

DEVELOPMENT OF A STEREO VISION SYSTEM FOR OUTDOOR MOBILE
ROBOTS

By

MARYUM F. AHMED

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2006

Copyright 2006

by

Maryum F. Ahmed

ACKNOWLEDGMENTS

I thank Dr. Carl Crane, my supervisory committee chair for his immeasurable support, guidance, and encouragement. I thank Dr. Antonio Arroyo and Dr. Gloria Wiens for serving on my supervisory committee. I also thank David Armstrong for his support on this project.

I thank my fellow students at the Center for Intelligent Machines and Robotics. From them I learned a great deal about robotics, and found great friendships. I thank my family for their undying love and guidance. Without them, I would not be the person I am today.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
ABSTRACT.....	x
CHAPTER	
1 INTRODUCTION	1
1.1 Purpose of Research.....	1
1.2 Stereo Vision.....	2
1.2.1 Some Benefits of Stereo Vision.....	2
1.2.2 Basic Stereo Vision Principles.....	2
1.3 Statement of Problem.....	5
2 MESSAGING ARCHITECTURE.....	6
2.1 Joint Architecture for Unmanned Systems	6
2.2 Sensor Architecture.....	7
3 REVIEW OF RELEVANT LITERATURE AND PAST WORK	11
3.1 Mars Exploration Rover.....	11
3.1.1 Overview.....	11
3.1.2 Algorithm.....	12
3.1.3 Additional Testing.....	13
3.2 Nomad.....	13
3.2.1 Overview.....	13
3.2.2 Lighting and Weather	14
3.2.3 Terrain.....	15
3.3 Hyperion	15
3.3.1 Overview.....	15
3.3.2 Filtering Algorithms	16
3.3.3 Traversability Grid.....	17

3.4	Previous Development at CIMAR.....	17
3.4.1	Videre Design Stereo Hardware.....	17
3.4.2	SRI Small Vision System.....	17
4	HARDWARE.....	19
4.1	Lenses.....	19
4.1.1	Iris.....	19
4.1.2	Focal Length.....	21
4.2	Cameras.....	21
4.3	Image Transfer.....	22
4.3.1	Video Signal Formats.....	22
4.3.2	Frame Grabbers.....	23
4.4	System Chosen.....	23
5	SOFTWARE.....	26
5.1	Image Rectification and Camera Calibration.....	26
5.2	Calculation of 3D Data Points.....	28
5.2.1	Subsampling and Image Resolution.....	28
5.2.1.1	Single pixel selection subsampling.....	29
5.2.1.2	Average subsampling.....	29
5.2.1.3	Maximum value subsampling.....	30
5.2.1.4	Minimum value subsampling.....	30
5.2.2	Stereo Correlation.....	30
5.3	Traversability Grid Calculation.....	32
5.4	Graphical User Interface.....	35
6	TESTING and results.....	39
6.1	Subsample Method.....	39
6.2	Image Resolution.....	40
6.3	Multiscale Disparity.....	41
6.4	Pixel Search Range.....	43
6.5	Horopter Offset.....	43
6.6	Correlation Window Size.....	45
6.7	Confidence Threshold Value.....	46
6.8	Uniqueness Threshold Value.....	47
6.9	Final Parameters Selected.....	48
6.9	Range.....	49
6.10	Auto-Iris.....	49
7	CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK.....	51
7.1	Conclusions.....	51
7.2	Recommendations for Future Work.....	52

APPENDIX

A SAMPLE CALIBRATION FILE.....53

B IMAGES FROM TESTING55

C RESULTS FROM FINAL SELECTED STEREO PROCESSING
PARAMETERS.....65

LIST OF REFERENCES76

BIOGRAPHICAL SKETCH78

LIST OF TABLES

<u>Table</u>	<u>page</u>
2-1. Meaning of grid cell values	10
5-1. Traversability values assigned to each dihedral angle range.....	35
6-1. Parameter values for subsample method test.....	39
6-2. Number of pixels correlated for each subsample method.....	40
6-3. Parameter values for image resolution test.....	41
6-4. Parameter values for multiscale disparity test	42
6-5. Number of pixels correlated with and without multiscale processing.....	42
6-6. Parameter values for pixel search range test.....	43
6-7. Parameters for horopter offset test.....	44
6-8. Parameters for correlation window size test.....	45
6-9. Parameters for confidence threshold value test	46
6-10. Parameters for uniqueness threshold value test	47
6-11. Final Stereo Processing Parameters.....	49

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1-1. Vehicles developed for the first two Defense Advanced Research Projects Agency (DARPA) Grand Challenges.....	2
1-2. Geometry of stereo vision.....	3
2-1. Overview of sensing system.....	8
2-2. World and traversability grid.....	9
3-1. Videre Mega-D Wide Baseline stereo cameras mounted on the Navigation Test Vehicle.....	18
4-1. Diagram of hardware and interfacing chosen for system.	25
5-1. Image pairs before and after rectification.....	27
5-2. Known target calibration images.....	27
5-3. SRI Calibration application.	28
5-4. Single pixel selection subsampling.....	29
5-5. Average subsampling.....	29
5-6. Maximum value subsampling.....	30
5-7. Minimum value subsampling.....	30
5-8. Coordinate transformations.....	33
5-9. Stereo Vision Utility.....	36
5-10. OpenGL windows showing the 3D point clouds.....	37
5-11. OpenGL window displaying the best fitting planes.....	38
6-1. Acceptable horopter values for a series of images.....	44
6-2. Acceptable correlation window size values for a series of images.....	45

6-3. Acceptable confidence threshold values for a series of images.....	47
6-4. Acceptable uniqueness threshold values for a series of images	48
6-5. Scene without auto-iris function.	50
6-6. Scene with auto-iris function	50
B-1. Original Images from Subsample Test.....	56
B-2. Disparity image results from testing with and without multiscale disparity processing	57
B-3. Disparity image and traversability grid results from testing with different image resolutions.....	58
B-4. Disparity image and traversability grid results from testing with different pixel search ranges.	62
C-1. Results from stereo processing.....	66

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

DEVELOPMENT OF A STEREO VISION SYSTEM FOR OUTDOOR MOBILE
ROBOTS

By

Maryum F. Ahmed

August 2006

Chair: Carl D. Crane, III

Major Department: Mechanical and Aerospace Engineering

A stereo vision system was developed for the NaviGator, an autonomous vehicle designed for off-road navigation at the Center for Intelligent Machines and Robotics (CIMAR). The sensor outputs traversability grids defined by the CIMAR Smart Sensor Architecture.

Stereo vision systems which have been developed in the past and previous research at CIMAR were examined. Hardware chosen for the system includes auto-iris lenses for improved outdoor performance, s-video cameras and a four frame grabber PCI card for digitizing the analog s-video signal.

Software from SRI International was used for image rectification and the calculation of camera calibration parameters. The SRI stereo vision library was then used for 3D data calculation. With the 3D data, a least squares plane fitting algorithm was used to find the slope of the terrain in each traversability grid cell. This information was used to give the cell a traversability rating.

Tests were performed to find the best image subsampling method and image processing resolution as well as the benefit of multiscale processing. Tests were also performed to find the optimal set of stereo processing parameters. These parameters included pixel search range, horoptor offset, correlation window size, confidence threshold and uniqueness threshold.

CHAPTER 1 INTRODUCTION

The Center for Intelligent Machines and Robotics (CIMAR) in the Mechanical and Aerospace Engineering Department at the University of Florida has researched many aspects of autonomous ground vehicles. This study focused on developing a stereo vision system for autonomous outdoor ground vehicles. This vision system was designed to tackle the specific problems associated with such vehicles and to be integrated into the CIMAR sensor architecture.

1.1 Purpose of Research

This study had two separate goals: first, to support Team CIMAR in the Defense Advanced Research Projects Agency (DARPA) Grand Challenge; then, to support the Air Force Research Laboratory (AFRL) autonomous ground vehicle program.

The DARPA Grand Challenge was a Department of Defense initiative designed to advance research in the field of high-speed outdoor mobile robotics. The competition was to develop an unmanned ground vehicle that could navigate the rough terrain of an approximately 140 mile race course through the Mojave Desert. The vehicles were allowed no outside influence other than Satellite retrieved Global Position System (GPS) data. Therefore, all obstacle avoidance, terrain estimation and path detection had to be done by sensors on the vehicle. The first race was in March 2004, and the second race was in October 2005. After each race, the ideas were applied to related applications at the Air Force Research Laboratory [1]. Figure 1-1 shows the 2004 and 2005 vehicles developed for the first two Grand Challenge events.



Figure 1-1. Vehicles developed for the first two Defense Advanced Research Projects Agency (DARPA) Grand Challenges A) The NaviGator for the 2004 event. B) The NaviGator for 2005 event.

1.2 Stereo Vision

1.2.1 Some Benefits of Stereo Vision

On a robot, stereo vision can be used to locate an object in 3D space. It can also give valuable information about that object (such as color, texture, and patterns that can be used by intelligent machines for classification). A visual system, or light sensor retrieves a great deal of information that other sensors cannot.

Stereo vision is also a passive sensor, meaning that it uses the radiation available from its environment. It is non-intrusive as it does not need to transmit anything for its readings. An active sensor sends out some form of energy into the atmosphere, which it then collects for its readings. For example, a laser sends out light that it then collects; and radar sends out its own form of electromagnetic energy. A passive sensor is ideal when one wants to not influence the environment or avoid detection.

1.2.2 Basic Stereo Vision Principles

Artificial stereo vision is based on the same principles as biological stereo vision. A perfect example of stereo vision is the human visual system. Each person has two eyes that see two slightly different views of the observer's environment. An object seen by the right eye is in a slightly different position in the observer's field of view than an object

seen by the left eye. The closer the object is to the observer, the greater that difference in position. Anybody can see this for oneself by holding up a finger in front of his or her face and closing one eye. Line the finger up with any object in the distance. Then switch eyes and watch the finger jump.

An artificial stereo vision system uses two cameras at two known positions. Both cameras take a picture of the scene at the same time. Using the geometry of the cameras, the geometry of the environment can be computed. As in the biological system, the closer the object is to the cameras, the greater its difference in position in the two pictures taken with those cameras. The measure of that distance is called the disparity.

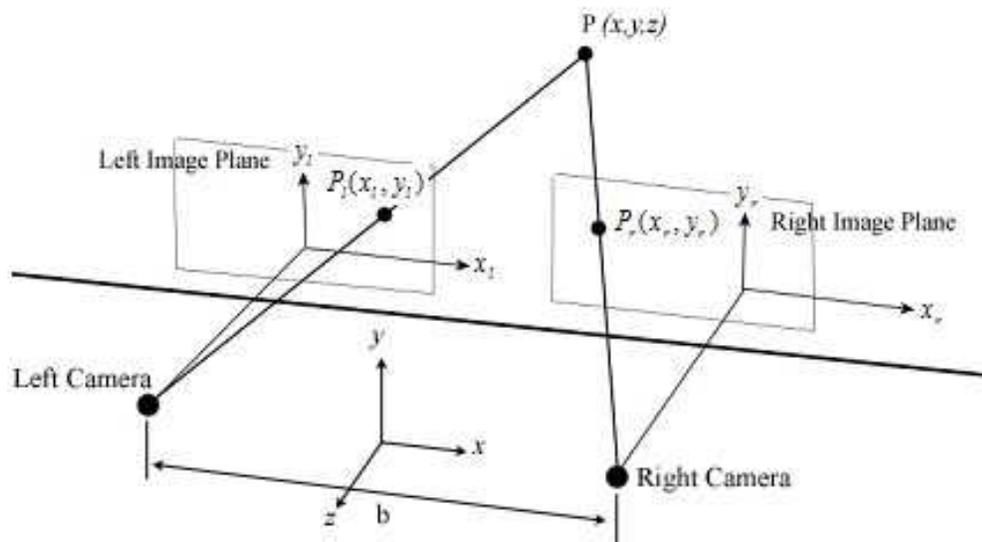


Figure 1-2. Geometry of stereo vision

Figure 1-2 illustrates the geometry of stereo vision. In this example, the optical axes of the cameras are aligned parallel and separated by a baseline of distance, b . A coordinate system is attached in which the x -axis is parallel to the baseline and the z -axis is parallel to the optical axes. The points labeled “Left Camera” and “Right Camera” are the focal points of two cameras. The distance f is the perpendicular distance from each

focal point to its corresponding image plane. Point P is some point in space which appears in the images taken by these cameras. Point P has coordinates (x, y, z) measured with respect to a reference frame that is fixed to the two cameras and whose origin is at the midpoint of the line connecting the focal points. The projection of point P is shown as P_r in the right image and P_l in the left image and the coordinates of these points are written as (x_r, y_r) and (x_l, y_l) in terms of the image plane coordinate systems shown in the figure. Note that the disparity defined above is $x_l - x_r$. Using simple geometry,

$$\frac{x_l}{f} = \frac{x + b/2}{z} \quad (1-1)$$

$$\frac{x_r}{f} = \frac{x - b/2}{z} \quad (1-2)$$

$$\frac{y_l}{f} = \frac{y_r}{f} = \frac{y}{z} \quad (1-3)$$

Note that

$$\frac{x_l - x_r}{f} = \frac{b}{z} \quad (1-4)$$

These equations can be rearranged to solve for the coordinates (x, y, z) of Point P.

$$x = b \frac{(x_l + x_r)/2}{x_l - x_r} \quad (1-5)$$

$$y = b \frac{(y_l + y_r)/2}{x_l - x_r} \quad (1-6)$$

$$z = b \frac{f}{x_l - x_r} \quad (1-7)$$

Equations 1-1 through 1-7 show that distance is inversely proportional to disparity and that disparity is directly proportional to the baseline. When cameras are aligned horizontally, each image shows a horizontal difference, $x_l - x_r$, in the location of P_r and P_l , but no vertical difference. Each horizontal line in one image has a corresponding

horizontal line in the other image. These two matching lines have the same pixels, with a disparity in the location of the pixels. The process of stereo correlation finds the matching pixels so that the disparity of each point can be known.

Note that objects at a great distance will appear to have no disparity. Since disparity and baseline are proportional, increasing the baseline will make it possible to detect a disparity in objects that are farther away. However, it is not always advantageous to increase the baseline because objects that are closer will disappear from the view of one or both cameras [7].

1.3 Statement of Problem

The task of developing a stereo vision system presents many issues with both software and hardware. If the system is to be used outdoors, problems with variable lighting and weather are added. A system where the scene in the images is not stationary adds timing issues with respect to image capture. Mounting this system on a robotic platform which traverses a rugged landscape adds vibrations to the system which can sometimes be intense. The stereo vision system must accomplish the following tasks:

- Capture two images of the scene: This requires two cameras and two camera lenses. This is mostly a hardware issue (Chapter 4).
- Transfer these images to a computer for processing: This may be done with capture cards or some other way of digital data transfer such as firewire. This is both a hardware and a software issue (Chapter 4).
- Process the images for 3D data: This requires stereo processing software which may be purchased (Chapter 5).
- Process the 3D data into a traversability grid: This requires grid computing software which must be written for this task (Chapter 5).
- Send the grid to the Smart Arbiter: This requires the application of the CIMAR Sensor Architecture (Chapter 2).

CHAPTER 2 MESSAGING ARCHITECTURE

2.1 Joint Architecture for Unmanned Systems

The Joint Architecture for Unmanned Systems (JAUS) is an initiative to create an architecture for unmanned systems and is mandated for use by all programs in the Joint Robotics Program. This messaging architecture is for communicating among all computing nodes in an unmanned system. JAUS must satisfy the following constraints [5]:

- **Platform independence:** No assumptions about the vehicle are made (i.e. tracked vehicle, omnidirectional vehicle).
- **Mission isolation:** Developers may build their systems for any mission with any set of tasks.
- **Computer Hardware Independence:** Any computing and sensing technology may be used. Computing power on individual systems can be upgraded throughout the system's lifespan.
- **Technology Independence:** Like computer hardware independence, the technology used in system development should be unrestrained (i.e. vision or range finding for obstacle detection).

JAUS defines a system hierarchy which consists of the following levels:

- **System:** A system is a grouping of at least one subsystem. An example system might include several vehicles along with an operator control unit (OCU) and several signal repeaters.
- **Subsystem:** A subsystem is an independent unit which may be something such as a single vehicle, an OCU, or a single signal repeater. The NaviGator vehicle is a subsystem.

- **Node:** A node is a “black box” that contains all the hardware and software for completing a specific task for the subsystem. The stereo vision computer, software, cameras, internal messaging and interfacing hardware together make up one node. The specific node configuration is left to the developer to design.
- **Component:** A component is a single software entity which performs a specific function. The Stereo Vision Smart Sensor software is a component.
- **Instance:** Instances allow for component redundancy. Several instances of the same component may run on the same node.
- **Message:** A message is a communication between components. In order for a system to be JAUS compatible, all JAUS defined components must communicate with JAUS defined messages.

JAUS does not define an adequate messaging protocol for environmental sensors and so the CIMAR Sensor Architecture was developed.

2.2 Sensor Architecture

The CIMAR sensor architecture was designed to compliment JAUS by conforming to all the above mentioned constraints. The architecture integrates many different sensors seamlessly and painlessly. Each sensing technology has different performance capabilities under different conditions. Many perform different functions. A robust outdoor robot often incorporates many of these sensors. Therefore it is necessary to develop methods of combining each sensor’s view of the world into one world view which can be used by the path planning components to make decisions.

Figure 2-1 shows the flow of information through the CIMAR sensor system. Each sensor collects data in its inherent format and processes that data into traversability grids with its own computer. Together, the sensor and computer are known as a Smart Sensor Node. Each Smart Sensor and the Spatial Commander (which contains all a priori knowledge about the world and can be thought of as a “pseudo-sensor) sends its traversability grid results to the Smart Arbiter. The Smart Arbiter makes decisions about

which data is the most reliable and fuses those that it deems reliable into one grid. It then sends this grid to the Reactive Planner for path segment calculations. Figure 2-2 shows the vehicle in the world with the resulting traversability grid.

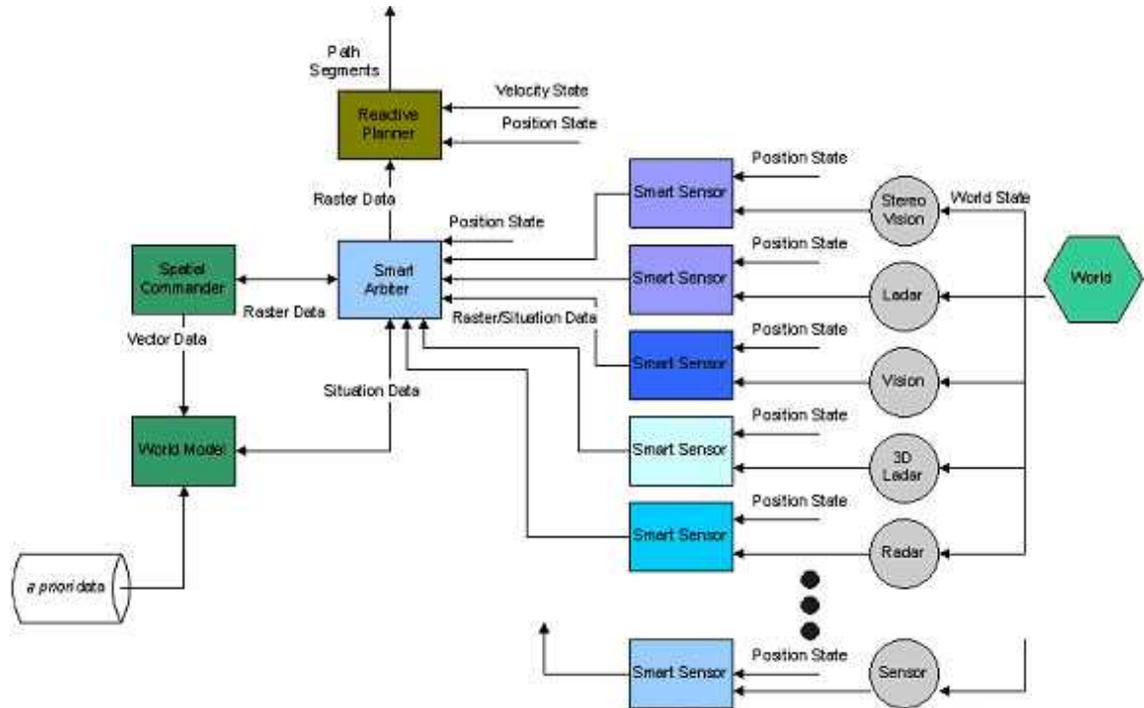


Figure 2-1. Overview of sensing system

The attractiveness of the CIMAR sensor architecture is in its modularity and efficiency. Any sensor may be removed from or added to the system without a significant impact on the Smart Arbiter, hence complying with the requirements set forth by JAUS. To do this, each sensor sends the same size traversability grid using the same messaging rules. The disadvantages of the current system are (1) the resolution of the traversability grid is currently fixed at two values and (2) better estimates of traversability may be made in some instances by considering the raw data from more than one type of

sensor. This last case, however, can easily be incorporated into the system as the two sensor component can be implemented as one “super” smart sensor component.

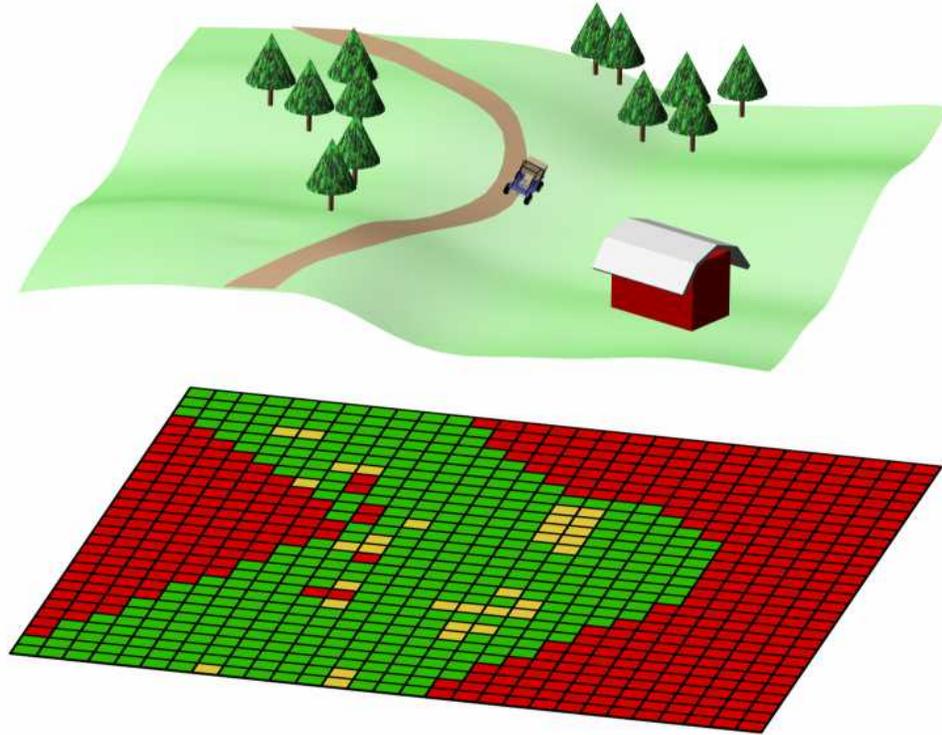


Figure 2-2. World and traversability grid. The smart arbiter fuses data from the Smart Sensors and the Spatial Commander. The final grid contains information about traversability (obstacles and terrain) as well as which areas are out of bounds or in bounds.

The grid has 121 cells \times 121 cells. It is in global coordinates so that the positive horizontal axis points East and the positive vertical axis points North. There are two different resolution modes. The first is a low resolution, long range mode where each cell is 0.5 m \times 0.5 m (making the entire grid 30 m \times 30 m). The second is a high resolution, short range mode where each cell is 0.25 m \times 0.25 m (making the entire grid 15 m \times 15 m). The short range mode should be used when the vehicle is traveling at a slower speed over more challenging terrain. Each grid cell contains a number from 0 to 15. Table 2-1 shows the meaning of each value.

Table 2-1. Meaning of grid cell values

Cell Value	Meaning
0	Only used by the World Model Knowledge Store to indicate out of bounds
1	Nothing to report
2 through 12	Traversability values with 2 meaning absolutely non-traversable, 7 meaning neutral, and 12 meaning absolutely traversable
13	Reserved for “failed/error” – this value tells the recipient that there was some kind of problem in (re)calculating the proper value for that cell
14	Reserved for “unknown” – this value tells the recipient that the traversability of that cell has never been estimated
15	Reserved for marking a cell as having been traversed by the vehicle and is mainly used for display purposes

The Smart Arbiter also outputs the same information in the same format. If the use of only one sensor is desired for the system, it is possible to send the information directly to the reactive planner and bypass the arbiter completely [11].

CHAPTER 3 REVIEW OF RELEVANT LITERATURE AND PAST WORK

The purpose of this chapter is to discuss prior stereo vision systems developed for other outdoor mobile robots. The systems that will be discussed were developed by NASA's Jet Propulsion Laboratory (JPL) for the Mars Exploration Rover missions [3][6], and Carnegie Melon University (CMU) for their Nomad vehicle [15] and Hyperion vehicle [14][16].

3.1 Mars Exploration Rover

3.1.1 Overview

The high-profile and highly successful Mars Exploration Rover missions used autonomous passive stereo vision to create a local map of the terrain to be used for navigation. There were many reasons that stereo vision was chosen for the task. One reason is that stereo vision is a passive sensing technology, thus the sensor requires less power for operation than an active sensor that must emit a signal. Also, if the cameras do not have a wide enough field of view, multiple cameras may be added to view the scene and thus, no moving parts are necessary. This reduces the number of failure points. With the idea of minimizing failure points in mind, the two stereo cameras were mounted rigidly on a camera mast rather than a moving head.

Gennery's CAHVORE formulation was used for camera calibration. This method uses a pair of images of a known calibration target to create geometric models of the camera lenses. It assumes that the system will maintain its geometry over a long period of time.

3.1.2 Algorithm

The first step in their algorithm is to reduce the image size using pyramid level reduction. Each level of the reduction decreases the image size by half the length and half the height by averaging the pixel values. Each image reduction reduces the computation by a factor of eight: two from each spatial dimension and two from a reduction in the number of disparities that must be searched [3]. Additional advantages of lowering the image resolution are that stereo correlation is less sensitive to lens focus and errors in calibration. The downside is that the depth resolution loses precision causing a decrease in 3D range accuracy [6].

Pairs of images are then rectified using the camera lens models created earlier in pre-deployment. The Laplacian of the images is computed to remove pixel intensity bias. A one dimensional correlator is then used to find potential matches for the pixels in the images. The correlator uses a square pixel window [3]. Testing showed that for most cases, a smaller window size such as 7×7 worked better with a more textured scene and a larger window size such as 29×29 worked better with a less textured scene [6]. The disparity range is the range of pixels to be searched for a match. It is derived from the range of depth values (the range in front of the cameras) to be searched. The larger the searchable pixel range, the smaller the minimum distance required for detection between the cameras and the object. The downside is that when the disparity range is large, it takes longer to search for matches. The disparity values found are then tested for reliability by several filters so that mistakes may be thrown out. The disparity value and the camera model described in Chapter 1 are used to project the 3D points [3].

The system used for navigation is called the Grid-based Estimation of Surface Traversability Applied to Local Terrain (GESTALT), which is modeled after Carnegie

Mellon's Morphin algorithm. GESTALT uses a grid model of the world in local coordinates. Each square cell of the grid is equally sized and spaced and is approximately the size of one rover tire. Each cell holds an 8 bit value, which is a measurement of the terrain's "goodness" and "certainty". The cell may also be marked as "unknown". [3]

3.1.3 Additional Testing

As mentioned previously, a performance analysis and validation of the system tested parameters such as image resolution and correlation window size. Other factors such as the effects of vertical misalignment, focus issues and stereo baseline were also tested.

Vertical misalignment is caused by poor calibration parameters, and thus incorrect image rectification. This was tested by intentionally shifting one image and measuring the number of correctly calculated disparities. As is expected, the error in disparity values increased as the misalignment increased.

Focus was tested by blurring one image in the pair. It was found that good focus was very important for accurate disparity calculation. Cameras with a narrow field of view were especially sensitive to focus issues. As mentioned previously, the effects of bad focus could in some cases be offset by image subsampling.

Researchers at JPL anticipate that their analysis and experimental validation will be useful in the development of other correlation based stereo systems. [6]

3.2 Nomad

3.2.1 Overview

Nomad was a robot developed by CMU, mostly in the late 1990's and in 2000 for the Robotic Antarctic Meteorite Search program. It was designed to autonomously navigate the harsh polar terrain and to find and classify meteorites. For navigation,

Nomad used a combination of stereo vision, monocular vision and laser range finders for terrain classification and obstacle avoidance. The stereo vision system consisted of two pairs of black and white stereo cameras (four cameras total) mounted on a camera mast 1.67 m above the ground [15]. The optimal configuration for the stereo cameras and mast was determined by a nonlinear programming formulation specifically designed for this project. The optimal baseline for the cameras was found to be 0.59 m [4].

3.2.2 Lighting and Weather

Because the Antarctic terrain consists greatly of snow and ice it is highly reflective. During the summer season there is always daylight, but the sun stays low in the sky. These two factors cause a significant amount of glare and light saturation in images. Luckily, the horizon in their deployment area was occupied by hills that blocked direct sunlight from the cameras. Researchers found that they were able to regulate the ambient light with the camera's iris and shutter and produce good images. They used linear polarizing filters to reduce glare. However, testing showed that the reduction in glare did not significantly increase the number of pixels matched in stereo processing.

To test the effects of the sun on stereo processing, images were taken while the vehicle was driven in circles. At one point in the circle the vehicle faces into the sun and at the opposite point in the circle it faces completely away. Sun position had minimal effect. It was shown that there was very little variation in the number of pixels matched at different positions in the circle.

The number of pixels matched on sunny days versus overcast days was also compared. This showed a more drastic change. On average, stereo processing matched about twice as many pixels on sunny days as on overcast days. However, the researchers

did note that on overcast days the terrain had so little contrast that even humans had great difficulty perceiving depth.

Images were also taken during a third weather condition, which may not have a significant correlation to this project but is still interesting to note. Blowing snow seemed to have no effect on stereo processing. The snow was difficult to see in the images and so stereo processing found about as many pixels with the blowing snow as without.

However, the laser range finders were significantly impacted by the presence of blowing snow. They failed to provide accurate data under these conditions [15].

3.2.3 Terrain

Stereo vision was tested on snow, blue ice and moraine. Moraine is a rocky terrain. These three different types of terrain are common in Antarctica. Results showed that terrain type had very little effect on the number of pixels that stereo processing was able to match [15].

3.3 Hyperion

3.3.1 Overview

Hyperion is a robot developed by CMU as an experiment in sun-synchronous robotics. A sun-synchronous robot must expend minimum energy and gather maximum solar energy while completing its mission. Hyperion uses a stereo vision based navigation system designed for robustly crossing natural terrain.

A route is generated from the mission planner, which uses a priori elevation maps and knowledge of the movement of the sun. The resolution of the elevation map is typically 25m or greater so the mission planner can only navigate around very large obstacles like hills and valleys. For smaller obstacles, there is a motion planner (also

called navigator) for more precise navigation. The navigator uses maps built from stereo vision.

The system also uses a laser range finder that acts as a “virtual bumper” to warn the vehicle that danger is imminent. If it detects an obstacle it issues an immediate stop command.

3.3.2 Filtering Algorithms

Areas of low texture in an image provide poor results for stereo matching and therefore unreliable three dimensional data. Most stereo vision systems filter out this unreliable data and are unable to report information for image areas of low texture. This causes the navigation system to have no information by which to make decisions. Most navigation systems would err on the side of caution and not attempt to traverse these areas. However, the designers of Hyperion decided to treat undetected terrain as safe rather than dangerous. This assumption was based on the nature of the terrain. They expected to find sparse obstacles and softly rolling terrain. The laser also added an element of safety.

The terrain was assumed to be a smoothly varying $2\frac{1}{2}$ dimensional surface. Large spikes in the data were assumed to be noise and were discarded. The disparity of each pixel was compared with the disparity of its neighbors and thrown out if the difference was larger than a threshold. This allowed small patches of data to be thrown out while large patches that are more likely to be accurate can remain.

A second filtering method based on the distance of each three dimensional data point from the assumed ground plane was used. If the distance was too great, the point was thrown out. These filtering methods allowed for a reduction in errors while still maintaining dense point clouds [14].

In testing, most of the terrain was detected by the stereo vision system and all of it was detected by the laser range finder [16].

3.3.3 Traversability Grid

Navigation was done based on a traversability grid that the stereo vision system created. Each cell gave an estimate of roll, pitch, and roughness for that area. Each cell was approximately the vehicle's size. The roll and pitch was computed using the data in the entire cell. The roughness is estimated by looking at much smaller sub-cell's. With this information, the Morphin Algorithm determined the preferred path [13].

3.4 Previous Development at CIMAR

3.4.1 Videre Design Stereo Hardware

Previous stereo vision work at CIMAR has been done using hardware from Videre Design. Videre Design specializes in stereo vision hardware and software for embedded applications. The current work began with testing of Videre cameras and software and progressed from there in an attempt to address the specific needs of the NaviGator robot.

Two of the Videre camera rigs used at CIMAR are the Mega-D Wide Baseline and the Mega-DCS Variable Baseline. As the names imply, the Variable Baseline cameras can be moved with respect to each other; the Wide Baseline stereo rig has two cameras at a wider, fixed position with respect to each other. The variable baseline cameras must be calibrated after each move. The Mega-D camera pair can be seen in Figure 3-1. They have a Firewire IEEE 1394 interface. Hence, the computer hardware used for the task must also have an IEEE 1394 interface [2].

3.4.2 SRI Small Vision System

One huge advantage of using these cameras is that Videre Design works closely with SRI International's Artificial Intelligence Center. SRI has developed the SRI Stereo

Engine, an efficient implementation of stereo correlation. The Stereo Engine Library provides a C++ library of functions for adding stereo processing to user written applications [9]. The Stereo Engine is incorporated into the SRI Small Vision System (SVS), a standard development environment that runs on Linux and Windows operating systems. SVS also contains libraries for image rectification and camera calibration.

Videre camera rigs have interfaces to SVS so the camera hardware and stereo software can be easily integrated. When the cameras are connected to the computer's IEEE 1394 interface, SVS library functions can be used to grab and process image data in a very user friendly way.



Figure 3-1. Videre Mega-D Wide Baseline stereo cameras mounted on the Navigation Test Vehicle

The Videre Design system can be used to accomplish tasks 1 through 3 listed in Chapter 1. The results are acceptable when used indoors in constant, controlled lighting. The system breaks down when used in variable lighting conditions.

CHAPTER 4 HARDWARE

This Chapter details the hardware necessary for steps 1 and 2 listed in Chapter 1 and describes some of the options available for this hardware. Finally the options chosen to improve the system are described.

4.1 Lenses

The camera lens is the interface between the environment and the sensor. A properly chosen lens will improve the quality and range of the results.

4.1.1 Iris

A large issue with the use of computer vision in an outdoor environment is variable lighting. Whether monocular or stereo, if the cameras being used create images from the visible light spectrum this will be an issue. Image processing will yield different qualities of results based on the lighting situation.

In conditions where the camera is gathering too much light, the image becomes over-exposed and will appear washed out or even completely white. Conversely, if the camera does not gather enough light, the image is under-exposed and large areas will appear black. In an indoor testing environment, the amount of light in the room can be fixed. In an outdoor environment, the lighting changes substantially based on things such as time of day, weather, camera orientation with respect to the sun, and shaded area.

A camera's iris acts much like the iris of a human eye. The iris is an adjustable aperture, which can be made larger or smaller. With a larger aperture, more light is allowed to enter the camera. A smaller aperture allows less light. Camera lenses can have

a manual iris or an auto-iris. The manual iris is adjusted by the user while the auto-iris uses feedback from the camera to make adjustments.

The lens's f-stop is a measurement of the size of its iris aperture. The number represents the relationship between the diameter of the opening and the focal length.

$$F - stop = \frac{f}{d} \quad \text{Eq. 4.1}$$

So for example, f2 means that the diameter is half the focal length and f16 means that the diameter is 1/16th the focal length. Therefore, the larger the number, the smaller the aperture [10].

Auto-iris lenses come in two different types: DC and video. In DC lenses, the camera processes the image and sends a DC signal to the lens to open or close the iris. A video lens receives a video signal from the camera and does the processing by which it makes a decision about whether to open or close the iris. Basically, with DC lenses, the camera does the processing and with video lenses, the lens does the processing.

In fixed lighting conditions, the user may set the iris of the camera to gather an optimal level of light prior to performing the task. An optimal level for stereo vision is one in which the images show features with the greatest texture. In the outdoor setting, manually adjusting the iris before use is not good enough. As described before, the lighting changes based on many variables and images will not always have the proper exposure for image processing. If the stereo system were stationary this might not be such a large issue as the user could adjust the iris. However, a fully autonomous mobile robot must be a "hands-off" system during performance. Also, in an effort to protect the cameras from the environment, the cameras are enclosed in a protective casing that does not allow convenient access to the lens.

4.1.2 Focal Length

Camera lenses may have variable focal lengths or fixed focal lengths. A variable focal length lens can zoom in and out. For this application, fixed focal length lenses were desired as variable focal lengths would add great complexity to the system.

Lenses with larger focal lengths create images that are zoomed in farther. It was desirable to choose a focal length that would allow the system to detect objects far enough away to provide adequate time for obstacle avoidance. The trade off is that the greater the focal length, the narrower the field of view. The field of view (FOV) of a lens can be computed by

$$FOV_{horizontal} = 2 \tan^{-1}\left(\frac{x}{2f}\right) \quad \text{Eq. 4.2}$$

$$FOV_{vertical} = 2 \tan^{-1}\left(\frac{y}{2f}\right) \quad \text{Eq. 4.3}$$

where x is the horizontal width of the sensor, y is the vertical height of the sensor and f is the lens focal length.

4.2 Cameras

A stereo camera pair must have two identical cameras rigidly mounted so that they will not move with respect to each other. Cameras are available with a multitude of options. Some of the most important questions to consider are what kinds of outputs are required for the task, lens compatibility, shutter speeds, resolution, and ruggedness.

As mentioned previously, an auto-iris lens adjusts its aperture based on camera feedback. If this option is chosen for the lens the camera must contain the proper output (either DC or Video) for the auto-iris. If a manual iris is chosen, no output is required. Another feature that should be considered for lens compatibility is the lens mount type.

Lenses are available in C and CS types. The decision of which camera to use goes hand in hand with the choice of which lens to use.

The decision of which camera to use also goes hand in hand with the choice of method for image transfer from camera to computer. There are several formats for the signal that the camera sends containing the images. The format influences the speed of data transfer, image quality and resolution. As stated above, the Videre cameras use a firewire interface to send a digital image signal. Cameras that send analog signals must use a frame grabber (also called a capture card) to digitize the images. Three common analog video formats are described below.

4.3 Image Transfer

4.3.1 Video Signal Formats

One video signal format is s-video (separated video), also known as Y/C. In this format the camera sends two analog signals, one containing the image's luminance (intensity or Y) information, the other containing the image's chrominance (color or C) information. S-video is usually connected with a round 4-pin mini DIN connector. It has a resolution of 480 interlaced lines in NTSC format and 576 interlaced lines in PAL format.

Another format is composite video, also known as YUV. This format sends three components, one luminance (Y) and two color (U and V), in one composite analog signal. A yellow RCA type connector is usually used to transmit composite video. Like s-video, it can also be used with NTSC or PAL format with the same resolution.

Component RGB video sends three analog signals, one for red, one for green and one for blue. Sometimes one or two more signals are sent with synchronization

information. This format can send images with a resolution up to 1080 progressive scan lines and is better for tasks requiring very high resolution images.

4.3.2 Frame Grabbers

As mentioned previously, analog image signals must be converted to digital images for computer processing. Frame grabbers are used most often and typically consist of hardware that can be inserted into a PCI slot. A video to firewire or USB converter can also be used for getting the digital image. For stereo vision, two images must be transferred to the computer at the same time. If the system uses multiple pairs of cameras, it may be desirable to transfer all of the images to the same computer. This with the input capabilities of the computer hardware used for processing should be taken into consideration when choosing a conversion method.

Because the images will be processed and not simply stored or displayed, it is necessary to choose a frame grabber that comes with a library for programming user applications rather than just commercial software. The NaviGator component code is written in C and C++, therefore a C or C++ library is useful for easy integration. Also, all computers on the NaviGator run the Linux operating system, so Linux drivers are necessary for hardware added to a computer.

4.4 System Chosen

The decision was made to use auto-iris lenses because of the lighting issues discussed previously. Pentax auto-iris lenses were chosen because of their rugged metal threading and low cost. The irises of these lenses are DC type and have a range from F1.2 to F360. Focal lengths of 6mm and 12mm were tested to see which would provide the better data range.

This dictated the direction of the rest of the system hardware. The Videre cameras used previously do not have auto-iris output, so new cameras had to be chosen. Cameras with digital firewire output were the first choice for easy integration with the SRI library, but none were available with auto-iris output. Without the option of firewire, s-video was selected as the ideal image signal because the stereo correlation software only needs image intensity and not color for disparity calculation. Since s-video separates the two signals, it is possible to only grab the intensity signal. This allows for less data transfer and a faster system. If future versions of the system desire the use of color, it is a simple matter to add the color signal to the image capture.

Upon searching for a frame grabber that was suitable for the task, the Matrix Vision mvSIGMA-SQ was selected. This is a PCI card that has four separate frame grabbers with s-video inputs so it can capture up to four images at once. Having more than one frame grabber on the same card allows for software synchronization. Without the software synchronization, a gen-lock cable would be needed to synchronize image capture. It also allows for a more compact system as the computer is only required to have one PCI slot. Although this system uses one stereo pair, future work may incorporate two pairs and this card allows for an easy transition. The card is C programmable and runs on Linux and Windows.

After the lens and image transfer method were selected, the Appro CV-7017H camera model with the correct auto-iris output and s-video output was selected. This camera has been previously used and proven at CIMAR. It was used for monocular lane detection at AUVSI's International Ground Vehicle Competition and for the monocular

Pathfinder component on the first NaviGator during Grand Challenge 2004. Figure 4-1 shows a diagram of the hardware.

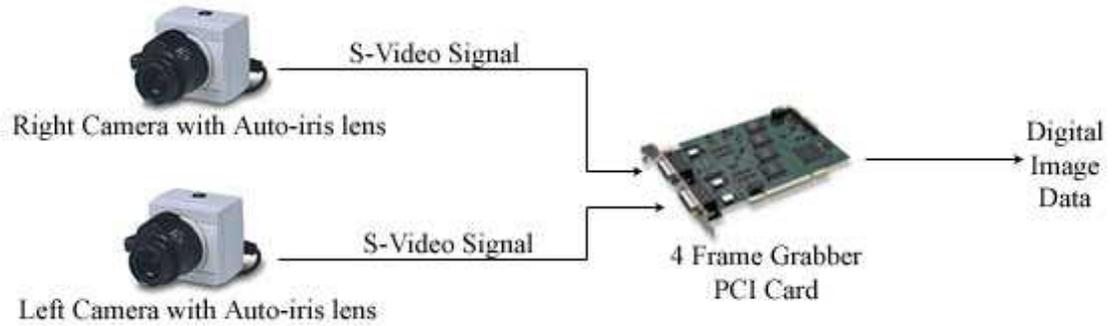


Figure 4-1. Diagram of hardware and interfacing chosen for system.

CHAPTER 5 SOFTWARE

This section describes the commercial stereo vision software that was used and the additional software developed for computing the traversability grids.

5.1 Image Rectification and Camera Calibration

In reality, the cameras will not have perfectly aligned optical axes. Images will also contain some distortion. The main form of distortion in images is radial distortion where the images are compressed towards the edges. This occurs most prominently in wide angle lenses. Another form of distortion is lens decentering where the center of focus of the lens does not line up with the center of the image.

The first step in the application of stereo vision is to change the imperfect images into an idealized stereo pair. Having idealized images makes the process of finding corresponding pixels in the two images easier. First the images are undistorted. Then they are rotated and scaled so that they fit the ideal geometry described in Chapter 1 [8]. Figures 5-1 (a.) and (b.) show a pair of images before rectification. Figures 5-1 (c.) and (d.) show the same image after rectification.

Even without the Videre cameras, the SRI Stereo Engine library was still used for image rectification, camera calibration, and stereo correlation. Rectification and calibration parameters were calculated by taking a series of images of a known target and running the SRI calibration application. Figure 5-2 shows a pair of calibration images. Figure 5-3 shows the SRI calibration application.

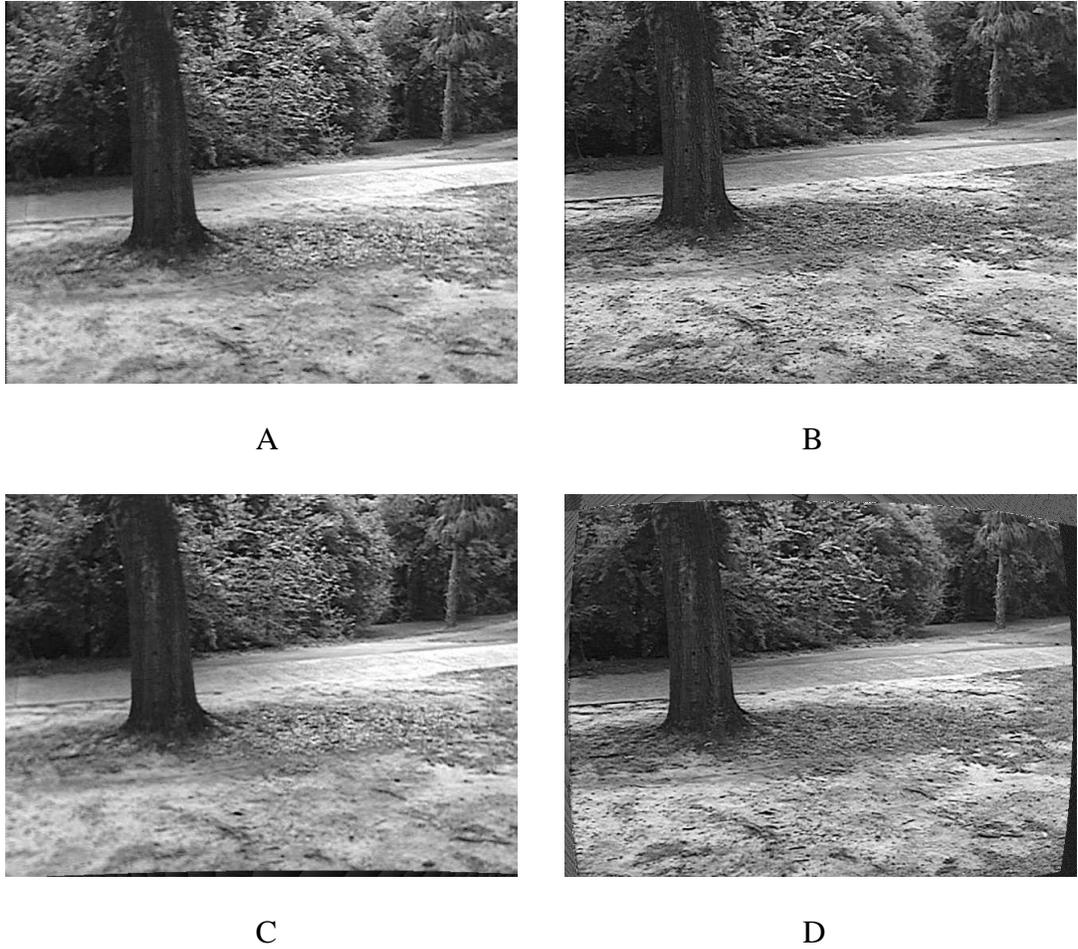


Figure 5-1. Image pairs before and after rectification. A) Left image rectification. B) Right image before rectification. C) Left image after rectification. D) Right image after rectification.

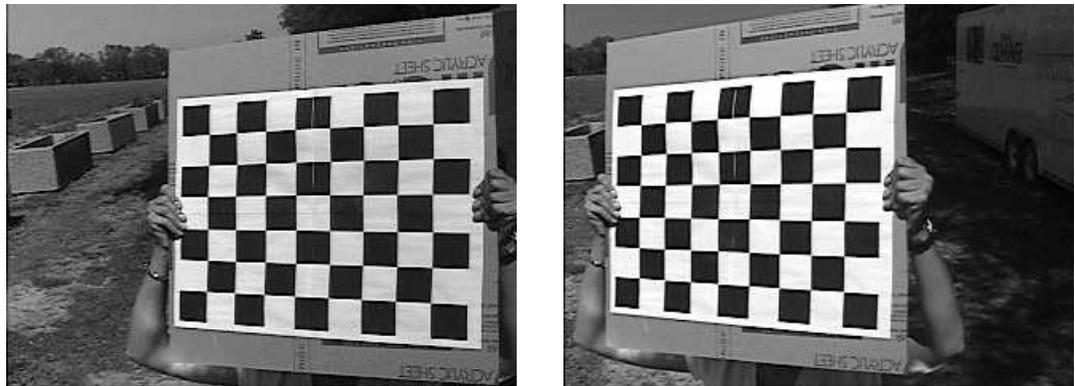


Figure 5-2. Known target calibration images. A series of images of the target are used to calculate image rectification and camera calibration parameters.

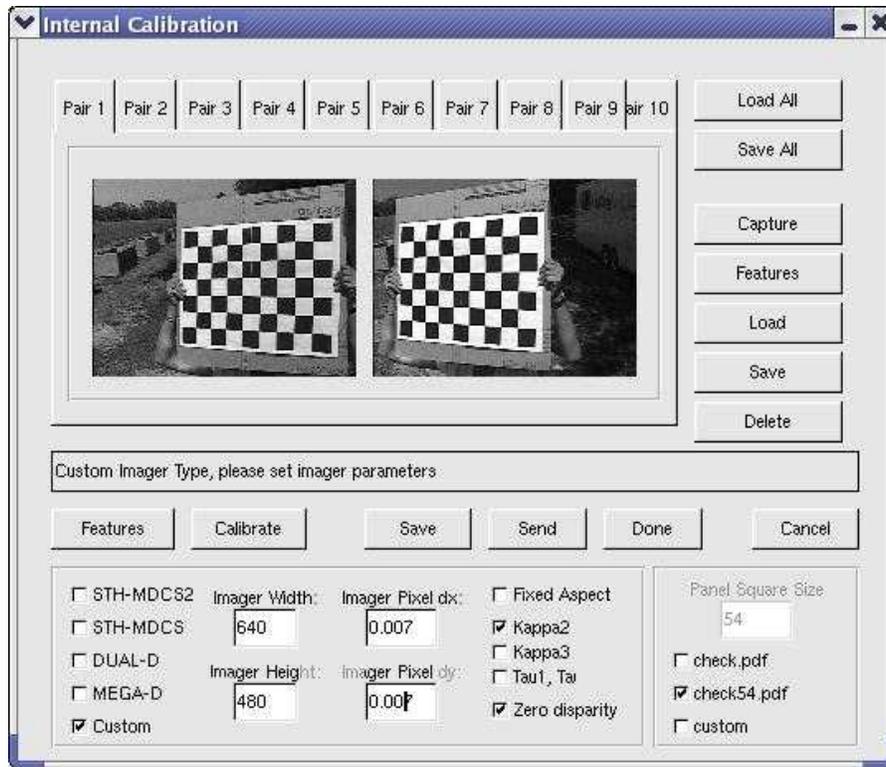


Figure 5-3. SRI Calibration application. Ten calibration image pairs can be loaded into the application and certain camera parameters are set. Then the application finds the rectification and calibration parameters.

A sample calibration file can be seen in Appendix A. When a change is made to the camera configuration, a new file must be computed. This file is then used whenever the component is run until the next time the cameras are moved or the lenses are changed.

5.2 Calculation of 3D Data Points

5.2.1 Subsampling and Image Resolution

The first step in processing was to subsample the images. Several methods of subsampling were tried. Also several image sizes were tried. The larger the image the more detail available for feature finding. However, the use of larger images significantly slows down the system. Images were captured at a resolution of 480×640 pixels. They were either left at this resolution or subsampled to a size of 320×240 or 160×120 .

Different methods of subsampling tested were single pixel selection, averaging, highest value, and lowest value. The results are presented in Chapter 6.

5.2.1.1 Single pixel selection subsampling

This method is computationally the least expensive. One pixel is chosen to replace each block of pixels that is to be reduced. Figure 5-4 illustrates single pixel selection where the upper left pixel is chosen to represent the local 2x2 area of pixels.



Figure 5-4. Single pixel selection subsampling. The upper left pixel of each local 2×2 area is chosen to represent the entire area.

5.2.1.2 Average subsampling

With the average subsampling method, each area is represented by the average value of all the pixels in that area. This method removes noise but does not preserve edges. Figure 5-5 illustrates average subsampling.



Figure 5-5. Average subsampling. The average pixel value of each local 2×2 area is chosen to represent the entire area.

5.2.1.3 Maximum value subsampling

With the maximum value subsampling method, each area is represented by the highest value of all the pixels in that area. With this method, the image will appear slightly lighter. Figure 5-6 illustrates this subsampling method.



Figure 5-6. Maximum value subsampling. The highest pixel value of each local 2×2 area is chosen to represent the entire area.

5.2.1.4 Minimum value subsampling

With the minimum value subsampling method, each area is represented by the lowest value of all the pixels in that area. With this method the image will appear slightly darker. Figure 5-7 illustrates this method.



Figure 5-7. Minimum value subsampling. The lowest pixel value of each local 2×2 area is chosen to represent the entire area.

5.2.2 Stereo Correlation

The SRI C++ library functions performed the stereo correlation. The functions were used for loading the subsampled images from memory, computing the disparity data and projecting the points into 3D space. The results of the SRI algorithms depend greatly

on many correlation variables. These variables can be changed by the user to get the best possible results:

- **Multiscale disparity:** If this option is turned on, the algorithm will calculate disparities with the original image and with an image of 1/2 the size. The hope is that each calculation will find some disparities that the other cannot. The obvious drawback is longer processing time.
- **Number of pixels to search:** the maximum pixel range that will be searched for a match. The larger the distance between matching pixels, the larger the disparity. If the range of pixels that are searched increased, larger disparity values can be found. A larger search range takes more processing time.
- **Horopter offset:** The horopter is the 3D range in front of the cameras that is covered by the stereo algorithm. It is a function of disparity search range, baseline, and the focal length of the lenses. It can be changed by setting an X-offset between the two images. Basically the same number of pixels will be searched but they will be different pixels.
- **Correlation window size:** Correlation compares areas of pixels in the two images. The size of this area is the correlation window size. For example a 7×7 window size attempts to find matching 7×7 areas of pixels in the two images. A larger window size reduces the noise in lower textured areas. The downside is that you lose disparity resolution. Since this system is looking for obstacles in relatively large $0.5\text{m} \times 0.5\text{m}$ areas, a loss of disparity resolution will probably not hurt the results for our application.
- **Confidence threshold value:** Areas are assigned a confidence value based on how textured the area is. The greater the texture, the higher the confidence that the matches found are correct. Areas with low texture can be thrown out if they are below a certain threshold. A high threshold will eliminate most errors, but will also get rid of a significant amount of good data.
- **Uniqueness threshold value:** The uniqueness filter attempts to throw out errors caused by the areas behind objects that can be seen by one camera but not the other. The minimum correlation value of an area must be unique, or lower than all other match values by some threshold. Usually, the areas around objects will have non-unique minima. [9]

The difficulty with selecting the best parameters is that different combinations work better in different situations. The task is to find the combination that gives the best results for the greatest number of situations.

After the correlation has been performed and the disparity has been calculated, there are additional SRI functions for projecting the pixels into 3D space. Those functions use the disparity values with the stereo vision geometry described in Chapter 1.

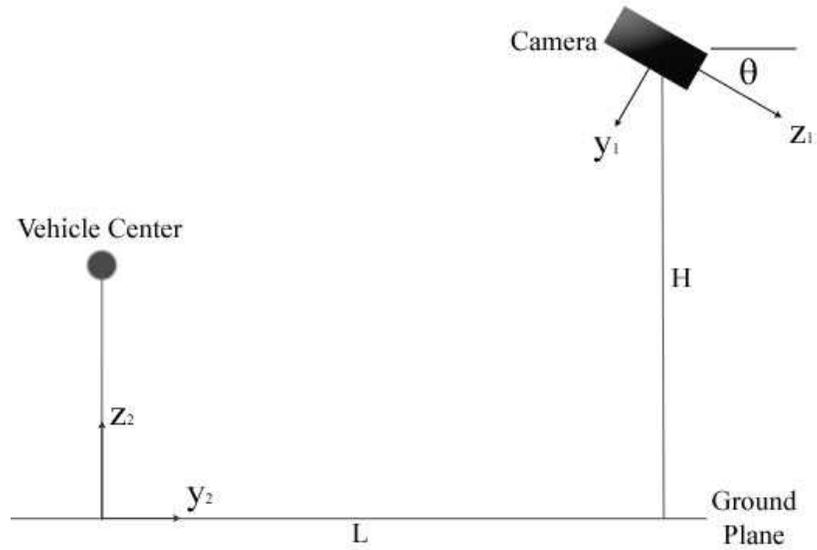
5.3 Traversability Grid Calculation

The next task is to take the 3D point clouds that are within the desired range and perform rotations and translations so that they are in the coordinate system of the traversability grid. Figure 5-8 shows the different coordinate systems involved in the transformation. Equations 5-1 and 5-2 state the two transformation matrices used for this transformation.

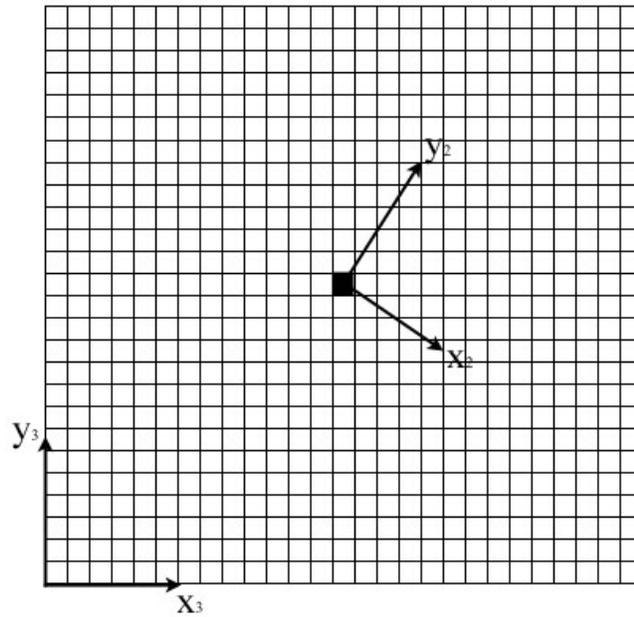
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & L \\ 0 & -\cos(\theta) & -\sin(\theta) & H \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \\ 1 \end{bmatrix} \quad (5-1)$$

$$\begin{bmatrix} \cos(\psi) & \sin(\psi) & 0 & \frac{GridWidth}{2} \\ -\sin(\psi) & \cos(\psi) & 0 & \frac{GridHeight}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ z_2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_3 \\ y_3 \\ z_4 \\ 1 \end{bmatrix} \quad (5-2)$$

Coordinate system 1 is centered on the left camera focal point at a height H from the ground; z_1 is parallel to the camera's optical axis, and y_1 points down relative to the center of the image. Coordinate system 2 is centered on the vehicle ground plane directly below the center of the vehicle; z_2 is up and y_2 points out of the front of the vehicle. θ is the angle between the cameras' optical axis and the horizontal. L is the horizontal distance from the vehicle center to the camera. The vehicle's yaw (ψ) is used to align the y axis with north in coordinate system 3. Coordinate system 3 is centered at the bottom left corner of the traversability grid.



A



B

Figure 5-8. Coordinate transformations. A) First step in coordinate transformation. The box represents the camera. Coordinate system 1 is the camera centered coordinate system B) Second step in coordinate transformation. Coordinate system 2 is the same as in the first step. y_3 points north, x_3 points east.

Once the points are in the correct coordinate system, the number of points that fall in each cell is counted. Then, for each cell, if the number of points is over a threshold value, the traversability value is calculated for that cell. Otherwise, a value of 14 meaning “unknown” is assigned to the cell.

The best fitting plane is found for the points in each cell using the least squares method. With this method, the least-squares error for the flat plane model should be minimized. The least-squares error is

$$LSE = \sum_{j=1}^n (f(x_j, y_j) - z_j)^2 \quad (5-3)$$

which becomes

$$LSE = \varepsilon(a, b, c) = \sum_{j=1}^n (ax_j + by_j + c - z_j)^2 \quad (5-4)$$

where

$$f(x, y) = ax + by + c \quad (5-5)$$

is the equation for the plane. The derivatives of Equation 5.4 should be taken with respect to each coefficient and set equal to 0.

$$\frac{\delta \varepsilon}{\delta a} = \sum_{j=1}^n 2(ax_j + by_j + c - z_j)x = 0 \quad (5-6)$$

$$\frac{\delta \varepsilon}{\delta b} = \sum_{j=1}^n 2(ax_j + by_j + c - z_j)y = 0 \quad (5-7)$$

$$\frac{\delta \varepsilon}{\delta c} = \sum_{j=1}^n 2(ax_j + by_j + c - z_j) = 0 \quad (5-8)$$

The equations can then be solved for a , b , and c [12].

The vehicle ground plane is assumed to be the true ground plane. The dihedral angle between the cell's best fit plane and the vehicle ground plane is found by comparing the normals to the planes. The angle is checked against threshold values associated with each traversability value. Table 5-1 shows the traversability value for each angle range. The assigned traversability value is then sent the Smart Arbiter.

Table 5-1. Traversability values assigned to each dihedral angle range.

Traversability Cell Value	Dihedral Angle
2	55°
3	50° < 55°
4	45° < 50°
5	40° < 45°
6	35° < 40°
7	30° < 35°
8	25° < 30°
9	20° < 25°
10	15° < 20°
11	10° < 15°
12	0° < 10°

5.4 Graphical User Interface

A stereo vision utility was created to assist in development and testing. The utility can be run with live stereo video or with a saved image. Figure 5-9 shows the graphical user interface.

The top left window shows the left camera image. The window beneath that shows the right camera image when stereo processing is turned off; it shows the disparity image when stereo processing is turned on (as in the figure). The window on the right shows the traversability grid output. This utility does not receive GPS so the vehicle is always assumed to be pointing north.

The user can click the "Save Images" button to save the left and right images. The user can load a saved image by clicking the "Use Stored Image" button. Much of the

testing was performed by saving images in the field and loading them later where the effects of stereo parameters can be analyzed.

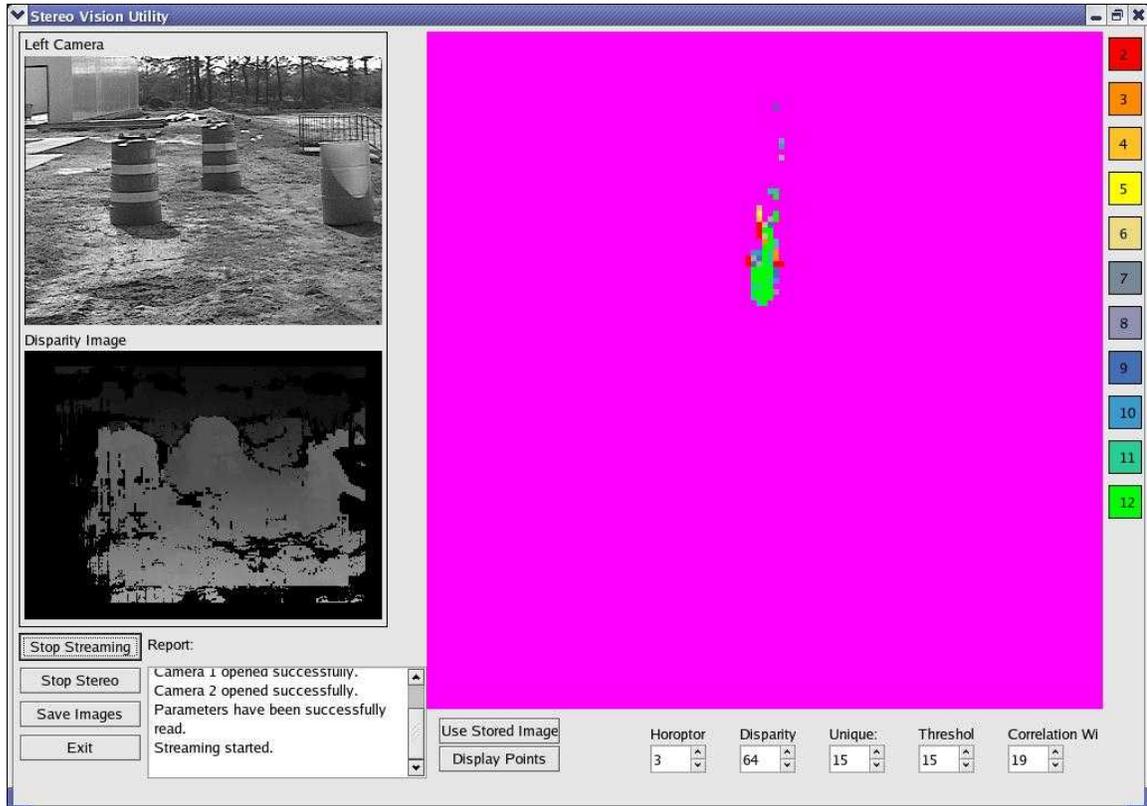
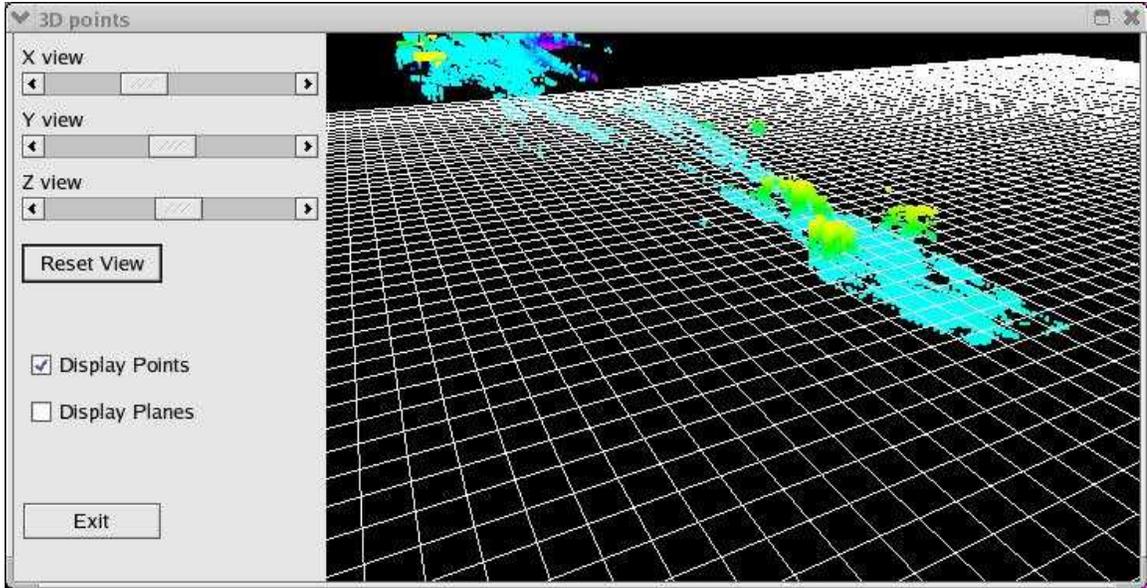


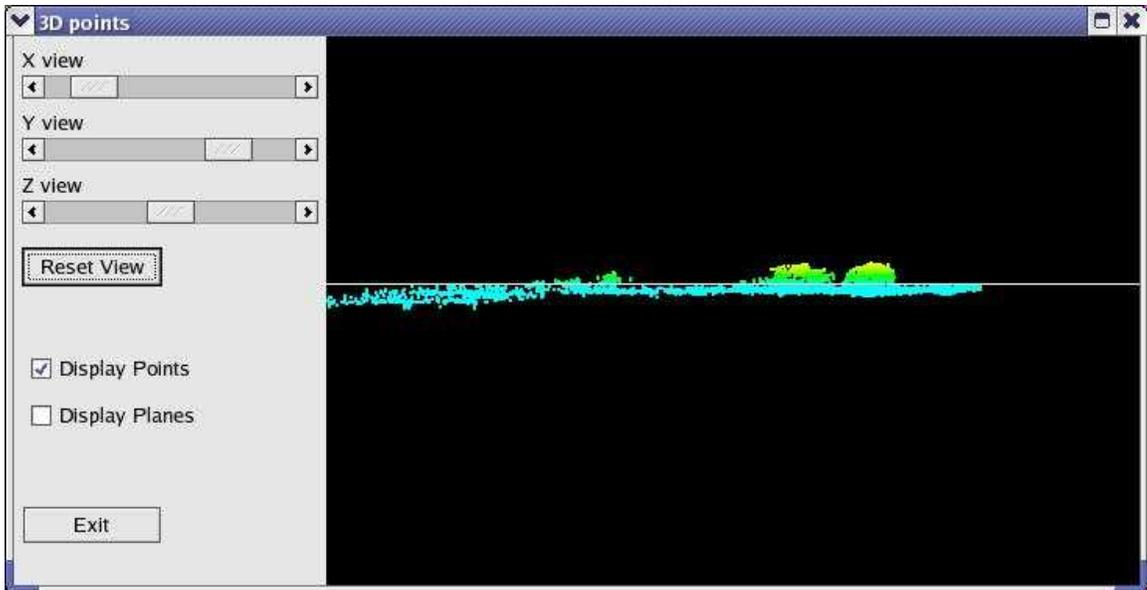
Figure 5-9. Stereo Vision Utility.

While stereo processing is turned on, the stereo parameters can be changed by using the spin boxes across the bottom right of the window.

The user can view the 3D points and the best fitting planes by clicking the “Display 3D” button. This button opens a window that uses OpenGL. The user has the option to display the 3D points and the 3D planes. Figure 5-10 shows the window displaying the 3D points. The color of each point indicates the height of the point. They range across the visible light spectrum from violet being the lowest points to red being the highest points. The side view in Figure 5-10B shows the points sloping downward from the



A



B

Figure 5-10. OpenGL windows showing the 3D point clouds. A) Top view. B) Side view.

vehicle ground plane. That is the slope of the actual ground.

Figure 5-11 shows the window displaying the best fitting planes. The colors are the same as the ones in the traversability grid display (Figure 5-9) and indicate the traversability value that each cell is assigned.

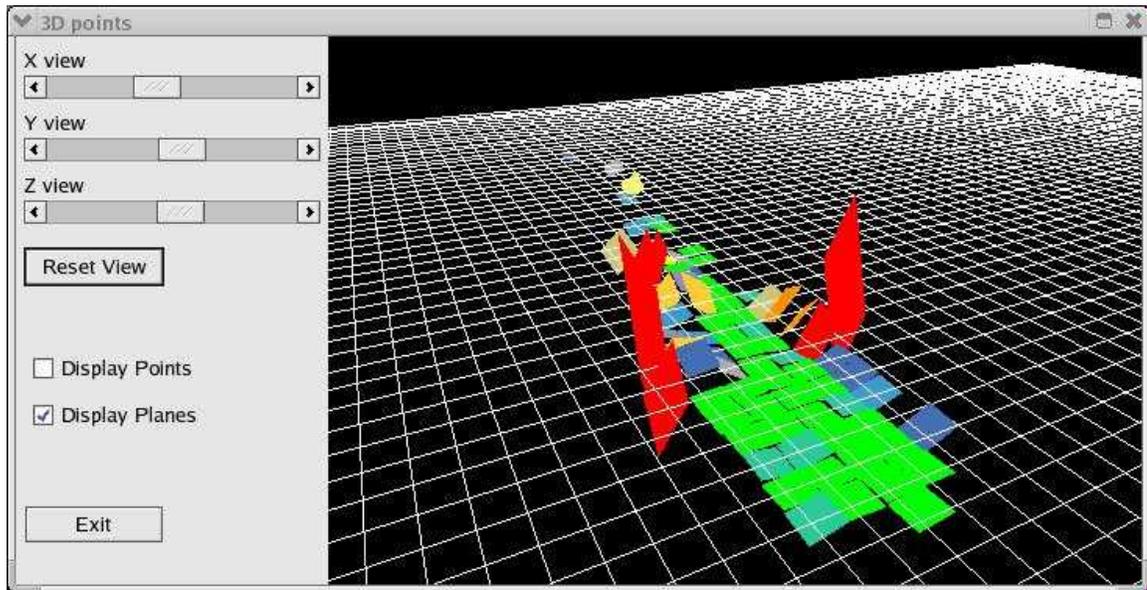


Figure 5-11. OpenGL window displaying the best fitting planes.

CHAPTER 6 TESTING AND RESULTS

For testing, several sets of images were taken with the cameras in different positions on the vehicle, different lenses and different lighting conditions. Tests were performed statically on these images and the results were compared to find the best combination of the stereo processing parameters described in Chapter 5. For most conditions, there were combinations of parameters that performed very well and produced very accurate traversability grids. However, those same parameter values returned very poor results under different conditions. The challenge was to find the set of parameter values that performed as well as possible in most conditions.

6.1 Subsample Method

Different images were processed with the four different subsample methods: single pixel selection, average of pixels, minimum pixel, and maximum pixel. Table 6-1 shows the values of the parameters for this test.

Table 6-1. Parameter values for subsample method test

Parameter:	Set to:
Subsample Method	Variable
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	64
Horopter Offset	0
Correlation Window Size	17
Confidence Threshold Value	17
Uniqueness Threshold	14

Table 6-2 shows the number of pixels correlated for eight images and each method. The highest number of pixels for each pair is in bold. The original images can be seen in Appendix B.

Table 6-2. Number of pixels correlated for each subsample method

Image Pair	Single Pixel	Average	Minimum	Maximum
1	35,654	35,407	34,943	35,107
2	24,981	23,753	26,839	22,168
3	19,710	19,803	20,468	18,856
4	24,728	24,239	25,499	23,799
5	24,112	23,651	24,737	22,634
6	21,146	20,335	21,997	19,602
7	45,864	45,484	45,049	47,246
8	47,071	46,673	45,595	48,658

In most cases the minimum value subsampling performed slightly better. Note that images 1 through 6 were taken in sunny, open conditions and images 7 and 8 were taken in shady, sun-dappled conditions. The minimum value method seemed to work best in sunny conditions whereas the maximum value method worked best in shady conditions. Since this project is geared towards the Navigator, which is designed to perform in the desert, the minimum value method was selected as the optimal method. Comparisons of update rates showed that the subsampling method had no effect on the speed of the system.

6.2 Image Resolution

For the image resolution test, images were processed at resolutions of 640×480, 320×240 and 160×120. Table 6-3 shows the parameter values used during this test.

Processing was done with multiscale disparity turned on, so the disparity images are a combination of values found from processing an image of the specified resolution and one of half that size. For example, if the resolution is set to 160×120, the disparity

results are a combination of results from processing the 160×120 images and 80×60 images.

Table 6-3. Parameter values for image resolution test

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	Variable
Multiscale Disparity	On
Pixel Search Range	64
Horopter Offset	0
Correlation Window Size	17
Confidence Threshold Value	17
Uniqueness Threshold	14

The threshold for the minimum number of points in a cell for calculating the cell's traversability was lower for the lower resolution images. Since each reduction creates an image with $\frac{1}{4}$ the number of pixels, the threshold was $\frac{1}{4}$ of the original threshold.

The resulting disparity images were compared for noise and the traversability grids were compared for the effects of that noise. Some of the disparity images and traversability grids can be seen in the in Appendix B.

The 640×480 disparity images had far too much noise. Many false obstacles were calculated in the traversability grid as a result. The 320×240 disparity images were generally clean with very little noise. The resulting traversability grids show better results. The 160×120 images did not produce sufficient disparity information for accurate traversability results.

6.3 Multiscale Disparity

Images were processed with and without multiscale disparity. The stereo parameters that were held constant were set to the values in Table 6-4.

Table 6-4. Parameter values for multiscale disparity test

Parameter:	Set to:
Subsample Method	Single Pixel Selection
Image Resolution	320×240
Multiscale Disparity	Variable
Pixel Search Range	64
Horopter Offset	0
Correlation Window Size	17
Confidence Threshold Value	17
Uniqueness Threshold	14

Some of the images can be seen in Appendix B with two versions of their disparity image. One was calculated with multiscale processing, the other without multiscale processing. In the disparity image, the lighter pixels represent points with higher disparity. Black areas are areas where the disparity could not be calculated. Table 6-5 shows the number of pixels correlated for several images with and without multiscale processing.

Table 6-5. Number of pixels correlated with and without multiscale processing.

Image Pair	Without Multiscale Processing	With Multiscale Processing
1	24,442	35,654
2	13,096	23,780
3	17,393	28,081
4	16,342	24,981
5	16,848	25,765
6	15,125	25,813

It is clear that multiscale processing adds a great deal of disparity data. The average update rate for both methods was 14.65 Hz. Adding multiscale processing does not impact the system's speed; however, the results are significantly better. Therefore, multiscale processing should be left on.

6.4 Pixel Search Range

When multiscale processing is turned on, the search ranges of 32 and 64 are the only ones that return valid results. Images were tested with both of these pixel search ranges. The parameter values are shown in Table 6-6.

Table 6-6. Parameter values for pixel search range test

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	Variable
Horopter Offset	0
Correlation Window Size	17
Confidence Threshold Value	17
Uniqueness Threshold	14

The disparity images and the traversability grids were compared for range. The update rates were also compared. Some of the disparity images and traversability grids can be seen in the in Appendix B.

From the disparity images, it can be seen that objects and ground in the foreground are only detectable with the search range of 64. The average update rate with the 32 pixel range was 17.43 Hz. The average rate with the 64 pixel range was 14.65. The traversability grids show processing must be done with a 64 pixel range for meaningful results.

6.5 Horopter Offset

Several images were tested over the range of possible horopter offset values with the parameters shown in Table 6-7.

The acceptable horopter offset values were recorded for each image. An acceptable value is one that produced an accurate traversability grid. They varied greatly from

Table 6-7. Parameters for horoptor offset test

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	64
Horoptor Offset	Variable
Correlation Window Size	17
Confidence Threshold Value	17
Uniqueness Threshold	14

image to image, but 3 seemed to be acceptable for most images. Therefore, 3 was selected as the optimal value for the horoptor offset. A chart of acceptable horoptor values for a series of images is shown in Figure 6.1.

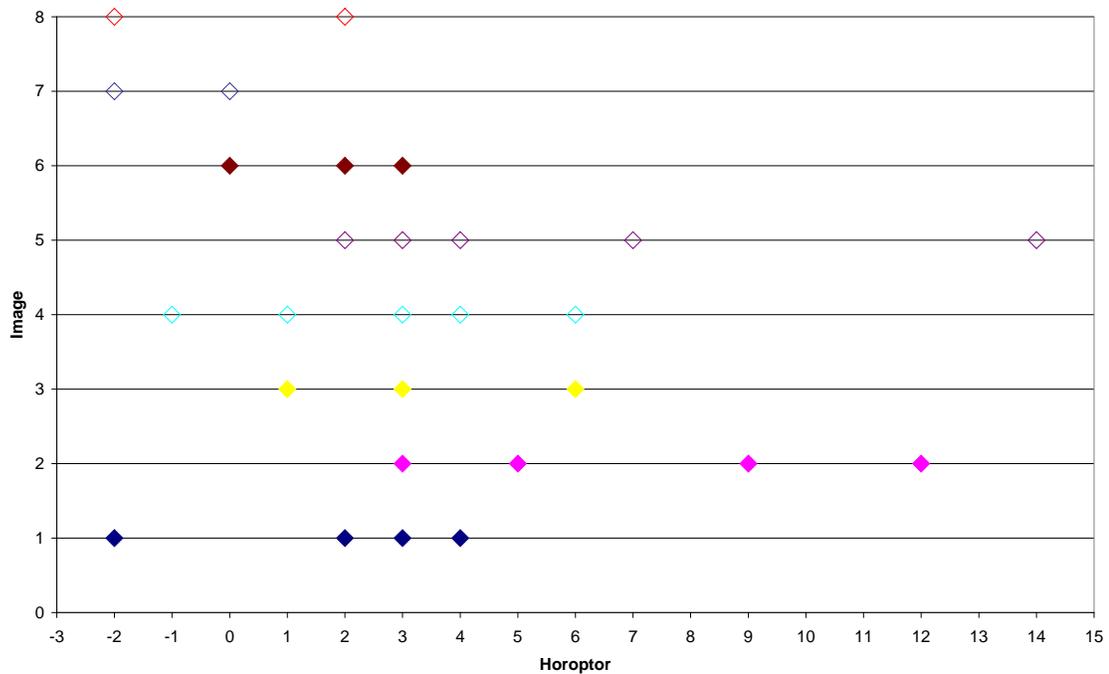


Figure 6-1. Acceptable horoptor values for a series of images are indicated by the marks on the chart.

6.6 Correlation Window Size

Several images were tested over the range of possible correlation window size values with the parameters shown in Table 6-8. Possible values range from 5 to 21 in increments of 2.

Table 6-8. Parameters for correlation window size test

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	64
Horopter Offset	3
Correlation Window Size	Variable
Confidence Threshold Value	17
Uniqueness Threshold	14

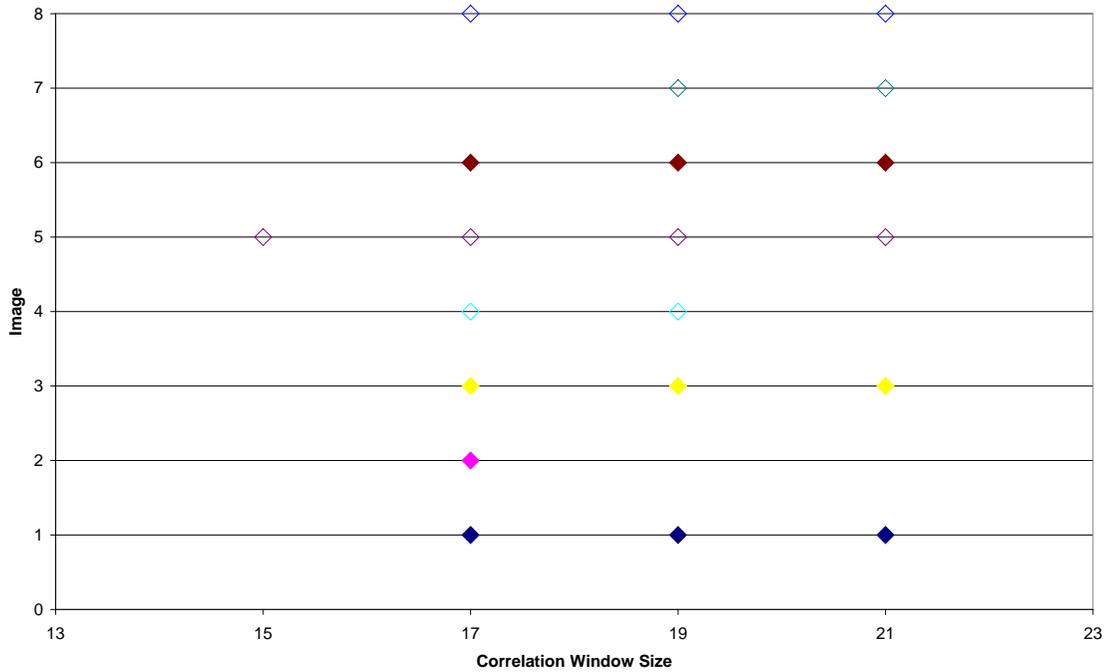


Figure 6-2. Acceptable correlation window size values for a series of images are indicated by the marks on the chart.

The values that returned acceptable traversability grids were recorded for each image. The acceptable values were fairly consistently in the range of 17 to 21. A chart of acceptable correlation window size values for a series of images is shown in Figure 6.2.

Since 19 was the average acceptable value, it was chosen for the optimal correlation window size.

6.7 Confidence Threshold Value

Several images were tested over the range of possible confidence threshold values with the parameters shown in Table 6-9. Possible values range from 0 to 40.

Table 6-9. Parameters for confidence threshold value test

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	64
Horopter Offset	3
Correlation Window Size	19
Confidence Threshold Value	Variable
Uniqueness Threshold	14

The values that returned acceptable traversability grids were recorded for each image. A chart of acceptable correlation window size values for a series of images is shown in Figure 6.3.

In all cases values over 25 contained too little data to compute a meaningful traversability grid. The upper limit for most of the images was 15, so this was chosen as the optimal confidence threshold value. This value was low enough to calculate sufficient data for the traversability grid and high enough to eliminate most noise.

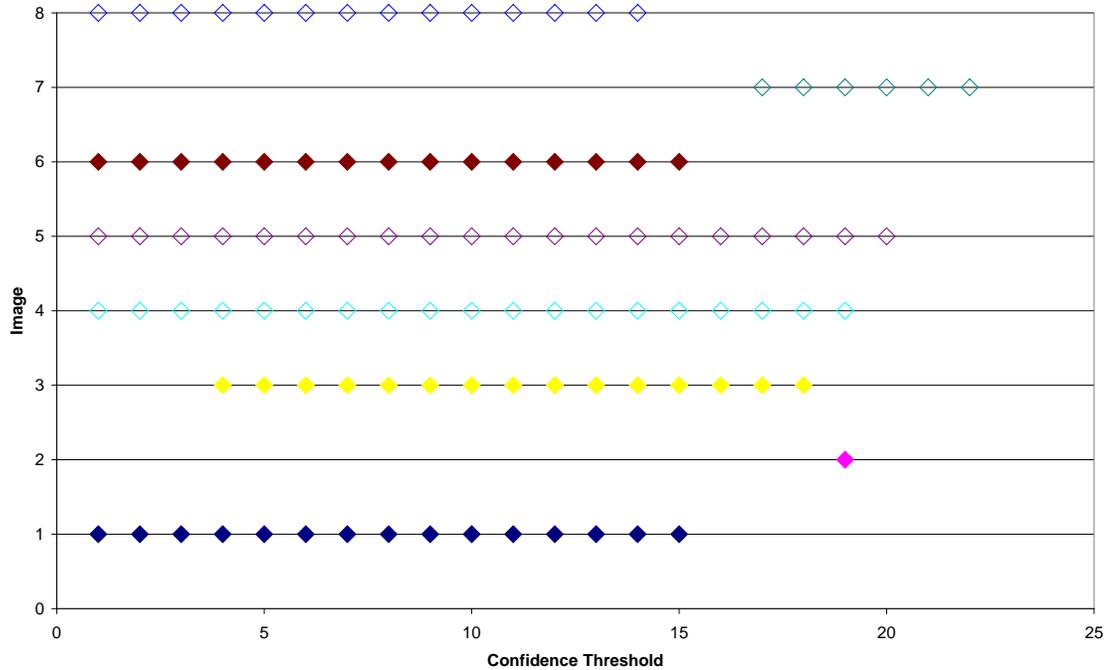


Figure 6-3. Acceptable confidence threshold values for a series of images are indicated by the marks on the chart.

6.8 Uniqueness Threshold Value

Several images were tested over the range of uniqueness threshold values with the parameters shown in Table 6-10. Possible values range from 0 to 40.

Table 6-10. Parameters for uniqueness threshold value test

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	64
Horopter Offset	3
Correlation Window Size	19
Confidence Threshold Value	15
Uniqueness Threshold	Variable

The values that returned acceptable traversability grids were recorded for each image. A chart of acceptable correlation window size values for a series of images is shown in Figure 6.4.

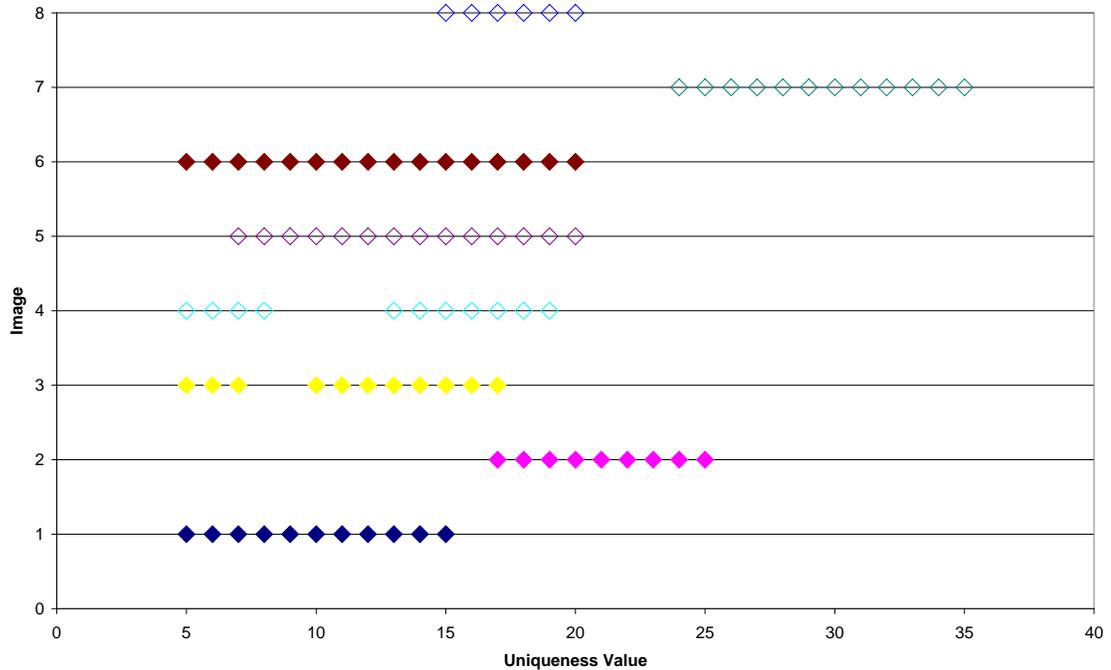


Figure 6-4. Acceptable uniqueness threshold values for a series of images are indicated by the marks on the chart.

A value of 15 was chosen for the optimal uniqueness threshold value. For most cases, this value was low enough to calculate sufficient data for the traversability grids and high enough to eliminate most noise.

6.9 Final Parameters Selected

The tests described above give the best possible overall results for different lighting conditions and image texture. The final parameters selected are listed in Table 6-11.

Disparity images and traversability grid results calculated using these parameters can be seen in Appendix C.

Table 6-11. Final Stereo Processing Parameters

Parameter:	Set to:
Subsample Method	Minimum Value
Image Resolution	320×240
Multiscale Disparity	On
Pixel Search Range	64
Horopter Offset	3
Correlation Window Size	19
Confidence Threshold Value	15
Uniqueness Threshold	15

6.9 Range

The 12mm focal length lens is able to detect objects farther away than the 6mm focal length lens, but it cannot detect objects that are close to the vehicle. It also has a much narrower field of view. The traversability grids with the 12mm focal length lens contain data in about half the area of the traversability grids from the 6mm focal length lens. Because of the limited space for the stereo vision cameras on the NaviGator sensor cage, the different camera configurations tested did not result in a significant impact on the grid range. Some recommendations for increasing the range of the stereo vision system will be discussed in Chapter 7.

6.10 Auto-Iris

To demonstrate the benefit of having an auto-iris rather than a manual iris, the auto-iris function was turned off and images were taken. Figure 6.5 and Figure 6.6 show the same scene with and without the auto-iris function.

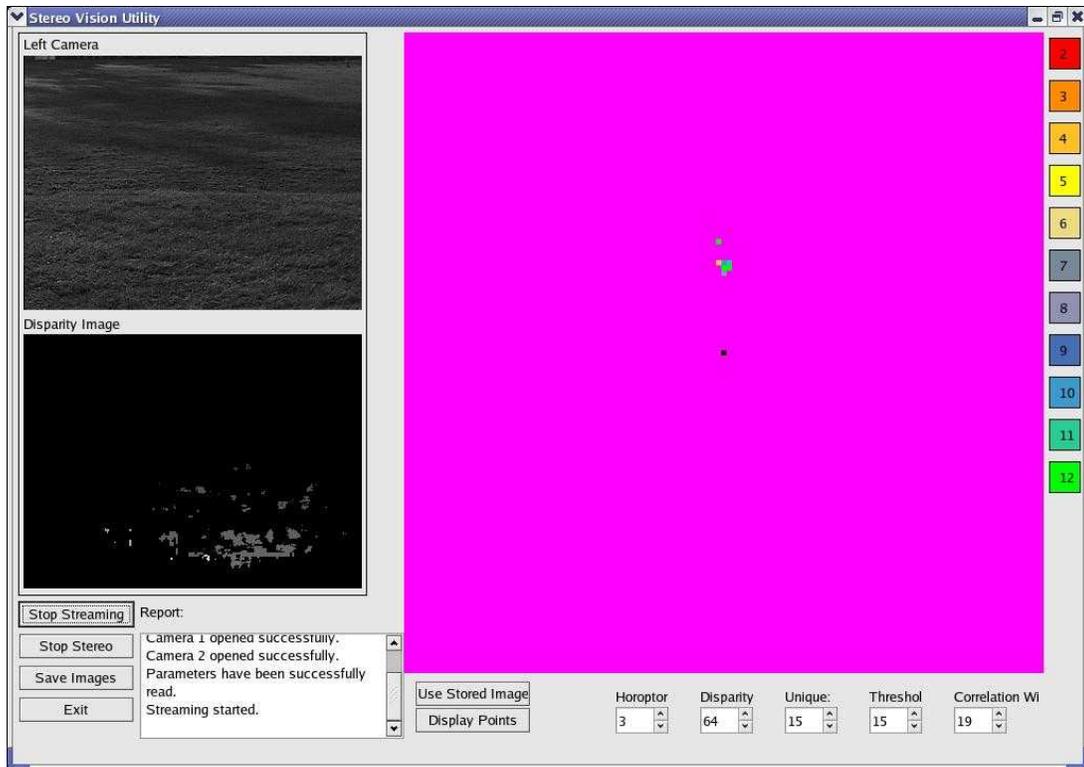


Figure 6-5. Scene without auto-iris function.

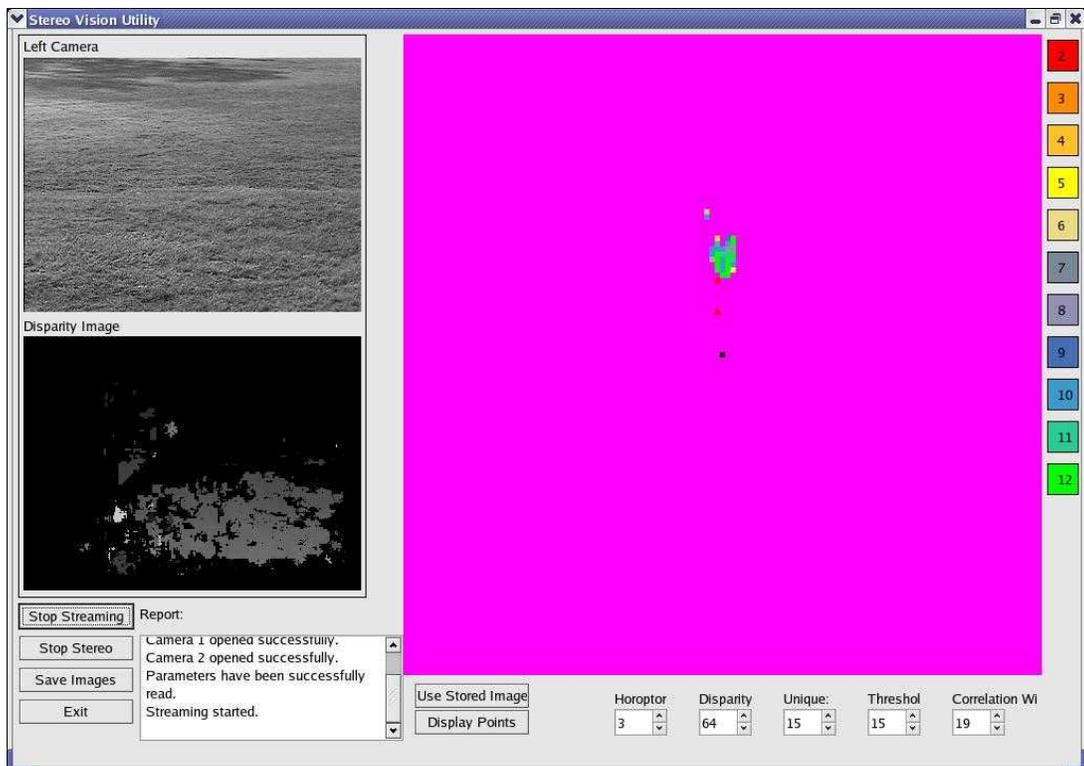


Figure 6-6. Scene with auto-iris function

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

7.1 Conclusions

This work focused on selecting the hardware and developing software for outputting CIMAR Smart Sensor traversability grids using stereo vision.

The first step was to select the hardware. Although the stereo vision system was not used in the DARPA Grand Challenge, a monocular vision system was used for path finding. The monocular vision system used the same hardware, which proved capable of the task.

The next step was to develop the software for computing traversability grids. The previous CIMAR stereo vision researcher used the manual iris Videre stereo cameras and found that slight changes in lighting greatly degraded the system's results making it completely unusable. This system is capable of delivering traversability grids with a moderate level of accuracy in different lighting conditions, though at times the disparity data does contain enough noise to create false obstacles. Also, the range and field of view are quite limited. In order to use this system successfully on an autonomous vehicle, future work must deal with these issues.

This work provides a hardware setup, an algorithm for computing traversability grids and an optimal set of stereo processing parameters. This is a starting point for a more robust stereo vision system to be developed at CIMAR.

7.2 Recommendations for Future Work

Future work should attempt to increase the field of view and range of the system. The simplest way to do this would be to add more cameras. These cameras should be positioned in such a way as to capture data from different regions around the vehicle. Using multiple pairs of cameras with different focal lengths and different baselines would increase the range. It is recommended that 6mm focal length lenses be used with something higher than 12mm focal length lenses. The difference in range between these two lenses was not high enough to make a significant impact. A wider baseline would increase the range but may not be possible with the current NaviGator sensor cage.

Future algorithm improvements should investigate the possibility of comparing the slope of each grid cell to the slopes of its surrounding grid cells rather than the vehicle ground plane. This will help to eliminate traversable hills from being classified as non-traversable.

Another recommendation for future work is that the problem be limited by searching for known objects before computing stereo data. This recommendation is particularly geared towards work that will take place for the DARPA Urban Challenge in 2007, which will require vehicles to obey traffic laws. An Urban Challenge version of the stereo vision system could use pattern recognition methods to first detect lanes and street signs. Then the correlation could be performed on only those pixels containing the objects of interest. A set of stereo parameters might be found that has great success in correlating the pixels of those objects. The success of correlating unknown object pixels would no longer matter. This has the potential of greatly increasing the speed and accuracy of the results and to provide important classification information that other range finders (i.e. lasers, radars) cannot.

APPENDIX A SAMPLE CALIBRATION FILE

```
# SVS Engine v 4.0 Stereo Camera Parameter File
# top bar
# 6 mm lens

[image]
have_rect 1      # 1 if we have rectification parameters

[stereo]
frame 1.0        # frame expansion factor, 1.0 is normal

[external]
Tx -200.005429   # translation between left and right cameras
Ty -0.084373
Tz -5.470275
Rx -0.023250     # rotation between left and right cameras
Ry -0.042738
Rz 0.001794

[left camera]
pwidth 640       # number of pixels in calibration images
pheight 480
dpx 0.007000     # effective pixel spacing (mm) for this resolution
dpy 0.007000
sx 1.000000      # aspect ratio, analog cameras only
Cx 319.297412    # camera center, pixels
Cy 267.170193
f 815.513491
fy 813.086357
alpha 0.000000   # skew parameter, analog cameras only
kappa1 -0.204956 # radial distortion parameters
kappa2 -0.234074
kappa3 0.000000
tau1 0.000000    # tangential distortion parameters
tau2 0.000000
proj             # projection matrix: from left camera 3D coords to
left rectified coordinates
  8.130000e+02 0.000000e+00 3.322576e+02 0.000000e+00
  0.000000e+00 8.130000e+02 2.478598e+02 0.000000e+00
  0.000000e+00 0.000000e+00 1.000000e+00 0.000000e+00
rect             # rectification matrix for left camera
  9.998803e-01 -1.510719e-03 -1.539771e-02
  1.689463e-03 9.999313e-01 1.160210e-02
```

1.537912e-02 -1.162673e-02 9.998142e-01

```
[right camera]
pwidth 640      # number of pixels in calibration images
pheight 480
dpx 0.007000   # effective pixel spacing (mm) for this resolution
dpy 0.007000
sx 1.000000    # aspect ratio, analog cameras only
Cx 343.831089  # camera center, pixels
Cy 228.195459
f 812.178078   # focal length (pixels) in X direction
fy 807.398202  # focal length (pixels) in Y direction
alpha 0.000000 # skew parameter, analog cameras only
kappa1 -0.207084 # radial distortion parameters
kappa2 -0.042581
kappa3 0.000000
tau1 0.000000  # tangential distortion parameters
tau2 0.000000
proj           # projection matrix: from right camera 3D coords to
left rectified coordinates
  8.130000e+02 0.000000e+00 3.322576e+02 -1.626044e+05
  0.000000e+00 8.130000e+02 2.478598e+02 0.000000e+00
  0.000000e+00 0.000000e+00 1.000000e+00 0.000000e+00
rect          # rectification matrix for right camera
  9.996258e-01 4.218514e-04 2.735063e-02
 -1.041221e-04 9.999325e-01 -1.161727e-02
 -2.735369e-02 1.161008e-02 9.995584e-01

[global]
GTx 0.000000
GTy 0.000000
GTz 0.000000
GRx 0.000000
GRy 0.000000
GRz 0.000000
```

APPENDIX B
IMAGES FROM TESTING



Figure B-1. Original Images from Subsample Test

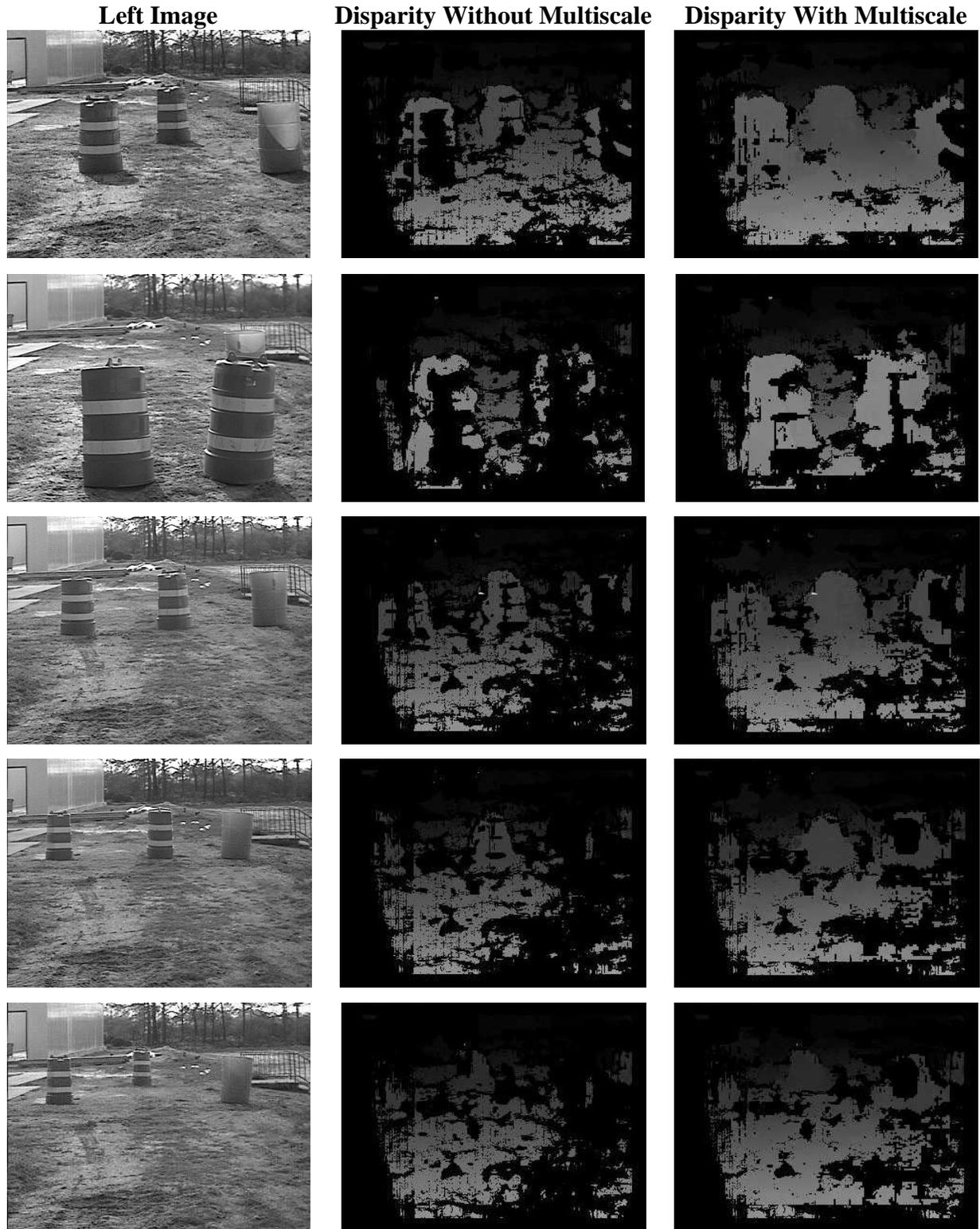


Figure B-2. Disparity image results from testing with and without multiscale disparity processing



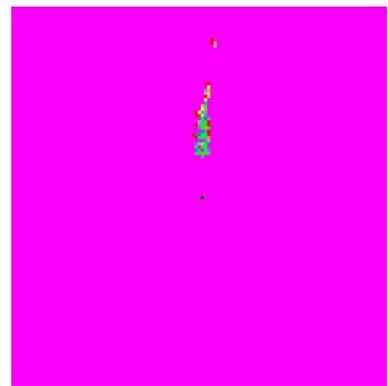
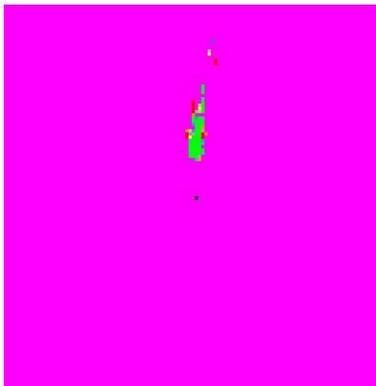
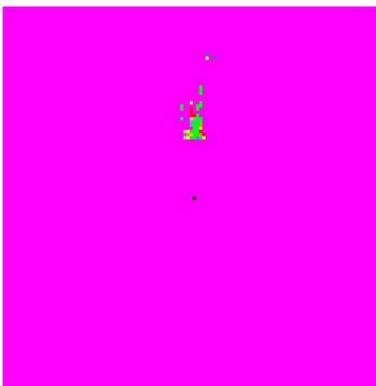
Figure B-2. Continued



640 x 480

320 x 240

160 x 120



A

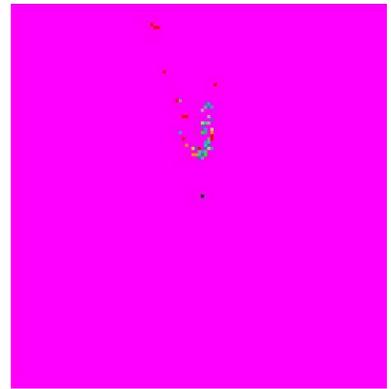
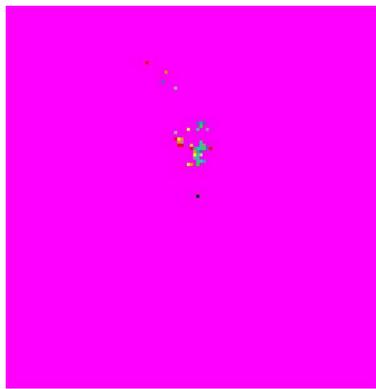
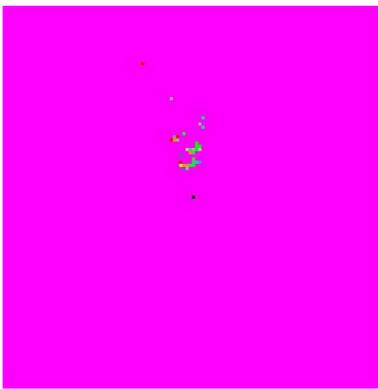
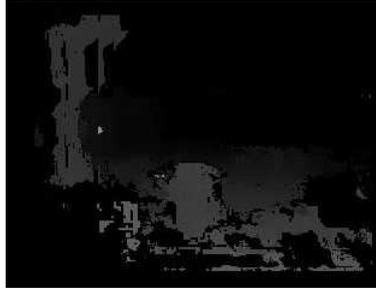
Figure B-3. Disparity image and traversability grid results from testing with different image resolutions.



640 x 480

320 x 240

160 x 120



B

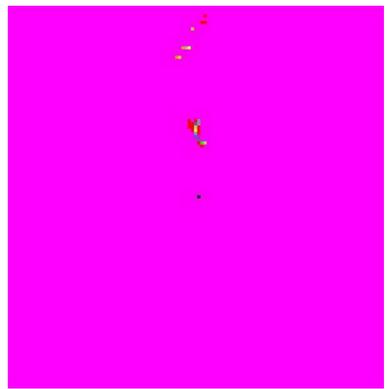
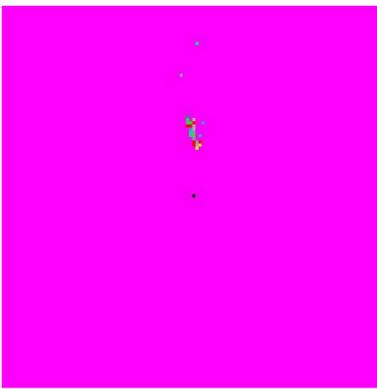
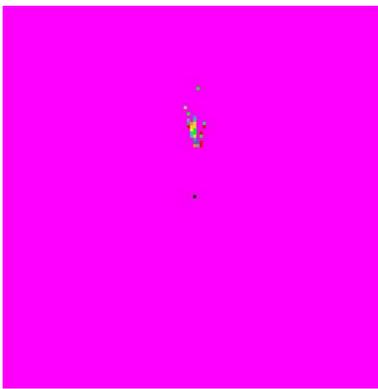
Figure B-3. Continued



640 x 480

320 x 240

160 x 120



C

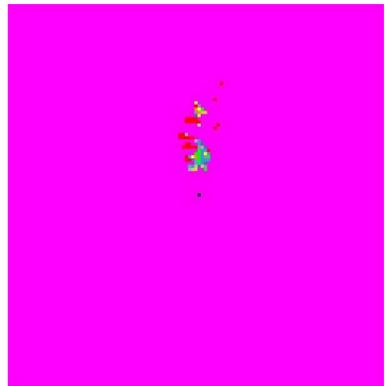
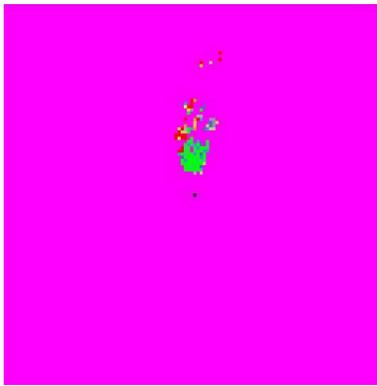
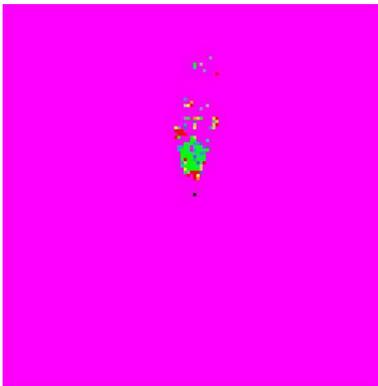
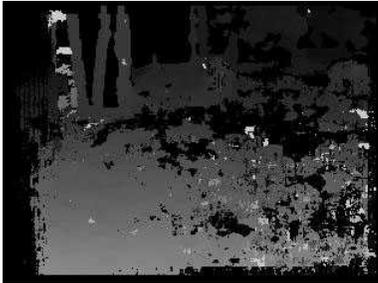
Figure B-3. Continued



640 x 480

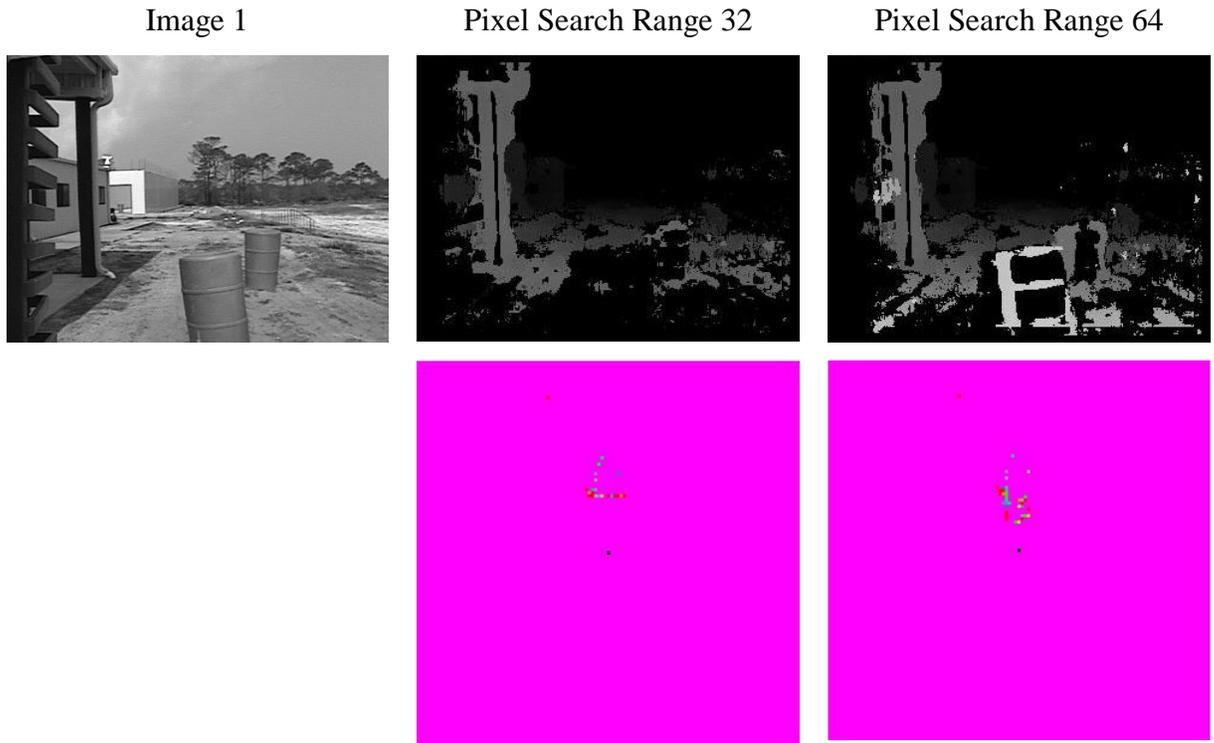
320 x 240

160 x 120



D

Figure B-3. Continued



A

Figure B-4. Disparity image and traversability grid results from testing with different pixel search ranges.

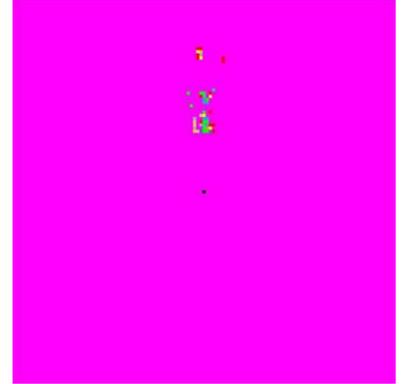
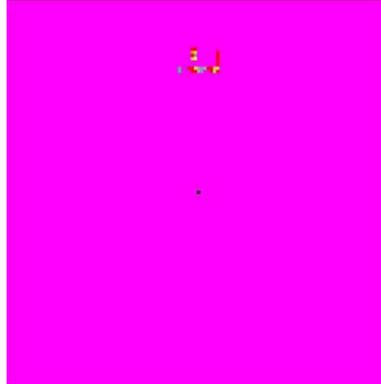
Image 2



Pixel Search Range 32



Pixel Search Range 64



B

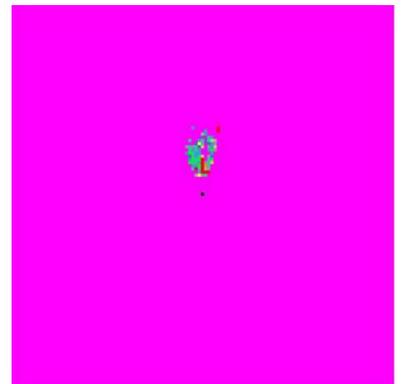
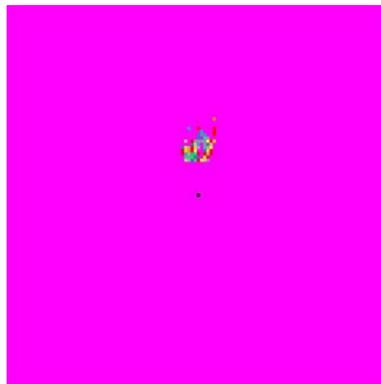
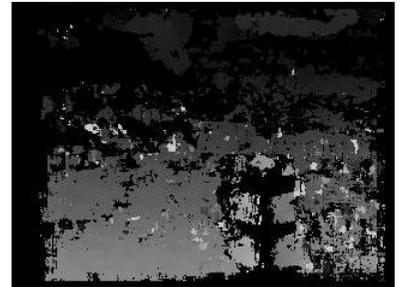
Image 3



Pixel Search Range 32

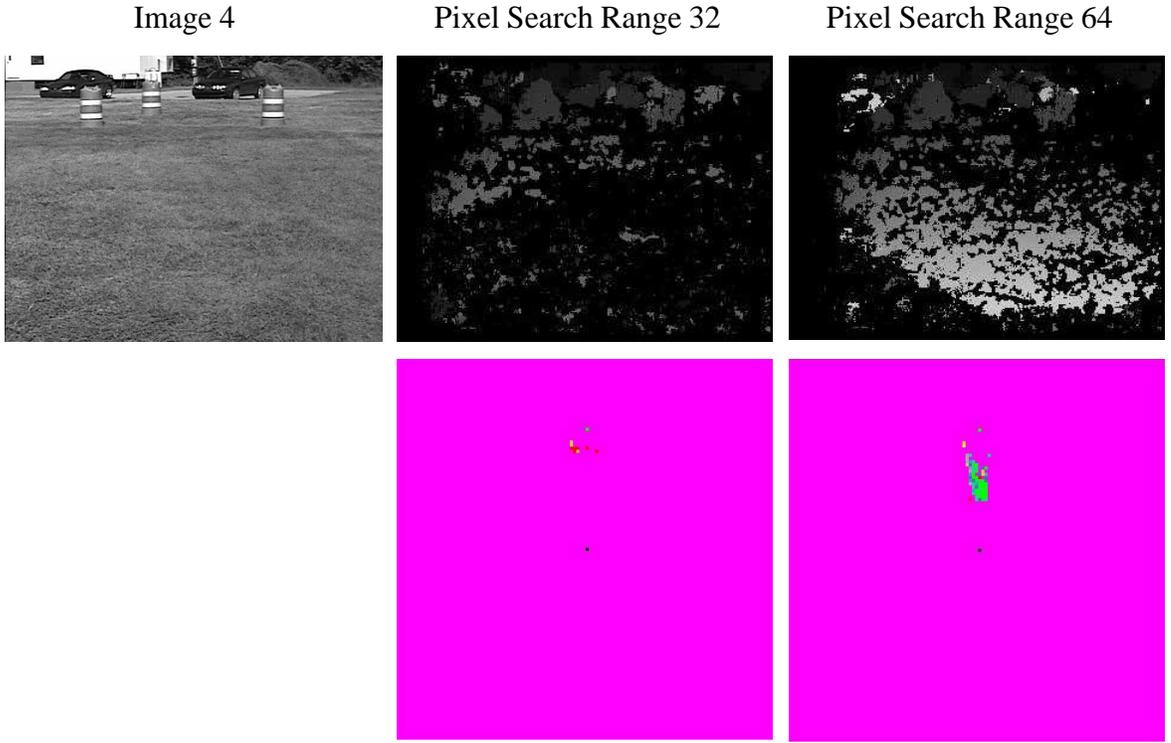


Pixel Search Range 64



C

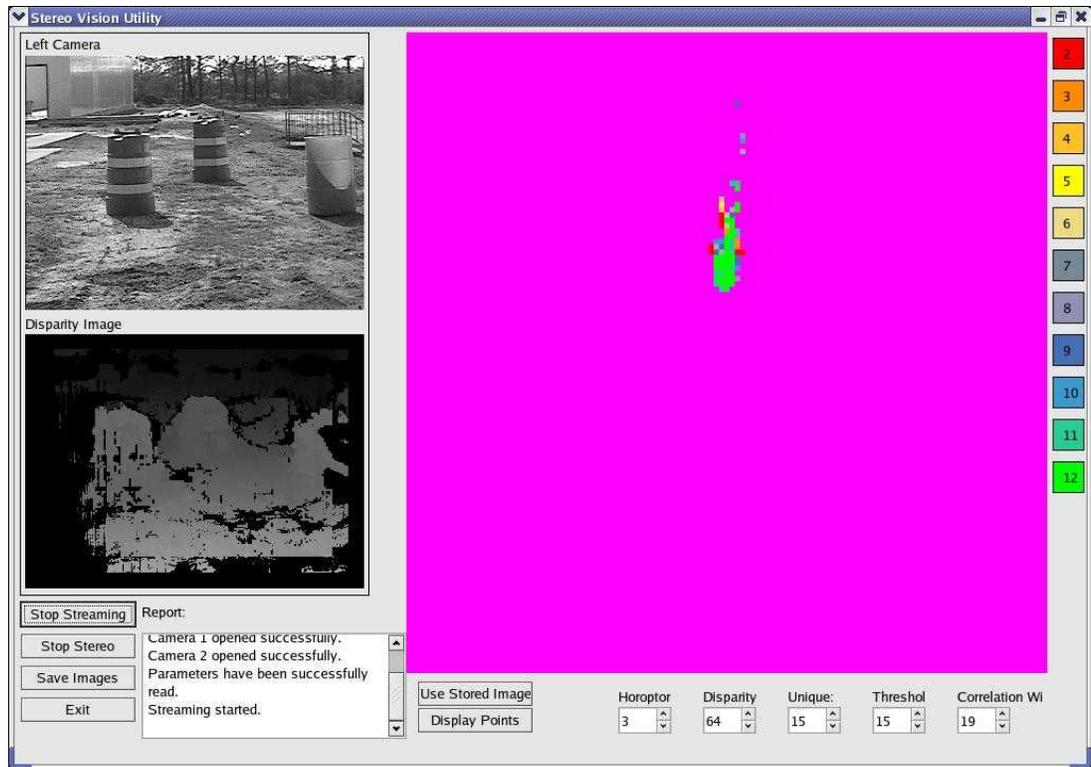
Figure B-4. Continued



D

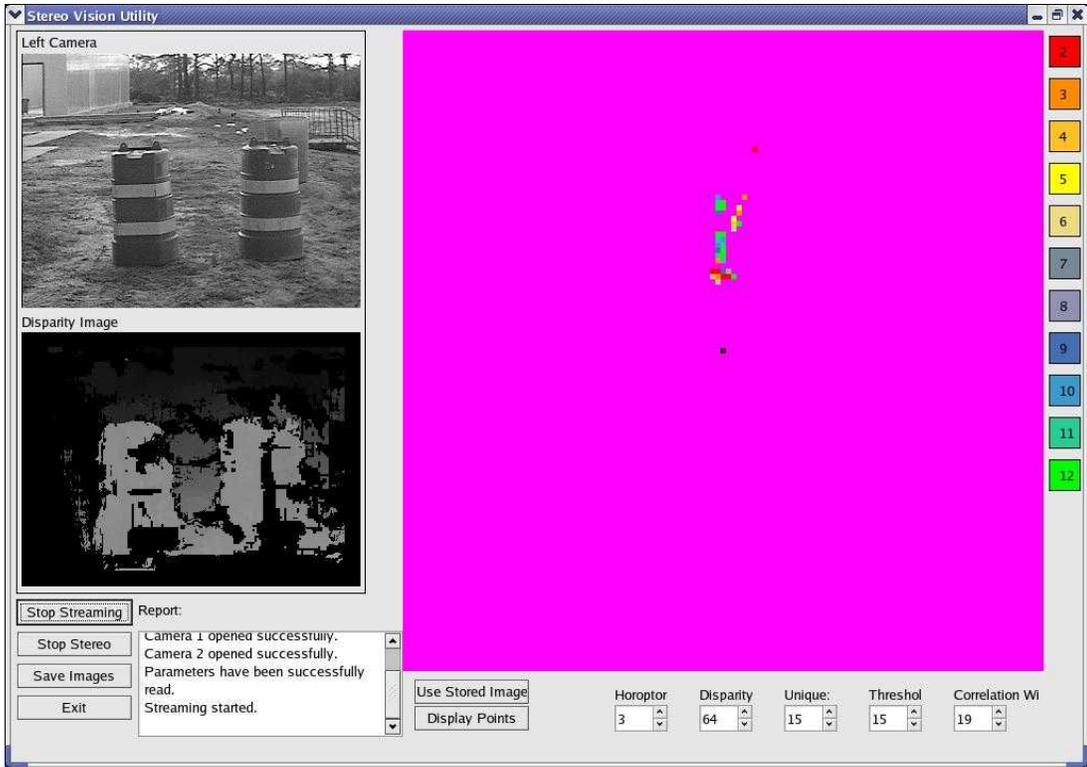
Figure B-4. Continued

APPENDIX C
RESULTS FROM FINAL SELECTED STEREO PROCESSING PARAMETERS

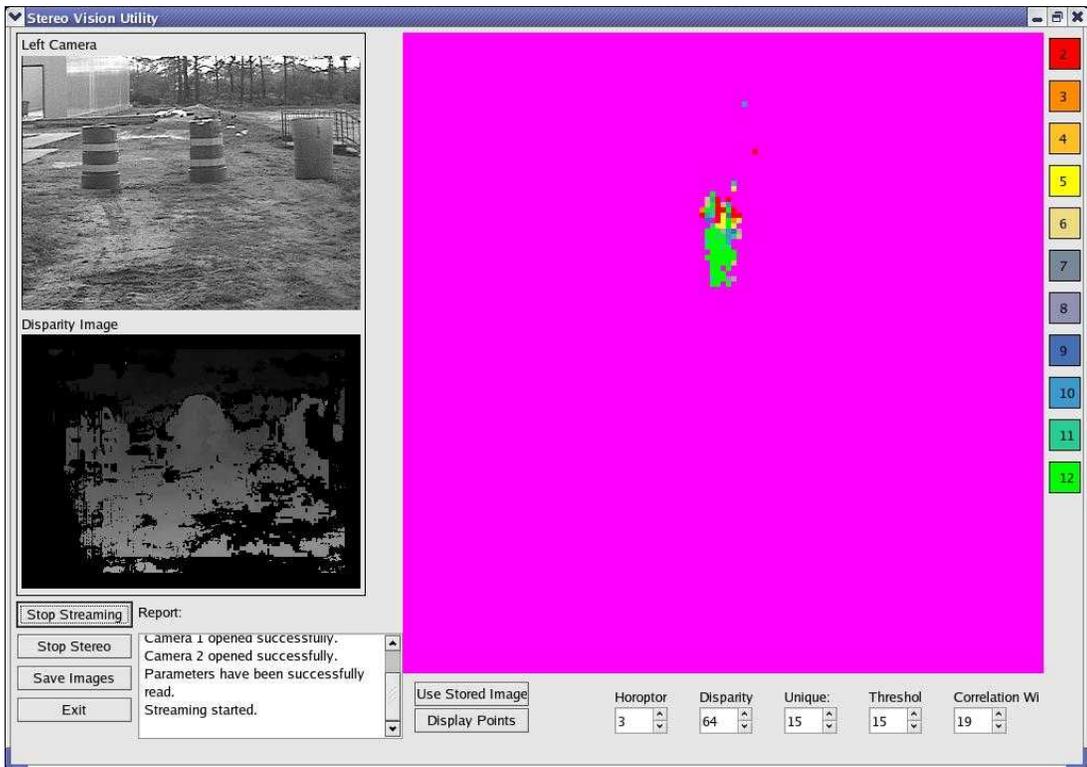


A

Figure C-1. Results from stereo processing. A through R show screenshots of the Stereo Vision Utility displaying the left original image, the disparity image, and the traversability grid calculated from the original image pair. The original images were taken of various scenes with the stereo processing parameters selected during testing.

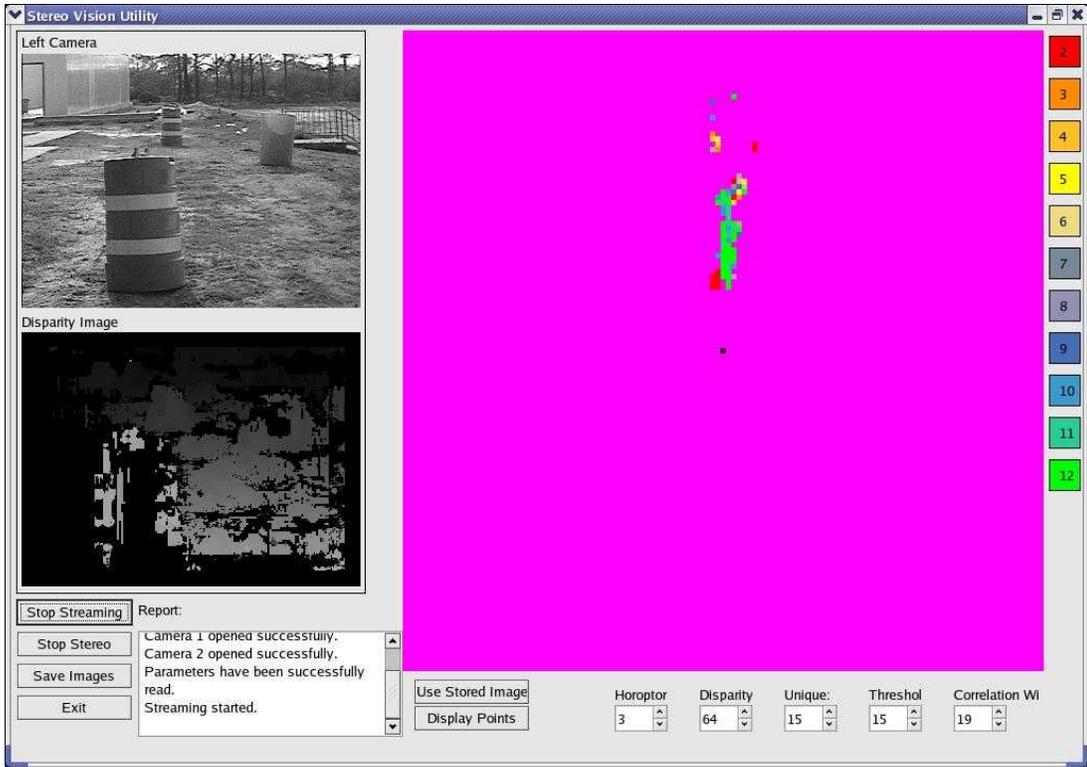


B

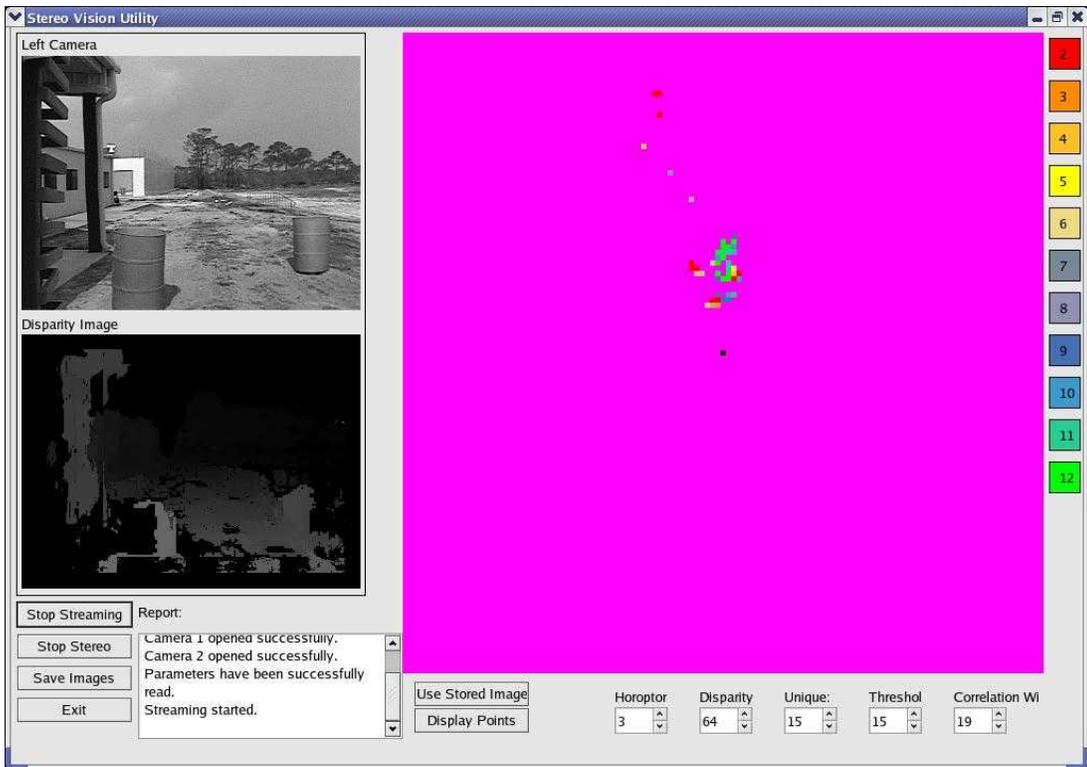


C

Figure C-1. Continued

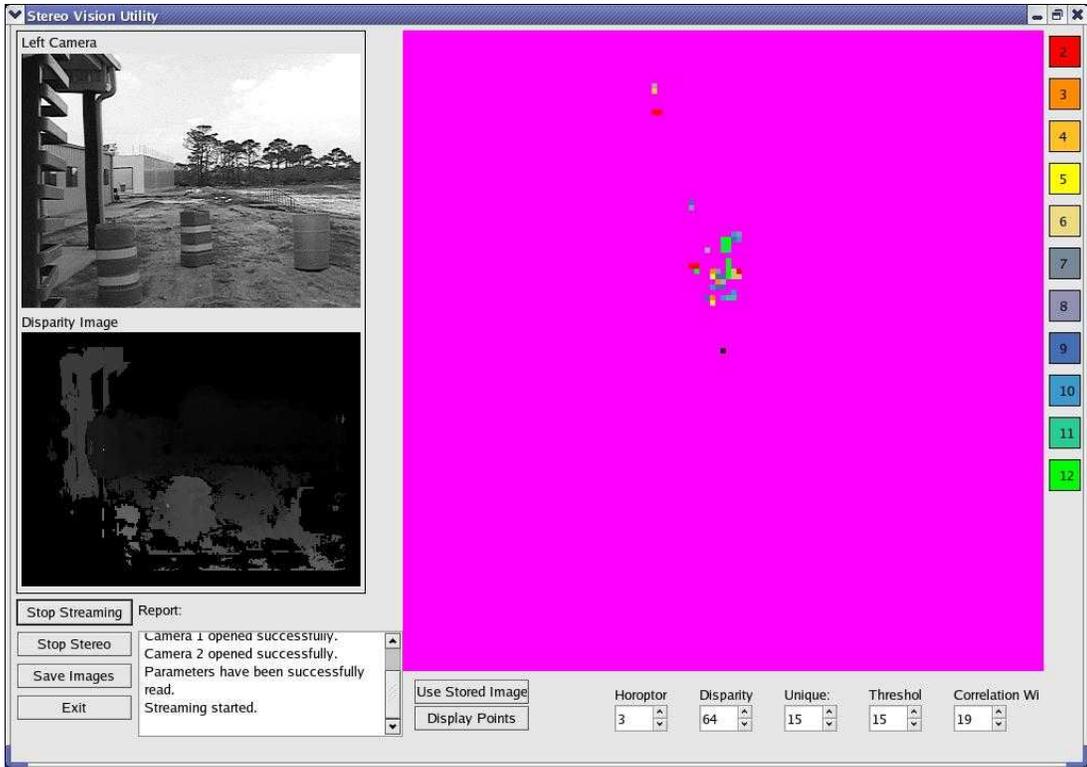


D

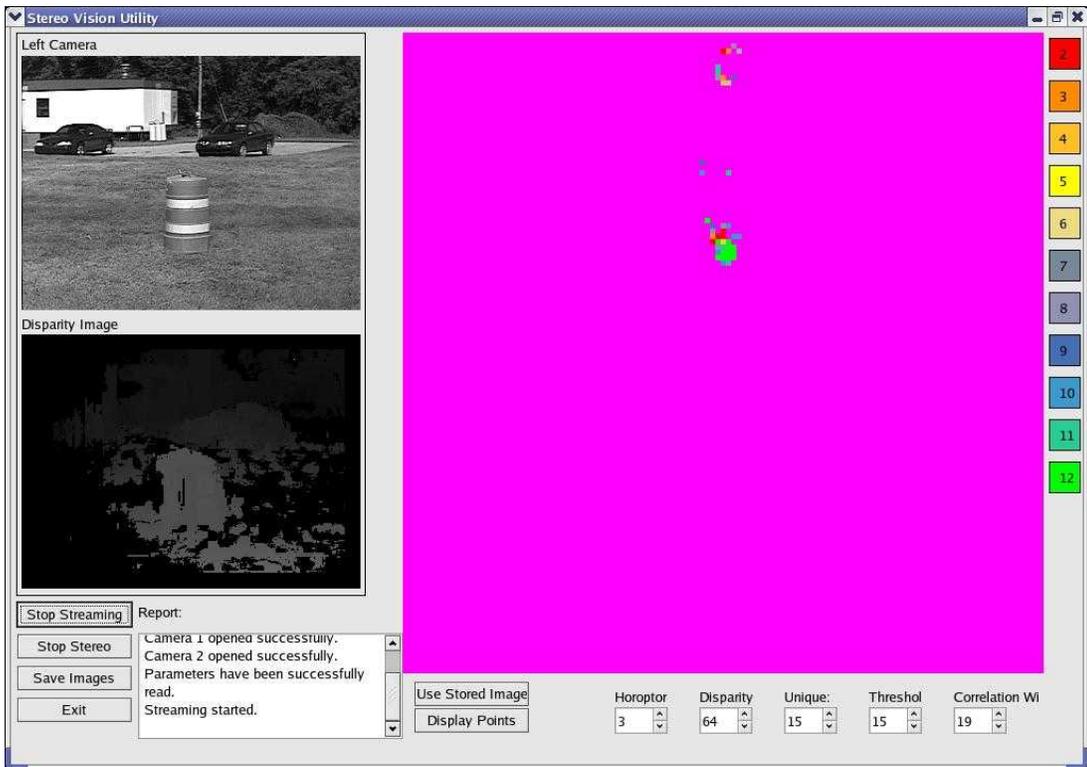


E

Figure C-1. Continued

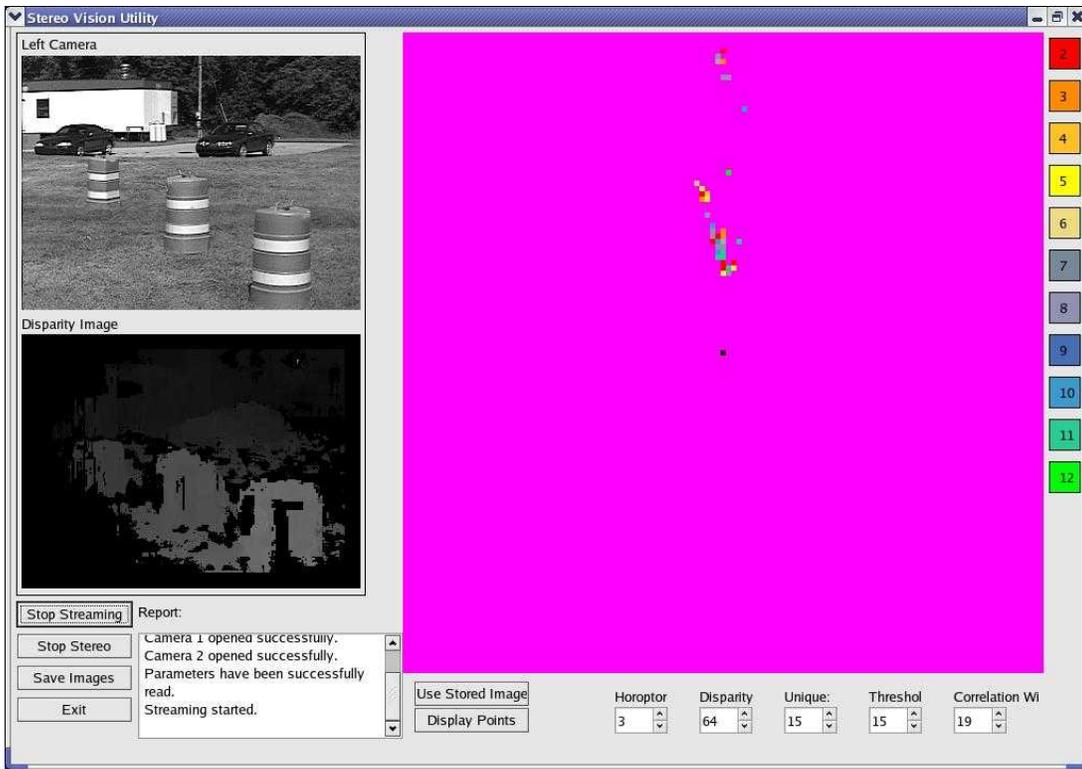


F

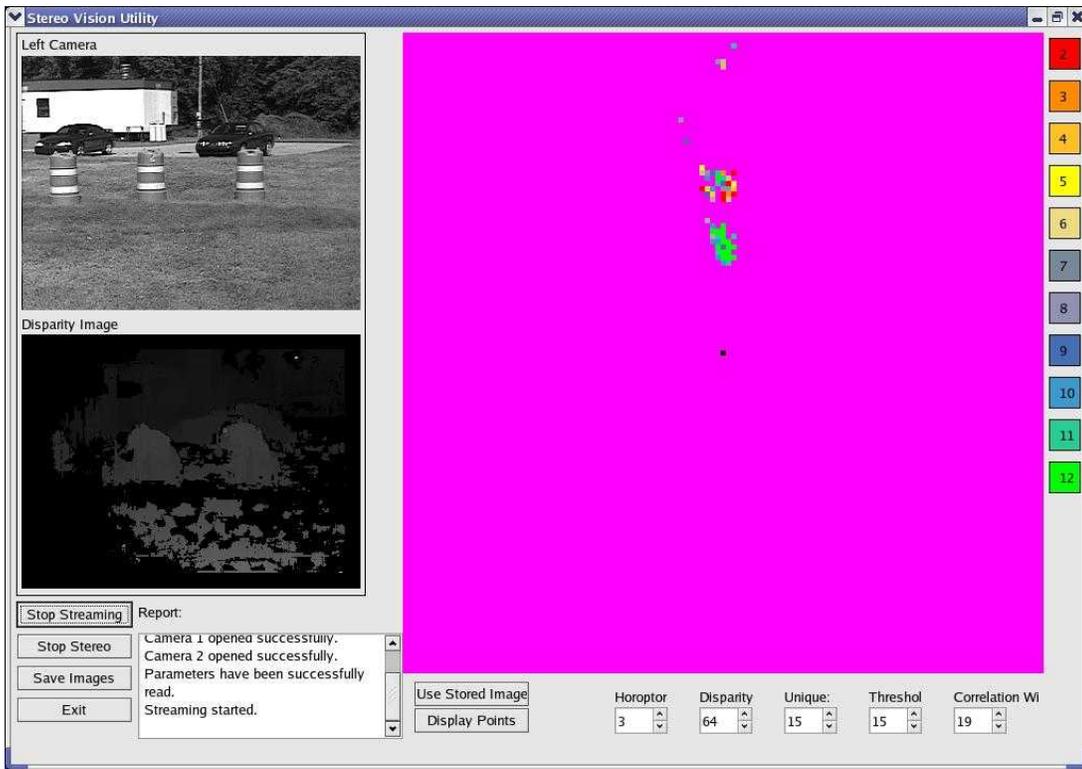


G

Figure C-1. Continued

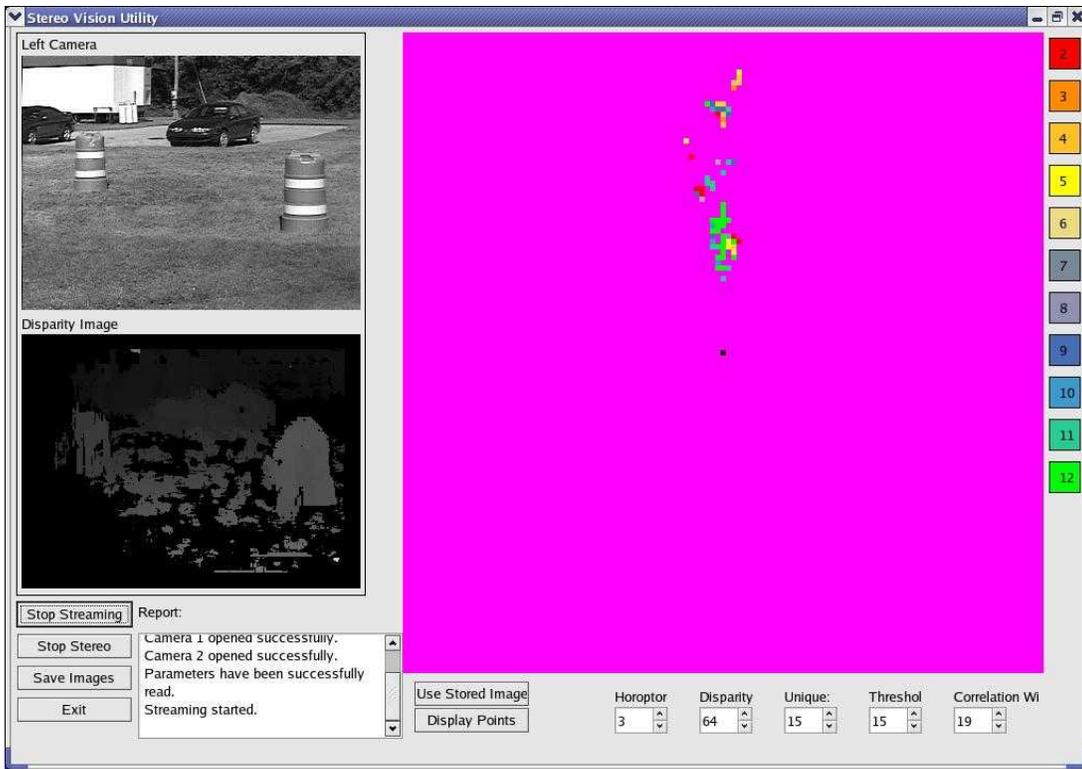


H

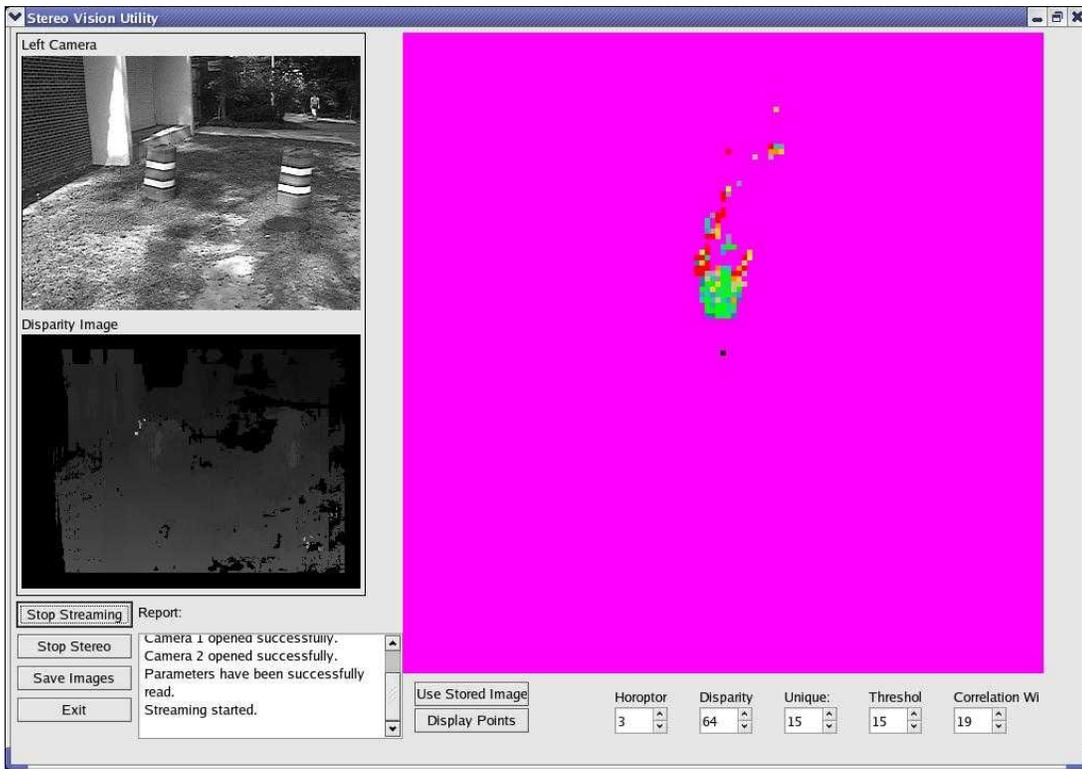


I

Figure C-1. Continued

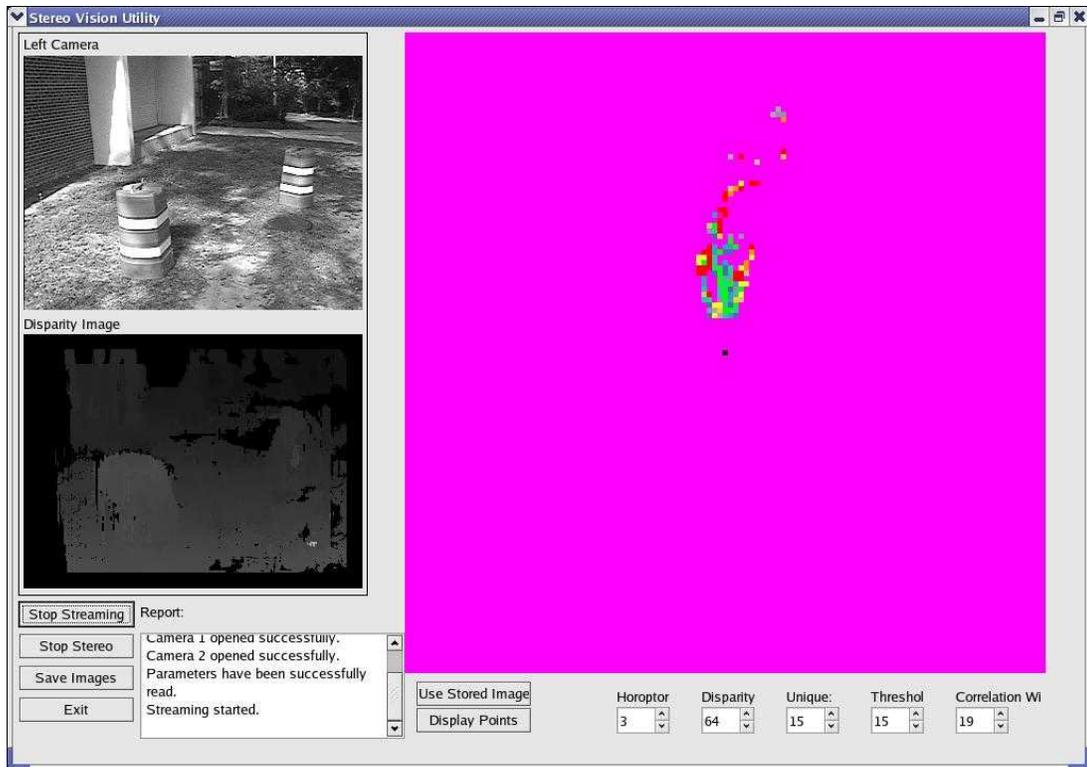


J

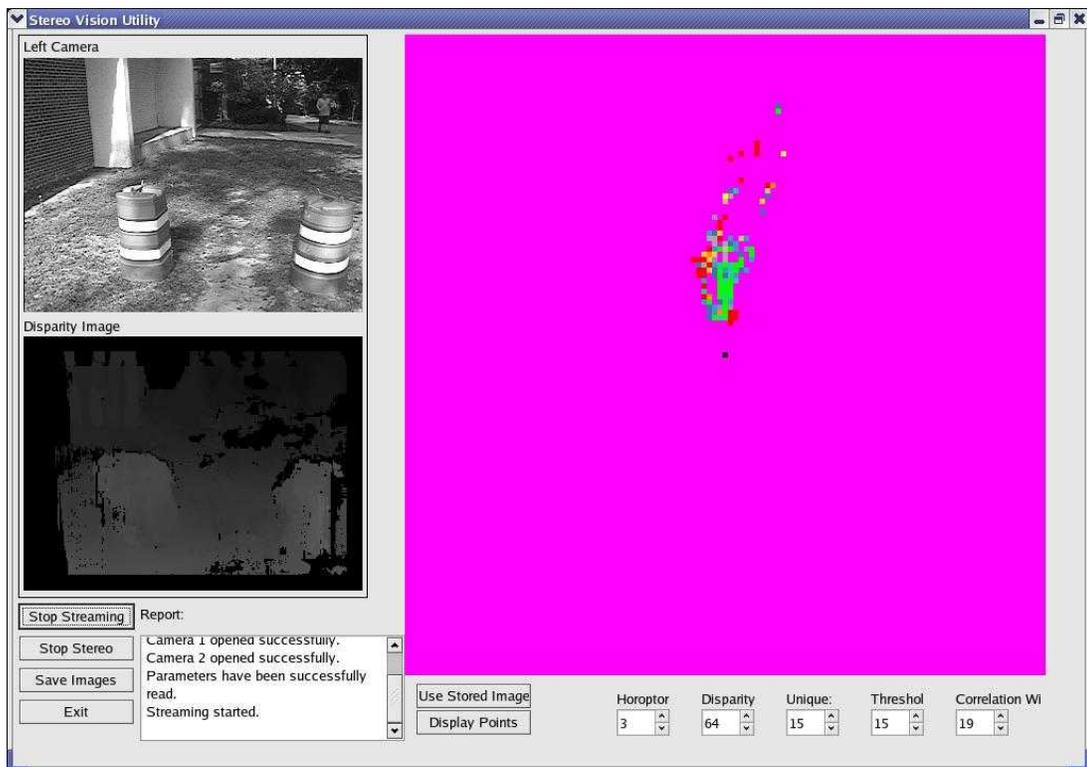


K

Figure C-1. Continued

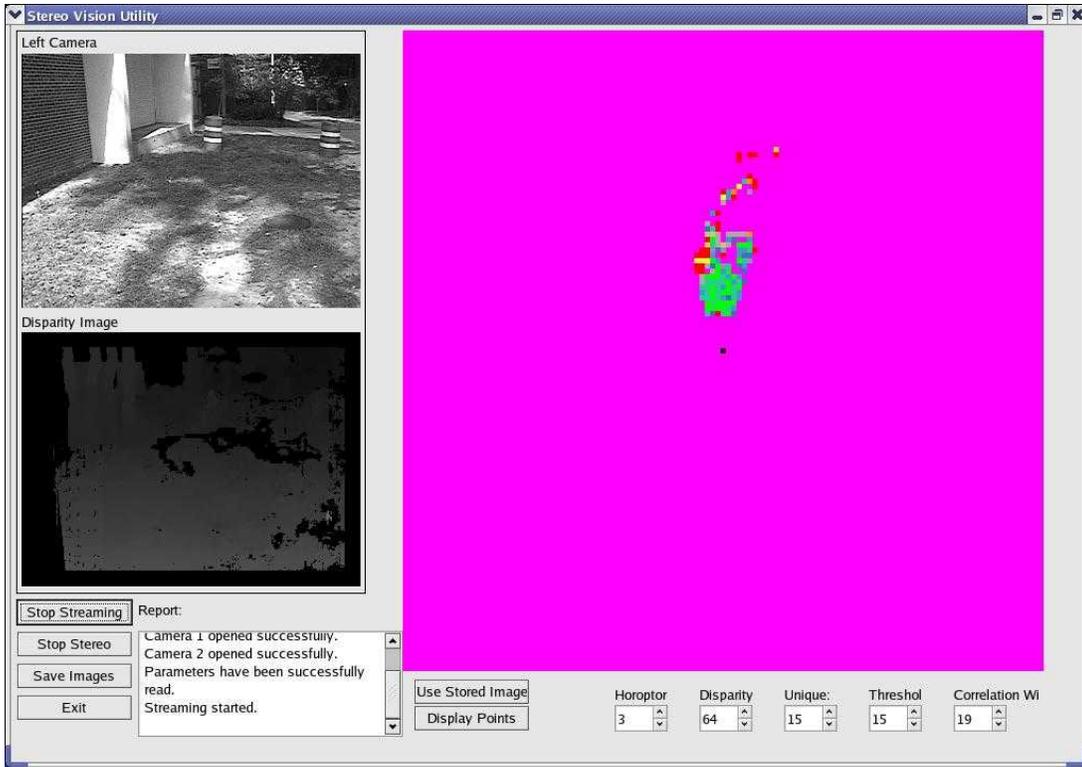


L

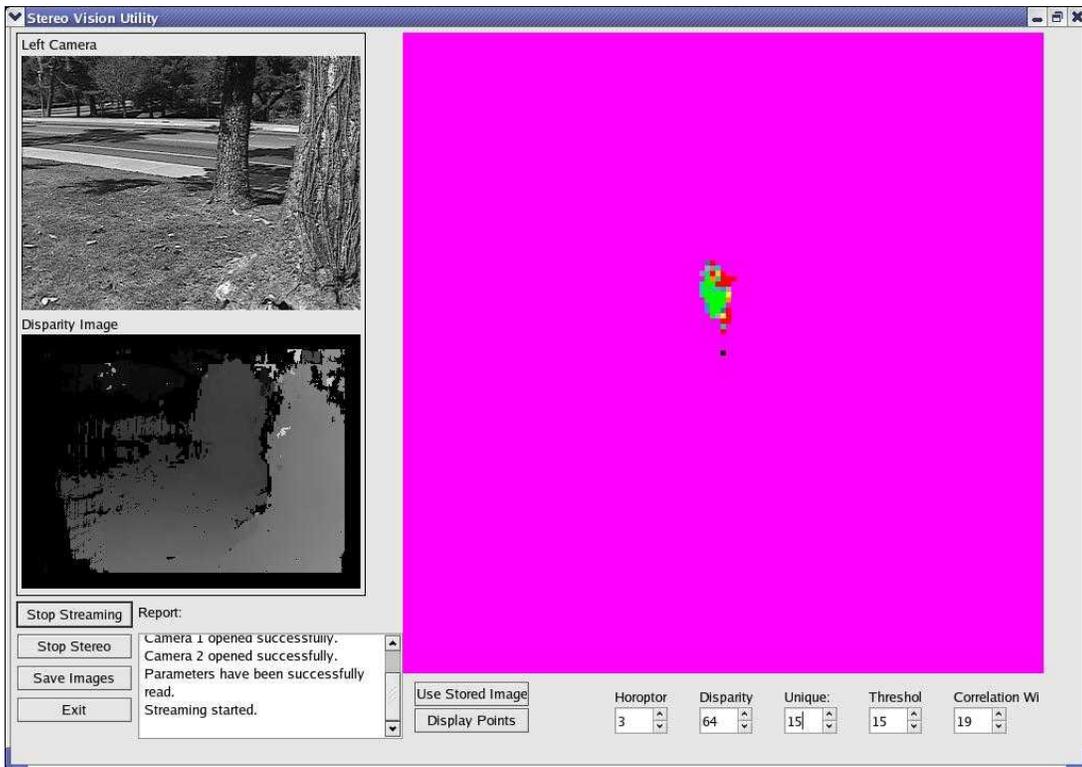


M

Figure C-1. Continued

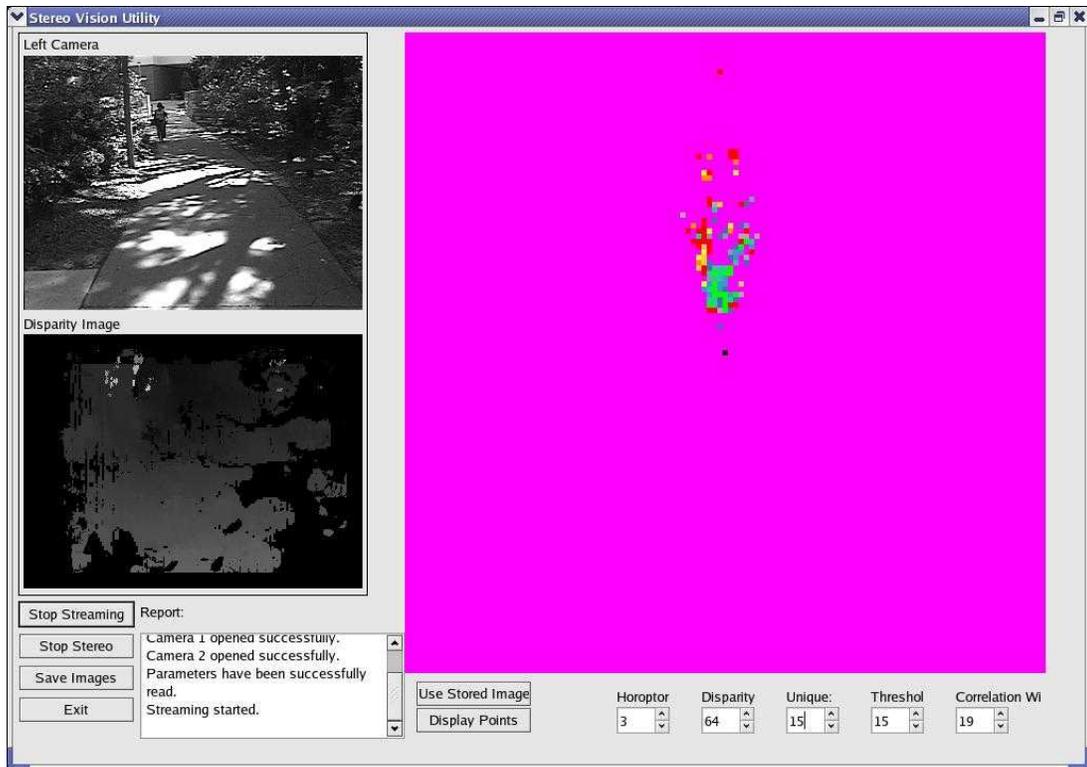


N

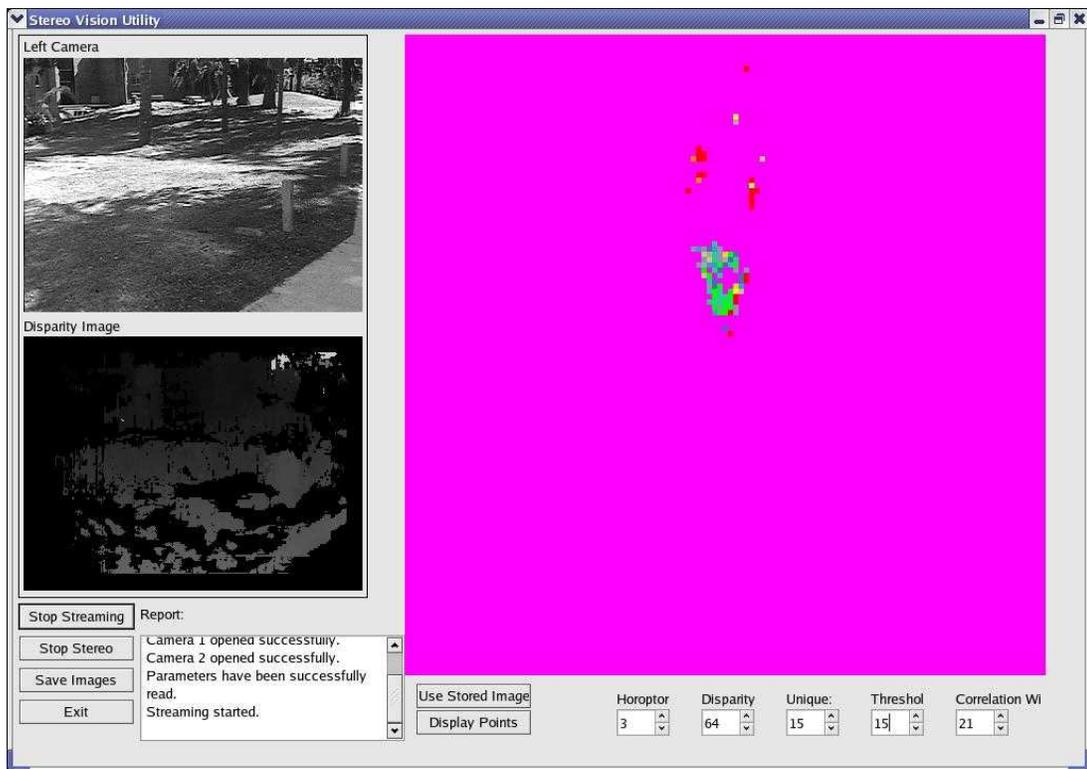


O

Figure C-1. Continued

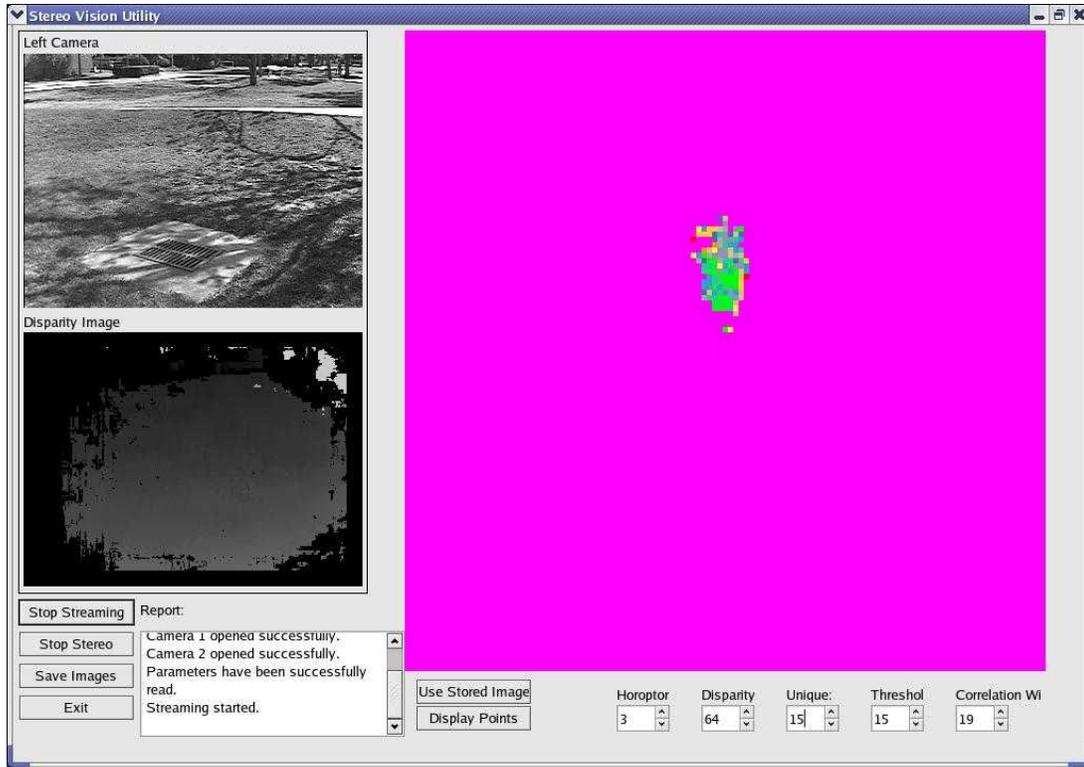


P



Q

Figure C-1. Continued



R

Figure C-1. Continued

LIST OF REFERENCES

1. C. Crane, D. Armstrong, M. Ahmed, S. Solanki, D. MacArthur, E. Zawodny, S. Gray, T. Petroff, M. Griffis, C. Evans, "Development of an Integrated Sensor System for Obstacle Detection and Terrain Evaluation for Application to Unmanned Ground Vehicles," SPIE Defense & Security Symposium, Vol. 5804, Pages 156 – 165, Orlando, FL, March 2005
2. C. Evans, "Development of a Geospatial Data Sharing Method for Unmanned Vehicles Based on the Joint Architecture for Unmanned Systems (JAUS)", M.S. Thesis, University of Florida, Gainesville, FL, 2005
3. S. Goldbert, M. Maimone, L. Matthies, "Stereo Vision and Rover Navigation Software for Planetary Exploration," IEEE Aerospace Conference Proceedings, Vol. 5, Pages 5-2025 – 5-2036 , Big Sky, MT, March 2002
4. W. Huang, E. Krotkov, "Optimal Stereo Mast Configuration for Mobile Robots," International Conference on Robotics and Automation, Vol. 3, Pages 1946 – 1951, April, 1997
5. JAUS Working Group, "Reference Architecture Specification, Volume II, Part 1, Version 3.2," The Joint Architecture for Unmanned Systems, <http://www.jauswg.org>, August 13, 2004
6. W. Kim, A. Ansar, R. Steele, R. Steinke, "Performance Analysis and Validation of a Stereo Vision System," IEEE International Conference on Systems, Man, and Cybernetics; Vol. 2, Pages 1409 – 1416, Hawaii, October 2005
7. B. Klaus, P. Horn, "Robot Vision (MIT Electrical Engineering and Computer Science Series)," MIT Press, McGraw-Hill Book Company, Cambridge, MA, 1986
8. K. Konolige, D. Beymer, "Calibration Supplement to the User's Manual Software version 3.2b," SRI International, Menlo Park, CA, June 2004
9. K. Konolige, D. Beymer, "SRI Small Vision System, User's Manual, Software version 4.1e," SRI International, Menlo Park, CA, September 2005

10. V. Meli, "News Spotlight, The Value of the Lens to the Camera," ADEMCO Video Systems, Louisville, KY
11. "Sensor Data Transfer Interface Control Document, Version 2.0" NaviGATOR Grand Challenge Architecture, University of Florida, Gainesville, FL, May 13, 2005
12. L. Shapiro, G Stockman, "Computer Vision," Prentice Hall, Upper Saddle River, NJ, 2001
13. S. Singh, R. Simmons, T. Smith, A. Stentz, V. Verma, A. Yahja, K. Schwehr, "Recent Progress in Local and Global Traversability for Planetary Rovers", IEEE Conference on Robotics and Automation, Vol. 2, Pages 1194 – 1200, San Francisco, CA, April 2000
14. C. Urmson, M. Dias, R. Simmons, "Stereo Vision Based Navigation for Sun-Synchronous Exploration," IEEE/RSJ International Conference on Intelligent Robots and Systems, Vol. 1, Pages 805 – 810, September, 2002
15. N. Vandapel, S. Moorehead, W. Whittaker, "Preliminary Results on the use of Stereo, Color Cameras and Laser Sensors in Antarctica," International Symposium on Experimental Robotics, Vol. 250, Pages 59 – 68, Sydney, Australia, March 1999
16. D. Wettergreen, B. Dias, B. Shamah, J. Teza, P. Tompkins, C. Urmson, M. Wagner, W. Whittaker, "First Experiments in Sun-Synchronous Exploration," IEEE International Conference on Robotics & Automation, Vol. 4, Pages 3501 – 3507, Washington, DC, May 2002

BIOGRAPHICAL SKETCH

Maryum Fatima Ahmed was born on December 25th, 1979 in Chicago, Illinois. She moved to Florida in 1992. In 1998, she graduated from Duncan U. Fletcher High School in Neptune Beach, Florida. She then began working on her Bachelor of Science degree in aerospace engineering at the University of Florida and received her degree in December 2002. She continued her education at the University of Florida and joined the Center for Intelligent Machines and Robotics. She received her Master of Science degree in mechanical engineering with a minor in electrical engineering in August of 2006. Maryum will begin working for Northrop Grumman Corporation in Melbourne, Florida during the summer of 2006.