting the orientation parameters, \(\delta, \theta,\) and \(\beta\), all equal zero and the magnification factor, \(m = 1\), the fisheye lens can then be modeled as an object plane incident with the lens surface whose central axis is aligned with the optical axis of the fisheye model. This configuration is shown in Figure 4.1. The appropriate equations from Chapter 2 for \(A, B, C,\) and \(D\) reduce to the following:

\[
A = 1 \\
B = 0 \\
C = 0 \\
D = -1
\]

and therefore, the non-linear distortion equations become

\[
x = \frac{R u}{\sqrt{u^2 + v^2 + R^2}}
\]

(4.1)
Figure 4.1: The simplified Spherical Lens Model. The projection of an Image Object Plane through a spherical lens is simplified in the Figure above. Only the distortive projections are considered.
Thus, the ability to correct for the distortions evident in the captured sensor plane image is readily available. To do this, the inverse projection from the sensor plane to the object plane must be performed. Much simpler, equations (4.1) and (4.2) can be solved simultaneously for $u$ and $v$, defining the set of coordinate correction equations.

\[
y = \frac{Rv}{\sqrt{u^2 + v^2 + R^2}} \tag{4.2}
\]

\[
x = \frac{Rx}{\sqrt{R^2 - x^2 - y^2}} \tag{4.3}
\]

\[
v = \frac{Ry}{\sqrt{R^2 - x^2 - y^2}} \tag{4.4}
\]

Dewarping an image is now possible if the proper lens radius parameter $R$ (see figure 2.6) is known since $R$ is the adjustable parameter which controls the amount of distortion correction. With the selection of the appropriate lens radius or correction factor, equations (4.3) and (4.4) can be used to map image points to their undistorted locations. However, a problem occurs when using these equations directly to dewarp an image. Figure 4.2 shows an image corrected using the direct projection from $(x,y)$ to $(u,v)$ space. The gaps represent the progressive stretching and omission of the data when projecting points to the dewarped object plane. The appearance of such artifacts is easily predicted when the projection of a scene onto a camera image sensor is considered. Since the wide-angle lens transformation can be thought of as a compression of information from the scene into a rectangular finite number of sensor elements, the inverse projection of these sensor elements to the dewarped space is a nonlinear stretching of the data. Since the projection is one-to-one as shown in equations (4.3) and (4.4), the same number of elements are used to depict the image in a now larger area; gaps must be evident in the data. A method of avoiding this
Figure 4.2: Forward projection dewarping results. This figure demonstrates the omission of data resulting from the direct mapping of image points to the dewarped perspective. Because of this absence of data in the dewarped image, we will not be able to describe the dewarping using a one-to-one forward mapping.
intermittent omission of data when correcting distorted fisheye images is outlined in section 4.4.

4.1.1 Choosing the Spherical Model Correction Factor

The amount of distortion in a fisheye image is directly related to the lens radius of the spherical model. As a result, correction of the distorted view can be controlled through manipulation of the spherical model radius. For example, by selecting \( R \) large, one establishes little distortion in the forward projection through the surface. Imagine projecting a small planar surface onto the side of a much larger sphere; relatively little distortion will be evident. Conversely, a small lens radius will result in large distortions if a similar projection onto the smaller sphere is performed. As a result, to properly dewarp an image, the modeled spherical surface radius must be selected using the parameter which most closely approximates that of the actual lens. The following summary details a calibration procedure for determining the best lens radius parameter value for dewarping the distortions for a given camera and wide-angle lens system.

The objective of this calibration procedure is to choose a lens radius value, in pixels, which best corrects or linearizes the curved image appearance of an otherwise linear feature. To accomplish this linearization, a series of image coordinates known to be on a straight real-world object are located and iteratively corrected until a straight line is obtained. The value of \( R \) which produces this linearization of the curve represents the dewarp factor or lens radius for the system model. To perform this process accurately, however, a few issues must be addressed. First, the linear
object being imaged must be straight. Second, the choice of image points representing the linear object must possess a sufficiently small deviation from the actual curve. And finally, an accurate method of regression must be performed to ensure a linear representation of the corrected curve.

The first task is to choose a straight object. For tests in this research, the edge of an optical bench bread board is imaged with the edge of interest positioned between the half radius point and the border of the image. This placement insures that a significant amount of distortion is introduced by the lens system. Interior features demonstrate little distortion, and therefore, proper corrective characterization is difficult to quantify. The linear feature should also encompass a significant portion of the field-of-view to ensure a significant number of image points will lie along the distorted linear feature. For this procedure, the edge is oriented near vertical. The exact orientation is not crucial since all distortions are assumed radially symmetric by this model. It is also desired that the calibration procedure be able to automatically extract the points or distinguish the distorted line from the image. In this method, large contrast differences between the optical bench and the background are exploited to accurately describe the edge contour. An image is easily thresholded to obtain the binary representation shown in figure 4.3. The edge representing the binary transition can then be detected to provide an accurate point-wise representation of the curve as is shown in figure 4.3(b).

Once the edge is located and the coordinates of each edge pixel are stored, the curve of can be dewarped using equations (4.3) and (4.4), with $R$ varying over a user defined range. As $R$ is varied, a simple linear regression tool is used to test the deviation of the
Figure 4.3: Edge pixel detection of linear calibration feature. These images demonstrate how a representation of the warped feature is obtained. First, a binary image of a straight edge is obtained, image (a). The edge is located and the points of transition stored, shown in (b), using a simple vertical edge detector.

corrected set of points from a straight line. The best line in the process will possess the smallest absolute deviation between the representative points and the straight line model. An improved correction factor value can be obtained by repeating this procedure for variously positioned linear representations and averaging the resulting factor values to eliminate random error effects. This procedure, outlined in figure 4.4, is similar to the iterative line-based method described by Prescott [9]. Prescott's premise is that the projection of a straight line from the world space should be a straight line in image space if distortions are due only to nonlinearities in the lens. Using a spherical lens model, it can be seen that all points are projected through a surface of constant radius. Thus, finding the lens radius value which provides the most accurate correction of the linear feature will complete the model since all points share the same correction factor. Figure 4.4(a) represents the original warped representation. Figure 4.4(b) through (d) demonstrate the progression of the
Calibration Procedure:
1. Edge detection and curve representation.
2. Vary $R$ from a predetermined $\text{MAX}$ to $\text{MIN}$.
3. Perform a linear regression operation to fit a line to the resulting dewarped data.
4. Evaluate the best line fit representation. The value of the dewarp factor $R$ is the best factor for the particular camera and lens combination.

Figure 4.4: The calibration procedure for finding the dewarp factor for a particular image radius. The dewarp factor which provides the best correction of the fisheye imaged linear feature is selected as the dewarp factor.
dewarping for various values of the correction factor $R$. Figure 4.5 is a demonstration of the correction results using the lens radius $R$ obtained through calibration for a given lens. Because the lens used in these experiments is of high quality, the correction is quite good throughout the interior of the image. However, the correction fails severely near the limits of the field of view.\footnote{The error is quantified in Section 5 of this chapter.}

### 4.2 Description of a General Wide-angle Lens Model

While the spherical model of the lens system is demonstrated in the previous section to provide qualitatively acceptable dewarping, real lenses cannot be fabricated with enough precision so that the spherical lens model can be used for image metrology applications. For instance, as the radial distance increases in the images, the lens radius parameter $R$ needed to best correct distorted features will generally vary. This variation in the dewarp factor implies that the spherical surface used to depict the ideal fisheye lens is inaccurate for describing actual camera systems. Therefore, the projection transformations need to be generalized to account for deviations from the spherical surface model.

Two assumptions are made initially. First, all distortions are assumed radial as in the spherical model. In other words, the Azimuth Angle Invariability Postulate described in Chapter 2 still applies. This eliminates the need for angular distortion correction. Angular distortions are usually insignificant, and result primarily from poorly mounted lens systems. Second, lens surfaces are assumed to be smoothly
Figure 4.5: Final correction results of test board. Shown here is an image corrected when using a spherical lens model. This ideal model obviously fails to accurately characterize the wider angle of the field-of-view.
Figure 4.6: Diagram of the general lens surface coordinate reference frame for describing the projection of a point in an object plane through a general lens surface and onto the camera sensor. This camera/lens model will be used to develop a radial distortion correction algorithm.

Varying in shape with no folds. These properties ensure that projections along a radial line are unique and can be described by continuous functions.

Figure 4.6 shows the coordinate reference frame used for the general distortion characterization. The object plane represents the undistorted image space and is perpendicular to the optical axis and aligned with the sensor plane coordinate system. Therefore, the center of the object plane can be described from the image plane origin by:

\[
\begin{align*}
  x &= 0 \\
  y &= 0 \\
  z &= R_0
\end{align*}
\]

where \( R_0 \) is the initial height (radius) of the defined surface. Defining the origin of the object plane as a vector relative to the sensor plane \((x,y)\), the following vector is
formed:

\[ \mathbf{O}_{xy} = [0, 0, R_0] \]  \hspace{1cm} (4.6)

The object point of interest, relative to the object plane origin can be represented in terms of image plane coordinates:

\[ x = u \]  \hspace{1cm} (4.7)

\[ y = v \]

\[ z = R_0 \]

thus giving the vector relative to the object plane origin:

\[ \mathbf{P}_{uv} = [u, v, 0] \]  \hspace{1cm} (4.8)

Therefore, relative to the image center the vector expression simply becomes the sum of the two independent vectors.

\[ \mathbf{P}_{xy} = \mathbf{O}_{xy} + \mathbf{P}_{uv} \]  \hspace{1cm} (4.9)

\[ \mathbf{P}_{xy} = [u, v, R_0] \]  \hspace{1cm} (4.10)

Normalized projection onto a surface of radius \( R(x, y) \) is determined by producing a surface vector \( \mathbf{S} \):

\[ \mathbf{S}_{xy} = \frac{R(x, y) \times \mathbf{P}_{xy}}{||\mathbf{P}_{xy}||} \]  \hspace{1cm} (4.11)

Substituting yields the following vector expression for the mapping of an object plane
point onto the surface:

\[ S_{xy} = \frac{R(x, y) \times P_{xy}}{\sqrt{u^2 + v^2 + R_0^2}} \]  

(4.12)

And thus, the projection onto the two-dimensional image plane becomes simply the \( x \) and \( y \) component of the surface vector. The distortion equations become:

\[ x = \frac{R(u, v)u}{\sqrt{u^2 + v^2 + R_0^2}} \]  

(4.13)

\[ y = \frac{R(u, v)v}{\sqrt{u^2 + v^2 + R_0^2}} \]  

(4.14)

The inverse projection can be easily found by solving the above equations for \( u \) and \( v \). The expressions for distortion correction are shown in the following:

\[ u = \frac{R_0 x}{\sqrt{R^2(x, y) - x^2 - y^2}} \]  

(4.15)

\[ v = \frac{R_0 y}{\sqrt{R^2(x, y) - x^2 - y^2}} \]  

(4.16)

From equations (4.13 - 4.16), notice that \( R \) must be parametrized with respect to both coordinate spaces. This parameterization process will be detailed in section 4.3 and 4.4.

Figure 4.7 shows a cross-section of the modeled system with an arbitrary surface inserted to enhance visualization of the projections. Since \((x, y)\) are known for each point, both \( R_0 \) and \( R(x, y) \) must be determined in order to properly dewarp the image. From figure 4.7, however, notice that \( R^2 = r^2 + h^2 \) or in cartesian coordinates \( R^2 = x^2 + y^2 + h^2 \). Substituting this relationship into the correction equations (4.15) and (4.16), the expressions simplify to the following:

\[ u = \frac{R_0 x}{h(x, y)} \]  

(4.17)

\[ v = \frac{R_0 y}{h(x, y)} \]  

(4.18)
Therefore, a description of the lens surface by height at a given sensor plane coordinate provides a simple means of projecting the pixel to its undistorted location.

4.3 Lens Surface Model Characterization

The previous section develops the transformation equations which describe the projection of points in an object plane through an arbitrary surface onto the sensor. However, the surface at this point has not been characterized. This section will develop a method for modeling a particular lens' physical surface profile. Furthermore, a calibration procedure will be discussed which provides a simple, direct method for determining the needed parameter, surface height, as a function of image coordinates.
For this development, the complexity of the descriptor will be limited to a quadric surface. Thus, the general lens surface [16] can be modeled using the following expression:

\[ Ax^2 + By^2 + Cz^2 + Dxy + Ex + Fy + Gz + H = 0 \]  

(4.19)

If the cross term is eliminated, then the second-order lens surface is assumed to be aligned with the specified cartesian coordinate system. Also, the \( z^2 \) term is dropped since the surface need only be defined in the positive direction. As a result of not having any cross terms, the surface equation can be decoupled and described separately as functions of \( x \) and \( y \). Therefore, the simplified equation in terms of the height, \( h \), instead of \( z \), becomes:

\[
\begin{align*}
    h(x, y) &= (Ax^2 + Ex + H) + (By^2 + Fy + H) \\
    h(x, y) &= h(x) + h(y) - H
\end{align*}
\]  

(4.20, 4.21)

From the above equation, it is clear that the lens height at any \((x, y)\) image coordinate can be calculated if \( h(x) \) and \( h(y) \) are known. Thus, characterizations of \( h(x) \) and \( h(y) \) must be recovered for any given lens in order to remove the nonlinear distortions.

The first step in calibrating our axial surface functions is to define the center of distortion. Several methods for locating this coordinate location have already been determined in previous research [17, 11, 10, 18], and the process is not reinvented here. A procedure similar to the method described by Basu [17] is incorporated in this calibration. Once, the center of distortion is located, the camera system is setup orthogonal to a calibration board with the center of the image frame buffer aligned with a row and column of the calibration points. The setup used for calibrating
the system is shown in figure 4.8. For this process, the concern is only with the points along the \( x \) and \( y \) axes as are marked in the figure. The most important aspect of the process is that the calibration points possess a known separation. This separation distance can then be used to calculate the desired pixel disparity between the undistorted locations of the desired pinhole imaged points. To calculate this separation, the existence of negligible distortion in the center of the image is used to obtain the unit length per pixel relationship between the image and board. For the test conducted in this research the size of the center calibration circle is utilized to determine this quantity. Once the length/pixel ratio is found, the desired pixel
distance between undistorted image calibration points is easily calculated.

After manually extracting the $x$ or $y$ coordinates for each point along the $x$-axis (y-axis), the distortion equations (4.13) and (4.14) are used to solve for the needed lens radius component $R$ which moves each uncorrected image point to its calculated position in the dewarped space. Considering the calibration only along the $x, u$ axis directions where ($y = v = 0$), the following expression is formed:

$$R(u) = \frac{x}{u} \cdot \sqrt{R_0^2 + u^2},$$

(4.22)

and likewise, in the $y$ and $v$ direction:

$$R(v) = \frac{y}{v} \cdot \sqrt{R_0^2 + v^2}$$

(4.23)

Since $(x, y)$ are found from the image and $(u, v)$ can be calculated from the undistorted pixel to unit length relationship, these two expressions can be used to determine $R(u)$ and $R(v)$ if the image center lens radius $R_0$ is known. However, since $R_0$ must be the same in both the $x$ and $y$ directions, the functions can be determined as follows. First, the radius value for each calibration point along the respective axis is found by employing equations (4.22) and (4.23) while using an arbitrary value for $R_0$. Functions for both $R(u)$ and $R(v)$ are then determined from this data set. If the constant terms in each fit are both not equal to $R_0$, the process is repeated with different choices of $R_0$ until the equality constraint is met. When this equality constraint is achieved, the functions $R(u)$ and $R(v)$ represent the change in lens radius along the undistorted coordinate axes. The results of this process for a given lens are shown in figure 4.9. As a side note, for all tests conducted to date, the best functional fit to these data points has proven to be linear. Future testing may find cases where this
Figure 4.9: Plots of the axial change in lens radius. A linear fit is used to characterize the change in the modeled lens radius as a function of the axial components of the dewarped space. The plot clearly shows a reduction in the model's radius off the optical axis.

A linear relationship is not applicable. However, higher order fits will not affect the calibration process.

The final stage of the calibration process is to characterize the relationship between the x and y axial components and the height of the surface. From figure 4.7, it is apparent that along the x-axis the tangent of the angle $\beta$ is equal to $\frac{h_x}{R_u}$. With this relationship, the $x$ and $h$ values corresponding to the desired $u$ are easily described:

$$x = \frac{R_u}{\sin \beta} \quad \text{and} \quad h_x = \frac{R_u}{\cos \beta}$$

The relationships are similar along the y axis. Plotting $h_x$ vs. $x$, and likewise $h_y$ vs. $y$, the corresponding axial lens cross sections are determined. The resulting second order curves are shown in figure 4.10. Furthermore, because the surface fits are generally smooth, the coefficients of all the odd-ordered terms are zero. Substituting equations (4.24) and (4.25) into equation (4.21), the final expression of the quadric
Figure 4.10: Plots of the reconstructed lens surface along respective axes. The plots above demonstrate the two axial cross sections of the lens surface. In the plots, the surface height is found as a function of $x$ and $y$, respectively, through the surface relationships previously developed.

The surface now becomes an elliptic paraboloid of the form:

$$h = Ax^2 + By^2 + R_0$$  \hspace{1cm} (4.26)

where $a_2$ and $b_2$ are found numerically. Note that this equation provides a complete characterization of the lens height above the sensor plane as a function of actual image coordinates.

An Inventor model depiction of the corresponding lens surface model is shown in figure 4.11(a). This lens model describes a surface characterization of the Nikkor, 16mm F2.8 fisheye lens mounted on the Kodak DCS460 digital camera.

### 4.4 Dewarping Implementation

With the formation of the wide-angle lens projection model and characteristic lens surface, a method for correcting lens distortions is readily available. Utilizing
Figure 4.11: Inventor reconstructions of the quadric and spherical lens models. The resulting Inventor models portraying the lens surface as characterized during the system calibration routine is shown in (a). Notice the deviation from the ideal spherical model (b).
the correction equations depicted by equations (4.13) and (4.14), a direct mapping of the distorted image pixels to an undistorted space can be performed. However, as exhibited in figure 4.12, gaps in the scene appear due to the omission of data during forward projection, as was discussed previously. To eliminate these gaps, a method of backward projection must be developed for the new quadric surface model. The distortion equations (4.15) and (4.16) provide the inverse projection through the lens surface, and thus a method can be implemented to remove the unwanted gaps in the corrected image. However, this correction scheme begins with only knowledge of the undistorted coordinates, and no description exists of the lens surface in terms of the dewarped spatial components \((u, v)\). That is, in the calibration process the lens surface is described in terms of the sensor plane or distorted coordinates \((x, y)\). As a result, the lens radius parameter \(R\) of the projection equations is not defined as a function of the corrected coordinates \((u, v)\). Various solutions to this problem are proposed.

The first two proposed techniques involve redefining the lens surface in terms of \(u\) and \(v\). To do this, a dense selection of points are mapped to the dewarped space according to the forward correction transformations. The known lens height found during forward projection can then be plotted versus the undistorted coordinate locations. An Inventor model depicting such a surface is shown in figure 4.13. Two different options for backward projection now exist: (1) create a dense Look Up Table (LUT) that relates the dewarped coordinates to the proper lens model height, or (2) fit a three-dimensional surface function to the dense data set. The use of a LUT is problematic due to the potentially enormous size of the table. The surface
Figure 4.12: Forward dewarping using the quadric surface lens model. Distortion correction resulting from the forward projection of image coordinates to the dewarped space. A back projection scheme will be implemented to avoid the omission of data.
fitting technique is also difficult, since high-order surface fitting is not always highly accurate near extremities.

To avoid the potential pitfalls of the two procedures, a third method has been devised that takes advantage of the second order function used to describe the surface in the uncorrected space. Based in vector calculus, the algorithm uses the physical description of the model already created in the following manner.\(^2\) With reference to equation (4.9) the undistorted coordinates are defined as a position vector \(\mathbf{R}\) in terms of the image space by:

\[
\mathbf{R} = \mathbf{P}_{xy} = [u, v, R_0]
\]  

This position vector is then scaled by the parameterization factor \(t\), thus defining a

---

\(^2\)Refer to figure 4.6 to aid in visualization of the procedure.
new surface vector $\mathbf{R}$ with magnitude equal to $R$, the local radius of the surface:

$$ \mathbf{P}_{xy} \times t = [u_t, v_t, R_0 t] $$ (4.28)

Since the height of the surface is already defined by equation (4.26), the following linear system can be written:

$$ u_t = x $$ (4.29)
$$ v_t = y $$ (4.30)

$$ R_0 t = Ax^2 + By^2 + R_0 $$ (4.31)

From the system of equations, the parameter $t$ is then found by solving for the roots of the following polynomial:

$$ (Au^2 + Bv^2)t^2 - R_0 t + R_0 = 0 $$ (4.32)

Once the proper root is found ($0 < t < 1$), equations 4.29 and 4.30 are then used to define the direct coordinate mapping, and the back projection dewarping algorithm is completed. The result of this dewarping procedure is shown in figure 4.14.

### 4.5 Statistical Error Analysis

In this section, a comparative error analysis is presented between the results of dewarping using the spherical lens model and the quadric surface model algorithms. For the purpose of comparison, an image of the point calibration board is dewarped using the spherical lens model (figure 4.5) and the quadric lens model (figure 4.14).
Figure 4.14: Final dewarping results using back-projection method. An uninterrupted and corrected perspective produced by back projecting the undistorted coordinates to their corresponding sensor plane location.

Qualitatively, the quadric surface lens characterization provides a better correction of the image throughout the field-of-view. In the image corrected using the spherical lens model, the error becomes progressively worse as the distance from the lens center increases. Notice the substantial improvement in the edge features in figure 4.14.

In the two graphs of figure 4.15, the corrected calibration points are plotted against their ground truth locations where the known locations are found as detailed in the calibration procedure development, presented in section 4.3. The error measures for each are shown in table 4.1. As can be seen in this chart, the quadric lens
Figure 4.15: Error plots of correction results. The plots above depict the dewarping results for both the spherical lens characterization (a) and the quadric surface lens description (b). The center of the black circles depict the known location of the undistorted coordinates. Significant improvement results from use of the more general lens surface description in (b).
Table 4.1: The above table quantifies the errors resulting from the correction of radial lens distortions, using both an ideal spherical model and a 2nd order quadric surface description. Substantial improvement is evidenced in the more general surface characterization.

<table>
<thead>
<tr>
<th>Surface Model</th>
<th>Error (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg (pix)</td>
</tr>
<tr>
<td>Spherical</td>
<td>18.5</td>
</tr>
<tr>
<td>Quadric</td>
<td>4.1</td>
</tr>
</tbody>
</table>

lens characterization not only qualitatively out performs the spherical model, but provides much better quantitative results as well.

### 4.6 Conclusions

In this chapter, an enhanced algorithm for correcting nonlinear lens distortions is presented. The algorithm is based on a physical description of the lens surface and avoids the high order point to point mapping routines discussed in previous literature. The ideal fisheye lens model is described by a normalized projection of a point in an object plane onto the surface of a sphere. In this distortion characterization scheme, the limitations of the spherical model are relaxed by allowing the surface
description of the lens to be any smooth, continuous function. The use of quadric surfaces to characterize the lens is thus developed to demonstrate the robust nature of this transformation. In fact, very good results are shown for the corrected test image. The advantage of this radial lens distortion model is the simple bi-directional mapping capability inherent in the algorithm's development. By giving the model a meaningful physical description, the process of mapping between distorted and undistorted spaces is simplified by using geometric vector relationships.

Another advantage to the bi-directional mapping capability of this surface modeling correction scheme is not readily evident when considering the model for use solely as a dewarping agent. The true advantage of this characterization process is evidenced with its incorporation into a wide-angle stereo vision system, detailed in the following chapter.
CHAPTER 5

OMNister: an Omnidirectional Stereo Vision System

By far the greatest challenge to any stereo vision system is correspondence and the matching of points and features from the respective pair of images. As a result, the incorporation of wide-angle optics would seem to only complicate an already immensely difficult task. The significant spatial distortions characteristic of wide-angle cameras eliminate the linear epipolar search constraint found when using traditional rectilinear camera systems for stereo. When wide-angle imaging is utilized, the epipolar relationships between linear baseline image pairs no longer result in linear image feature translations. Motion of a pixel between images is now characterized by a distinct nonlinear curve. Distorted motion, however, is not the only significant complication cause by using wide-angle image pairs. Since shape is also distorted by wide-angle lenses, image features do not maintain a consistent shape between disparate locations. That is, the shape of an imaged object will appear vastly different in disparate locations on the sensor (see figure 1.3). As a result, matching these warped features between image pairs is another complication and challenge for such stereo systems.

The typical solution to these challenging problems in wide-angle stereo has been to systematically eliminate the distortions in the image pair and to create two undistorted stereo images. Thus, the linear epipolar relationships between images are
recovered and imaged feature shapes in respective stereo images are again similar. In fact, many researchers have maintained that correction of the wide-angle angle lens distortions is essential to achieving accurate stereo correspondence.

In this chapter, however, a novel omnidirectional stereo vision algorithm and system, termed OMNIster, will be developed that employs a search strategy based on the curved epipolar relationships which exist between high-distortion stereo pairs. By performing the search in the warped space, the need to actually dewarp the entire stereo image pair in order to find matching feature points is eliminated. These techniques utilize the wide-angle lens surface characterization model described in the previous chapter to define a curved epipolar search path between images. Such a search technique is viable due to the physical surface model projection scheme from which bi-directional perspective transformations were defined. This chapter will first detail this distorted correspondence process development. A stereo test setup with depth estimation results and error analysis will then be provided.

5.1 Distorted Epipolar Correlation Strategy

As mentioned previously, the distortion evidenced in wide-angle images complicates the search strategy generally used for stereo correspondence. In customary stereo applications, a horizontal relationship between camera sensor locations is defined in order to reduce the search area between images to a single row of pixel elements. However, when wide-angle or fisheye lenses are used, the epipolar relationship between the images is no longer horizontal. The epipolar line is now
transformed to a curve, defined by the projection characteristics of the lens. As a result, defining an efficient search strategy between stereo images is significantly more complicated. Therefore, previous research into wide-angle stereo has eliminated this need for a distorted search path by fully correcting the high-distortion images and subsequently applying traditional stereo correspondence methods to these spatially corrected image pairs.

However, the inefficiencies of this implementation can be significant. First, both images must be entirely corrected at a notable computational time cost. Second, the two corrected images are now much larger than the respective original distorted images. The corrected image size, for instance, can be as much as three to four times larger than the original distorted image. For ordinary resolution images, such a cost to memory does not severely impact the performance of the processing machine. However, when high resolution images are being used for stereo processing, memory management problems can become computationally overwhelming. For example, the OMNister system utilizes a Kodak DCS460c camera which possesses the highest resolution of any digital camera on the market to date – 3060x2036 pixel elements. Using 8-bit grayscale, the memory storage requirements of the original distorted stereo images is nearly 12.5 MB. The undistorted full resolution images, on the other hand, require an enormous 42 megabyte memory capacity, a severe test for most computing systems. The processing of full 24-bit color images, furthermore, is simply unthinkable requiring 126 MB of memory. As a result, the potential for substantial memory savings exists as an inspiration for the processing of the distorted images for stereo.
However, this is not the only reason for performing distorted stereo. Accuracy issues also arise when correlating features between reconstructed images. Uninterrupted correction of a digital image involves a many-to-one mapping strategy. (This mapping procedure is detailed in Chapter 4.) As a result, several pixels in a corrected image can represent a single pixel in the original image. This can adversely effect many matching routines, especially correlation-based methods due to the potential comparison of multiply defined pixel features. The existence of such multiplicity is a direct result of the loss of resolution in the fisheye image towards the image extremes. When correcting this distorted image perspective, the dewarped image must be larger in order to contain an entirely corrected perspective. This requires that in cases where distortions are significant, more than one point in the dewarped space must represent a single point in the original image. In fact, the number of points in the object plane \((u, v)\) representing a single point in \((x, y)\) will generally increase with the radial distance from the image center as is demonstrated in figure 5.1. This figure shows the number of undistorted pixel locations that are mapped to each individual location in the distorted image. Figure 5.1 demonstrates that upto eight pixels are mapped to a single pixel during correction. Therefore, in a stereo vision application, the use of a dewarped fisheye image may result in erroneous point matching results when using the traditional correlation based matching strategy. For instance, consider a matching scenario in which a point of interest occupies a central location in one image and an outlying location in the stereo pair. Once the stereo image pair is corrected, the two corresponding points can be potentially very dissimilar in graylevels. Referring to figure 5.1, the featured point that is centrally located will
be represented by one pixel in its dewarped image. However, correction of points in
the outlying region may result in a representation of the feature by as many as eight
pixels or more. A difficulty now exists in accurately matching the corresponding
points and measuring disparity. The correspondence strategy developed hereafter
will include design features which will minimize the amount of over-correlation in
the point matching process.

As described previously, the matching routine developed is a point-wise, graylevel
correlation-based technique. Correlation is used due to its general applicability to
the stereo correspondence problem. However, the following novel matching strategy
The formation of the curved epipolar search path in the left image is defined in this three step process. The undistorted epipolar line is established by the corrective projection of the point of interest to the dewarped domain. The transformation of the coordinates along this row to their corresponding distorted locations forms the curved epipolar search path in the left image.

is not dependent upon using correlation as the matching criteria. The new technique simply attempts to redefine the epipolar relationship between a stereo pair of images according to the quadric transformation algorithm developed in the previous chapter.

Figure 5.2 depicts the three-stage bi-directional mapping technique that is used to define the distorted epipolar search path. Characterization of the curved epipolar relationship is accomplished by a projection to and iterative transformation from the dewarped image space, or image object plane as defined by the system model. A few assumptions concerning the stereo setup will be made before continuing. First, it is assumed that the stereo images are obtained using a single wide-angle lens camera system mounted to a one-axis translation element. The use of a single camera system eliminates a need for two or more camera models; however, this does not affect the
actual process development. Second, the translation of the system is assumed to be axial, and in this case horizontal, creating left and right stereo image pairs. And finally, the search direction is assumed to be from left to right, demanding that the initial point of interest be in the right image. The corresponding point is then found in the left image.

Once selection of the point of interest from the right image \((x_r, y_r)\) is made, the coordinate location of the pixel is transformed to the dewarped object plane \((u_r, v_r)\) using the correction equations, (4.17) and (4.18). In this space, the epipolar relationship between images is of course linear and horizontal, so \(v_l = v_r\). Also, it is evident that \(u_l = u_r + n\). Therefore, this undistorted coordinate location relationship in the dewarped space can define our search path in the distorted left image. By letting \(n\) vary between predefined limits and back projecting the coordinate locations of the undistorted epipolar line to distorted image coordinates in the left image using equations (4.13) and (4.14), the curved epipolar path is defined. This simple, yet useful, epipolar relationship provides an accurate description of the distorted search region.

Although the epipolar search path is easily recovered, disparate image features potentially possess very different distortions in the two images. As a result, straight mask correlation on the distorted images will produce inconsistent and poor matching results. For instance, figure 5.3 demonstrates the significant dissimilar appearances that can exist in corresponding regions of the distorted image pair. Correlation of these two regions would produce poor matching results and an exact match can definitely not be guaranteed. Therefore, the question is how to define the correlation window and its relationship between image spaces to account for local distortion.
The technique described is based in concept on an adaptive windowing correlation method described by Kanade [19]. The correlation strategy employed by Kanade employs a rectangular mask that is adjusted in size and dimensions to define optimal correlation performance. However, for application in distorted stereo, not only must the window’s size and dimensions be adjusted, but each window’s general shape must be distorted as well. This OMNIster system employs a warping window correlation strategy which distorts the shape of the correlation mask according to the local image distortions. The most straightforward method for implementing the warping window strategy is to define the mask region in the dewarped space of the right image and obtain the right correlation mask values by back projecting each pixel position of the mask to its distorted location in the right image. The left correlation mask is obtained similarly by iteratively moving the window along the epipolar path in the
Figure 5.4: *Diagram of a method of distorted correlation.* In this matching process, the correlation window is chosen around the corrected coordinate location of the point of interest. Back projection of the mask coordinates to the respective image forms the left and right correlation arrays.

As a result, the shape of the moving window adapts to changes in the distortion as it is pushed along the search path in the left image. Depicted in figure 5.4, each pixel in the mask is translated in the undistorted space and projected to the corresponding location in the left image. From this new distorted window, the adapted correlation results are calculated. However, from the previous discussion concerning the many-to-one mapping that is characteristic of the corrected to distorted plane projection, such a process still produces inaccurate point matches due to multiply defined point mappings. In essence, this correlation process is the same as performing the correlation on the corrected images. The only savings result from the reduced memory load this algorithm places on the machine.

As a result of the problems discussed in the previous paragraph, the following
Figure 5.5: Diagram of the warping window stereo correlation strategy. This correlation process is similar to the previous. However, selection of the initial mask is performed in the distorted image. This selection ensures the unique correlation window and minimizes the repeated correlation of mask points.

modified warping window correlation process, depicted in figure 5.5, has been developed to minimize the recorrelation of pixel locations in the distorted image. In many accounts, the process diagrammed in figure 5.5 is similar to the matching routine shown in figure 5.4. However, the formation of the original window is performed differently. In this algorithm, the right mask is originally defined in the distorted right image plane. To find the corresponding left correlation window, the mask’s pixel locations are mapped to the dewarped space, translated along the epipolar line, and then projected to a corresponding region along the search path in the left image. The advantage of this second warping window correlation routine is that the initialization of the mask in the warped image domain ensures a totally unique correlation window. All values in the window represent a unique location in the image. By then
adapting the shape of the window to conform to the distortion changes as the mask progresses along the curved search path in the left image, the errors introduced by the many-to-one nonlinear mapping are minimized.

5.2 Stereo Test Procedure and Results

Once corresponding pixel locations in the left and right stereo image are found, a simple triangulation projection strategy is used to calculate the depth to the point of interest. Currently, this triangulation strategy is based upon the linearity properties of the pinhole camera model. Since the undistorted locations of the matching image points can be recovered during the matching process, a pinhole projection model is easily implemented to find the three-dimensional location of the featured point. The lens surface characterization model developed in Chapter 4 could also be employed to develop stereo triangulation equations. However, initial tests show that no advantage is gained by using such a stereo projection model.

Evaluation of the stereo accuracy will proceed in the same manner as the stereo tests of the OMNIview system in Chapter 3. A stereo pair of images are obtained of the random pattern board using the wide-angle camera system. The camera and lens system used in this test is the Kodak DCS460c camera with a Nikkor f2.8, 16mm fisheye lens. Two stereo reconstructions of the test board, one using the spherical correction model and one using the quadric surface model, are shown in figures 5.6 and 5.7, respectively.

In figure 5.8, the residual error from the planar surface fitting is shown for both
Figure 5.6: Stereo reconstruction of the planar surface using spherical lens model is demonstrated in the Inventor model above. Significant curvature is exhibited in the surface. Furthermore, corner data is completely unrecoverable due to correlation failure.

Figure 5.7: Stereo reconstruction of the planar random pattern board using the quadric surface lens characterization is demonstrated in the point-wise Inventor model above. Error in the surface is greatly reduced and accurate depth information is obtained from the entire field-of-view.
Figure 5.8: Plot of the x-axis depth measurement errors. Shown above are error plots for the stereo range measurements obtained during depth estimation tests for both the spherical lens model (a) and the quadric surface characterization (b). Significant improvement is demonstrated in the quadric surface stereo results.
the spherically and quadrically based reconstructions. Notice the strong correlation between errors in the distortion correction results evidenced in figure 4.5 and the stereo reconstructions recovered using the spherical lens model. In the interior region of the plane, the reconstruction of the surface shows only minor errors displaying gradual curvature in surface errors. However, this curvature increases significantly towards the edge of the field-of-view until finally in the extreme regions, the correlation failed due to a poor characterization of the true search path. These errors in reconstruction are expected considering the dewarping results shown in Chapter 4 where the spherical lens model was used. When a constant radius is assumed, the least accurate corrections are exhibited in the corners of the field of view.

On the other hand, a significantly more accurate planar reconstruction is obtained using the quadric surface model. Notice that even in the farthest extremes of the diagonal, accurate depth measurements are obtained due to the more accurate lens characterization. Although some error is found in the interior of the plane, the average residual is significantly reduced. From the statistical error results shown in table 5.1, the maximum error in the spherical model stereo reconstruction is 9.48% where the maximum deviation from the planar fit is only 3.184% when the quadric surface characterization is employed. These results demonstrate a considerable increase in 3-D reconstruction accuracy and reliability when the quadric surface system model is used to guide the stereo matching process.
Table 5.1: The above table quantifies the errors resulting from stereo depth measurement during the tests described previously, using both an ideal spherical lens model and a 2nd order quadric surface description. The significant reduction in the maximum absolute error exemplifies the improved distortion characterization exhibited in the more general surface description of the lens.

<table>
<thead>
<tr>
<th>Stereo Depth Measurement Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera: Kodak DCS460c</td>
</tr>
<tr>
<td>Resolution: (3060 x 2036)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surface Model</th>
<th>Absolute Error</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg (mm)</td>
<td>Max (mm)</td>
<td>Min (mm)</td>
<td>Avg (%)</td>
</tr>
<tr>
<td>Spherical</td>
<td>3.635*</td>
<td>16.38</td>
<td>0.00</td>
<td>2.10</td>
</tr>
<tr>
<td>Quadric</td>
<td>2.340*</td>
<td>5.545</td>
<td>0.00</td>
<td>1.35</td>
</tr>
</tbody>
</table>
5.3 Conclusions

In this chapter, an efficient and accurate stereo vision system, termed OMNister is developed and tested. A novel strategy for describing the epipolar geometry between a pair of high-distortion stereo images is also presented. The search strategy described eliminates the need to systematically correct the distorted wide-angle image pair. As a result, this correspondence strategy alleviates a significant load on the memory needs of the computing system; overall, this reduction in memory requirements is on the order of 250% or more. Furthermore, the warping window correlation technique also ensures a higher degree of accuracy in point matching when corresponding regions between images possess significantly different levels of nonlinear spatial distortion. A large decrease in processing time is also achieved resulting from the elimination of the prior need to perform the distortion correction. This method of characterizing a search path between images is successful due to the bi-directional transformations described in the lens surface model that is detailed in the previous chapter. As expected, the results of this stereo vision system prove both more reliable and more accurate than the results obtained when using the ideal spherical lens model to describe the camera system. The small errors which are evident result from slight inaccuracies in the surface characterization of the lens.
In this research, a wide-angle stereo vision system has been successfully developed and implemented. The system, termed OMNister, is inspired in concept by the field-of-view limitations imposed by traditional parallel axis stereo systems. The goal of the OMNister system is therefore, to provide accurate passive depth measurements throughout a large field-of-view allowed by incorporating fisheye and wide-angle optics into the image acquisition system. The use of wide-angle optics, although beneficial in providing maximum visual information from the scene, results in a severe complications due to the “barrel-warped” image distortions which result due to the projection characteristics of the lens. An efficient and accurate method of characterizing the radial distortions of the wide-angle lens is therefore desired and needed before implementation of the fisheye stereo system can be performed.

The desire for a quick and efficient correction of wide-angle images lead to the investigation and evaluation of the OMNIview camera orientation system. This video monitoring system provides the capability to digitally correct for spherical distortions and orientation control throughout a field-of-view in real-time. To better define the algorithm employed by the system for image restoration, Chapter 2 first revisited the mathematical development of the spherical lens model. The following chapter then evaluates the systems accuracy and functionality in a simple stereo vision system.
The results of both the distortion correction evaluation and stereo implementation demonstrated that the system and lens model are not well suited for accurate vision-based metrology. Errors in the distortion correction algorithm caused by the spherical lens model directly and significantly impact the accuracy of the stereo system.

As a result, a new method of lens distortion characterization is needed to complete the development of the OMNIster system. The solution proved to be a generalization of the spherical projection model used by the OMNIview system. In this model, projection through the lens is described as a normalized projection onto a spherical surface. To reduce the limitations of the spherical projection algorithm, a more general surface characterization is described which is specifically defined for a particular camera and lens. The development of the general perspective transformation equations is completed in Chapter 4 and a calibration process is suggested for characterizing the general lens surface description. A distortion correction evaluation with accuracy analysis is also provided to demonstrate the substantial improvement in system image dewarping accuracy.

A novel stereo implementation of the generalized distortion characterization algorithm is described in Chapter 5. This system avoids the customary methods of wide-angle stereo which require correction of the stereo image pair prior to stereo analysis. A distorted correlation search strategy is defined utilizing the bi-direction projections described by the surface characterization distortion model. The correspondence algorithm redefines the epipolar relationships between image pairs to develop an accurate search path. A warping window implementation is also discussed which optimizes the correlation process. Final stereo test results provided
which show a significant increase in point matching reliability and stereo depth measurement accuracy.

Future work in this research would center primarily around a more accurate description of the surface characterization of the lens model. Initial investigation developed only quadric surface characterizations, specifically the elliptic paraboloid. A more detailed modeling strategy, incorporating improved surface descriptors, could further improve the accuracy, reliability, and efficiency of the OMNister wide-angle stereo vision system.
BIBLIOGRAPHY


"A simple calibration procedure for fish-eye (high distortion) lens camera,"
in *Proceedings IEEE INternational Conference on Robotics and Automation*,

[12] Steven D. Zimmermann,
"Design and implementation of an electronic pan/tilt/rotate/magnify camera sys­
tem for wide-angle viewing applications,"

[13] W. R. Walsh, P. Hansen, H. L. Martin, and M. Yamaji,
"Stereoscopic triangulation control of a robot using wide-angle imaging,"
in *Proceedings of the SPIE*. The International Society for Optical Engineering,

[14] R.C. Gonzalez and R. E. Woods,
*Digital Image Processing*,
Addison-Wesley, Reading, Massachusetts, 1992.

[15] M. A. Abidi,
*Optimization-Driven Machine-Based Motion Measurement*,

*Calculus and Analytic Geometry*,

[17] S. Basu, A.; Licardie,
"Alternative models for fish-eye lenses,"
*Pattern Recognition Letters*, vol. 16, no. 4, pp. 433–441, April 1995.

[18] S.A. Willson, R.G.; Shafer,
"What is the center of the image?,”

[19] T. Kanade and M. Okutomi,
"A stereo matching algorithm with an adaptive window; theory an experiment,”
in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, Ed.

[20] J. Weng, P. Cohen, and M. Herniou,
"Calibration of stereo cameras using a non-linear distortion model,”
in *10th International Conference on Pattern Recognition*. IEEE, June 1990,
vol. 1, pp. 246–53.
VITA

Keith B. Johnson was born in Knoxville, Tennessee in June of 1972. He lived in East Tennessee until 1983, moving with his parents to Atlanta, Georgia where he attended Crestwood High School. Graduating in 1990, he returned to Knoxville and began study toward his BSEE at the University of Tennessee. During his undergraduate career, Mr. Johnson also entered the cooperative engineering program during which he participated in four assignments with Eastman Chemical Company specializing in Controls Engineering. However, an opportunity to serve as a research assistant for Professor Mongi A. Abidi in the Imaging, Robotics and Intelligent Systems Laboratory during the final semester of his senior year gave him a chance to pursue his interest in the field of robotics and vision. Upon graduating with his BS in Electrical Engineering, he immediately continued his education enrolling in the MSEE program also at the University of Tennessee-Knoxville under the direction of Dr. M. A. Abidi, Dr. R. T. Whitaker, and Dr. P. W. Smith. He is expected to graduate in December 1997 with specializations in image processing and stereoscopic computer vision.