2011

Optimization of Proximity Judgment

Brian Day
University of South Florida, bjday@mail.usf.edu

Optimization of Proximity Judgment

by

Brian Day

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Computer Science and Engineering
College of Engineering
University of South Florida

Major Professor: Dewey Rundus, Ph.D.
Dmitry B. Goldgof, Ph.D.
Luther Palmer, Ph.D.

Date of Approval:
October 18, 2011

Keywords: Computer Vision, Stereovision, False Color, Depth Cues, Teleoperation

DEDICATION

To my wife, for putting up with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

As humans, we have evolved to see in three dimensions. Our ancestors developed two eyes that only look forward, which allows the visual area that can perceive depth to be most of the field of view. A variety of sensors have been developed which can determine depth in the environment. They range from producing individual points of depth to the depth of everything in the environment. These sensors have become cheap and can now reliably produce accurate depth. Research is needed to determine how to present the proximity information to the people using the sensors. Touch, sound, and vision have all been used to provide depth information to the users. This research focuses on vision and compares methods of visually presenting proximity information to a user. The methods examined are stereovision and false color visual proximity mapping. False color mapping proved most effective while, surprisingly, stereovision was not helpful.

CHAPTER 1

INTRODUCTION

Watching a bomb squad training day made it clear that a way to help the technicians perform their jobs of guiding a robot to approach and defuse a bomb, would be to increase their ability to judge proximity. Their technique employed a physical feeler attached to the robot. When the feeler contacted a target object, it would indicate to the operator that the object was a known distance away. Better techniques than the physical feeler, must exist.

An examination of the existing research relating to distance judgment reveals the majority is on perfecting the sensors and calculating depth from their output. Little research seems to address the presentation of the calculated proximity to the users. The focus of the current research is on that presentation.

Most commercial teleoperation, specifically robots, is designed only with mono-cameras. This design has the operator relying solely on monocular depth cues to understand the environment around the robot. This approach limits the operator's ability to operate in the environment because it does not allow the use of other information commonly used in depth judgment.

When depth sensors, or a stereocamera, are used to get the proximity information, only a few visual displays and fewer haptic interfaces for proximity judgment have been explained. The simplest presentation is to show the raw depth reading. If enough information is available, a proximity picture can be created. The picture is based on a mapping of the depth readings to other colors or to white intensity. Proximity information may also be delivered with either sound or vibrations to indicate proximity.

The interfaces this research compares are based on vision: the first is simple 2D, the second is 3D, the third is 2D again but with augmented visual depth mapping, and the fourth is 3D with augmented mapping. All the interfaces are shown using a stereo headset, and through shutter technology the 3D interfaces are shown. The standard 2D display is the control, which is what most teleoperation users use. The standard 3D is another control, similar to how our eyes naturally see the environment.

CHAPTER 2

HOW PEOPLE PERCEIVE DEPTH

People perceive distance primarily through vision, using a variety of depth cues. Depth cues are broken into two groups, monocular and binocular depth cues. Monocular depth cues are used when there is only one source of vision, such as one eye or one camera. Binocular depth cues are only available when two or more sources of vision are used, such as two eyes or stereo cameras. However, if two or more vision sources are available, all monocular depth cues can be used along with the binocular depth cues.

2.1 Monocular Depth Cues

Even though the standard way for a human to view the world is through stereovision, monocular depth cues are still used. At times we are forced to only use monocular depth cues. For example, a standard television, movie, or any video shot with a standard camera only gives a mono-view of the environment. However, we can still determine which objects are closer to the camera lens even though we cannot use stereovision, by using monocular depth cues which are: interposition, relative height, familiar size, texture gradient, shadow, linear perspective, motion parallax, and atmospheric blur [1-3].

Interposition is otherwise known as overlapping. When one object overlaps or obscures another object, as seen through the viewer's field of vision, the viewer interprets the obscuring object as closer than the object being obscured. [1-3] See figures 2.1 and 2.2. In figure 2.1 the bird is in front of the dog and is covering part of the dog, showing the bird is in front of the dog, interposition. In figure 2.2 there are three mountain ranges. The first range covers part of the second range and the second range covers part of the third range, showing interposition of the ranges.

Figure 2.1. This image shows interpolation, relative size, familiar size, and texture gradient.

Figure 2.2. This image shows atmospheric blur and texture gradient.

The use of shadows gives another depth cue. The rules of using shadows for depth cues are as follows. “With one source of light, all shadows lie in same direction. With an object on the ground, the shadow appears on the other side of the source. With a hole in the ground, the shadow appears on the same side as the source. The object covered by the shadow is perceived to be further away then the object in the light” [1]. The shadows add another feature to objects that can use interposition. If the light source is known, and where the shadow of one object lays in regards to a second object, the viewer can judge which object is in front of another [1-3]. For example, if the light is coming from the right and the closer object is to the right of the second object, the closer object's shadow will be in front of farther object. This allows the relative position of the two objects to be known even if the objects are not near each other as long as object one's shadow extends in front of object two.

Linear perspective is the visual effect of parallel lines seeming to meet at the point of convergence. This can be seen when walking down a street. The lines of the sidewalk, curbs of the street, and roof tops all seem to be running towards a point far in the distance. That is the point of convergence. So while the lines are parallel, they seem to be getting closer to each other. When two similar sized objects are at different distances from the observer, the farther object appears smaller than the closer object. In the absence of other information, the apparently larger of two objects is considered closer than the other object [1-3].

Relative height is the height of the object relative to the viewer. If there are two seemingly similar objects in the field of view, the object that is visually larger is perceived as closer than the smaller object. This effect is also seen in the linear perspective. Because parallel lines converge at a point in the distance, two objects of equal actual height at different distances along those lines will be perceived as having different relative heights to the viewer [1-3]. See figure 2.1. In this picture the bird looks larger than the dog; its height in the picture is greater than the height of the dog. This is consistent with the bird being in front of the dog.

Texture gradient is how the viewer perceives the textures of the objects in the environment. The roughness, shapes, and patterns of the objects visually change with distance. The closer the objects are, the better the textures can be seen and distinguished. As the distance increases, the ability to perceive the textures diminishes. The textures become finer and less visible as the distance increases [1-3]. See figure 2.1 and 2.2. In figure 2.1, if the animals were reversed you could see the dog's fur much better, but you would not be able to distinguish the feathers of the bird. In figure 2.2, each farther range has less and less visual texture to the range.

Motion parallax is the visual effect of the environment as either the observer or the environment moves. The object at the focal point of the viewing system will not move in the observer's perspective due to the system tracking it. Objects closer to the observer then the focal point will move in the opposite direction of the observer's perceived motion, however, the objects past the focal point will appear to move in the same direction of the observer's perceived motion [1-3].

Atmospheric blur is a visual effect due to water in the air. When a viewer looks into the far distance, there appears to be a haze or fog, causing the object's edges to be blurred. While viewing close objects, there is not enough water in the air to visually obscure the object. However, as the objects become farther and farther away from the viewer, the amount of water in the air between the observer and the object increase, blurring the object [1-3]. In figure 2.2 there are three mountain ranges. The closest range is the darkest of the ranges. The second range is slightly lighter. The third range is lightest and most blurry. This lightening and blur is due to the water vapor in the air that the reflected light must travel through.

The monocular depth cues that are important to this research study are: size of object, object movement, and perspective. These depth cues were programmed into the test simulation. They were chosen because they are some of the most commonly used depth cues. The other depth cues, while useful for depth perception, were not included in the experiment due to the constraints of the depth cues in the simulation, such as distance required for atmospheric blur.

2.2 Binocular Depth Cues

Stereopsis is based on calculating the difference in location of an object in two pictures taken from two vantage points. For the explanation please refer to figure 2.3. The distance between the two eyes is called the base line (b). As long as the base line is greater than zero, the view from each camera focused on an object will be different. This is due to the different angles of view which allow for the calculation of the z coordinate. To calculate the z coordinate of the point P, solve the equation $(b*f)/(X'l - X'r)$. The variables in the formula are b= base line, f is the focal length of the eyes (distance between retina and lens), and the $X'l - X'r$ is the disparity which is the distance of the object position between each image. The closer the object is to the eyes, the greater the disparity. This calculation is stereopsis and can be performed on any point that is visible in both images [4].

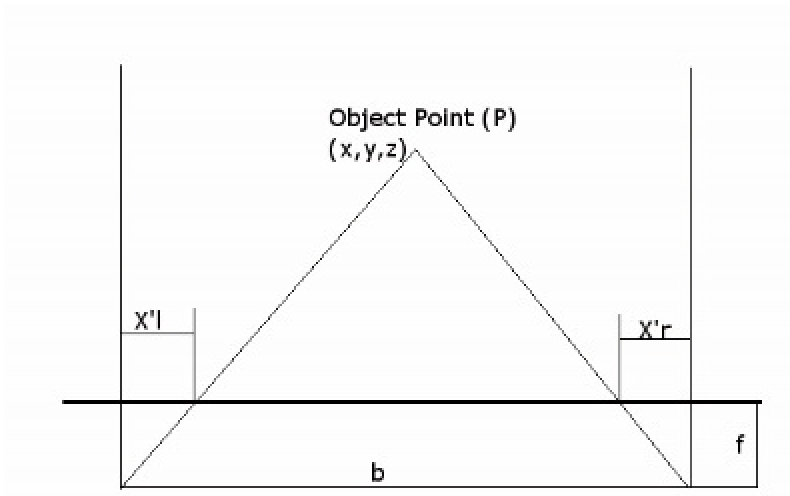

Figure 2.3. This image shows the necessary variables to calculate Z using stereopsis.

Convergence, the second binocular depth cue, is the finding of the angle between cameras necessary to have one point centered in both images. Human eyes converge when focusing on a given spot [2, 3]. Someone "crossing their eyes" is an extreme example of convergence, for a point very close to their face. We can judge the approximate distance to an object by noting how much our eyes need to converge to focus on it.

CHAPTER 3

RELATED WORK IN PROXIMITY JUDGMENT AND PRESENTATION

As technology has progressed, new distance sensors and display methods have been created. The drive for this progress is in part related to the fields of robotics and teleoperation. The presentation of proximity information to people who can't detect depth, such as the blind, has also been studied. Section 3.1 provides a discussion of common sensors used to obtain depth information. Section 3.2 considers common alternative visual depth displays. Section 3.3 considers haptic displays.

3.1 Sensors

Five main types of sensors are used for determining distances: sonar, laser, planer laser, stereovision, and miniature range sensors. Each has its advantages and disadvantages. Not only does each sensor type measure the environment differently, each also produces different forms of distance output. Sonar and laser sensors produce a single distance reading at a time. Planer lasers produce the distance readings in one plane; however, there are ways of moving the sensor to scan the whole environment. Stereovision sensor systems and miniature range sensors also allow for the calculation of depth of the whole environment.

The simplest sensor is sonar. Sonar sensors are active because they rely on sound pulses to determine the depth of the closest object in range. The sensor generates a pulse of sound, usually sounding like a "click", which bounces off the objects in the environment. The sensor listens for the first return "click", than calculates the time it took to return, giving the distance to the closest object in the sensor's range.

There are problems inherent with the sound based sonar sensors; a significant one is that of sound being absorbed or reflected away from the sensor [5, 6]. Another difficulty is that in order to create a 3D map of the environment, the sensor will have to scan all points in the environment, one at a time, moving or rotating the sensor after each reading.

There are also laser point sensors that replace the sonar with laser light to produce the singular depth value. The reason for using laser light instead of sound is precision. Sound spreads out as its moves away from the source and this covers a wide area. Laser light negligibly spreads from the source covering a tiny area in comparison to sound. The laser will return the distance to exactly what the laser is pointed at instead of the closest object in a wide area.

The planer laser sensor uses the equivalent of the laser point sensor to scan one plane of the environment. The sensor consists of a laser, spinning mirror, and a light sensor. The laser shines on the mirror; as the mirror spins it reflects the laser light into the environment in a series of points along the plane where it shines on objects. The returned light is picked up by the light sensor using time of flight to calculate the distance to the reflected point. The mirror only spins on one axis, causing the laser to travel only along one plane, hence a planer laser range finder.

Methods have been developed that allow the planer rangers to range the entire environment in front of the sensor. This is done by moving the plane of detection. This is accomplished by mounting the ranger onto a movable platform. The platform can other be spun around the Z axis or the platform may be raised or lowered, in a nodding motion, along the vertical axis. This movement allows the scanning plan to pass over the entire area forward of the sensor line by line, [7]. Look at Figure 3.1. The left picture is taken from a normal camera at the same point in space as the planer laser. The right picture is the 3D point cloud taken by the spinning planer laser at the same place as the normal camera. A point cloud is created by software that places the points detected by the senor into a 3D computer generated environment. The point cloud in figure 3.1 has been mapped to the color spectrum. Instead of mapping the color spectrum onto the total distance the sensor can detect, the spectrum was mapped onto a smaller distance, than repeated, giving the right image a striped look.

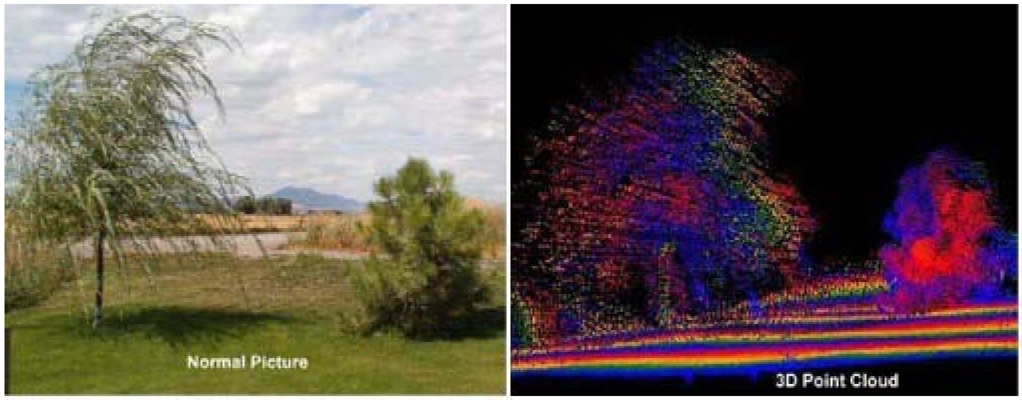

Figure 3.1. Output from the spinning Ladar. On the left is a standard picture of an environment. The right picture is an image of the depth map produced by the spinning laser mapped to a color repeating color spectrum. [7]

A stereovision camera system produces a full 3D depth map of the environment. Artificial stereovision systems retrieve three-dimensional information about their surroundings using the same principles as the human visual system. Two cameras each take a picture of a scene at the exact same time. Because these images are taken from different viewpoints, features appear at different locations in each image. The closer the feature is to the cameras, the greater the difference in location of that feature between the images; that distance is called disparity. Based on the disparity, the three-dimensional location of objects can be found. This means that if both cameras can see the object, then a depth value may be calculated for it. These sensors are passive and rely on the reflectivity of the objects [7].

There are two different types of stereovision cameras, converging and non-converging. Converging camera systems rotate to have both cameras point to the same focal spot. Converging camera systems more closely replicate the human eye. Non-converging camera systems use fixed cameras, but can still calculate distance. Fixed cameras tend to be more commonly used since they do not have parts that will wear out due to movement and require less power and fewer controllers because the cameras do not have motors. They tend to be more robust because the cameras are locked into the case and are designed not to move.

The miniature range sensors are new on the market. They produce a full 3D depth map in near real time. The sensors in essence take a depth picture of the area in their field of view. These active sensors use a flash of near infrared laser light that illuminates the environment. The light is then reflected back towards the sensor by objects in the environment. As was the case for laser range finders, time of flight is used to calculate depth. An array of CMOS chip sensors detect the phase shift of the returned light to calculate depth. The field of view is then split up between the two-dimensional array of CMOS chips. Each chip provides one depth reading, the closest object, for the area that chip is to cover. Each CMOS chip equates to one value in the finished depth map. This allows for all objects in the sensors' range, typically about 12m, and field of view, to have their depths obtained. Like sonar sensors, these can have problems with light being absorbed or not reflected back to the sensor. "The sensor emits laser pulses in the near infrared spectrum, if a material absorbs this energy instead of reflecting it the object is invisible to the sensor" [8].

3.2 Displaying Proximity Information

Sonar and point laser sensors provide a single depth reading for an object. This is typically displayed as a numeric value to the user. It is, of course, possible to produce other types of displays. Kay [9], for example, has developed a system that uses sound to give the user proximity information. Kay's system, Sonic Torch, uses a sonar sensor attached to a walking cane used by the blind. The Sonic Torch uses a wide angle sonic sensor to probe the environment. A high frequency tone is played followed by a low frequency tone. The time between the tones is an indication of depth. The greater the depth measurement is, the greater the time between the tones. An alternate type of visual display can also be created for the single point sensors. Instead of a numeric display, a visually displayed block could change color or intensity depending of the depth value.

The planer sensor has an interesting interface. The sensor scans 180 degrees from left to right but it only scans the height of the laser beam. This leads to two common visualizations for the interface. The first method is a top down look onto the Z axis. Please refer to figure 3.2.

In this image the entire sensing range is displayed. An object, a block in this example, is placed in the sensing range, cutting the laser beams, causing a “V” shape is the visualized sensing range. Another option of the look down method is to only show the sensed object and not the full sensing range. Either way the sensor can only detect the parts of the object that can be hit by its laser light.

A second interface method could be a false color display. The screen is divided into as many sections left to right as there are data points created by the planer sensor. Then the distance calculated by the sensor is used to color the corresponding sections using a color spectrum mapping from the data points.

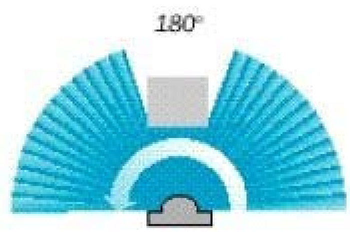

Figure 3.2. Top down view of the Z axis. [10].

The other methodology for using the planer ranger is to rotate the sensor on either the X or the Z axis. This method produces a depth map that no longer covers one plane, but produces a depth map of all objects forward of the sensor [7, 11]. This methodology produces a depth map equivalent to the kind produced by a stereovision system or by the miniature range cameras. There appears to have been little work done in displaying this kind of depth map to the users.

Only six ways, besides haptic Interfaces have appeared in the literature as methods for displaying proximity to the user: numeric values, point cloud, intensity, false color, anaglyph, and stereovision. Point cloud, anaglyph, intensity, and false color methods create images from the depth values by assigning pixel values per the depth values.

The false color method [12] assigns different colors to the depth values. To create an image to show to the user, each position in the depth map is assigned to a corresponding pixel in the image. Usually [0,0] in the depth map is assigned to pixel [0,0] in the image. Then the proximity-color mapping is used to determine what color the pixel should be assigned. Usually red represents the near color and blue the far color. A color mapping scale allows a user to determine both absolute and relative distance for objects in the scene. Figure 3.3 is an images produced by the Canesta sensor. On the left is an image of the returned intensity of a hallway and the right image is of the proximity of the hallway mapped to false color.

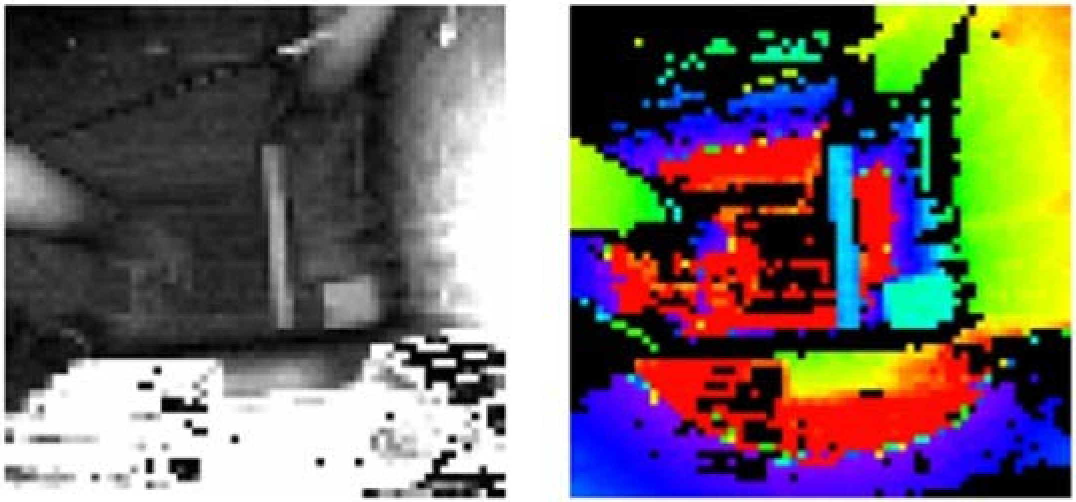

Figure 3.3. Example output from the EP208 sensor. Left: Active Brightness Image (grayscale). Right: Range Image. [8].

A second method is a point cloud [13]. This method utilizes software that maps each depth point into a 3D environment. This method is illustrated in figure 3.1 while showing both false color and point cloud methods combined. The left picture is taken from a normal camera at the same point in space as the planer laser. The right picture is the 3D point cloud taken by the spinning planer laser at the same place as the normal camera. The point cloud has been mapped to the color spectrum. This means that between zero and a set distance, the entire color spectrum is mapped. The difference between point cloud and false color is 3D vs 2D displays. For false color alone, a 2D display of the image is shown with pixels colored to represent depth. A point cloud display adds 3D dimensionality to the 2D display. Also, due to the 3D mapping of the points in the cloud the user can move the camera around in the point cloud. The software keeps track of where the planner ranger is scanning and then maps that depth value onto the display. Because the sensor can measure the environment deeper than the basic color mapping range, the color spectrum is repeated. This gives the viewer a recognizable way to view proximity. The user can then move around in the point cloud or manipulate the color mapping's min and max distance settings to get a sense of where everything is located. As seen in figure 3.1, each depth reading from the sensor is mapped and produced a 3D map of the environment.

A third method adjusts the intensity of the pixels, in the image, is changed based on proximity [14, 15]. In the first method of this type of display, the brighter the image is, the farther the scanned object is from the sensor. In the second method, the brighter the image is, the closer it is. In both methods, objects are identifiable but small changes in proximity are not easily identified. The images produced look like gray scale pictures taken in low resolution, see figure 3.4. The left picture is from a normal camera mounted approximately where the depth sensor is located. The right figure, where, instead of gray scale, green scale is used, is where the proximity measurements are mapped to a changing intensity where the greater the intensity, the closer the object is. The sidewalk is not visible in the right picture, except for the black spot in the middle top, due to sensor being unable to detect depth there. However the pole is clearly visible and has the same intensity throughout, meaning that the pole is straight up and the entire pole is equal distance to the sensor. It is also possible to use a selected hue rather than gray scale, in this figure, green.

Figure 3.4. Left: Left image in a stereo pair, Right: Disparity image. [7]

A fourth method is anaglyph [16, 17], which will take two stereo images from a scene and a proximity map of the scene and combine them so when the anaglyph glasses are used the scene is seen in 3D. The images are created using a stereo camera method. This is the same technology used in older 3D movies. See figure 3.5 for an example of an anaglyph image. The image to be used by the right eye is tinted cyan and the one for the left is tinted red. Glasses are worn with the left lens tinted cyan and the right tinted red. This tinting allows the one image shown to be split into two images, one for each eye. This allows for the stereovision that is the standard for human vision. The same principles that apply to stereovision apply to Anaglyph images.

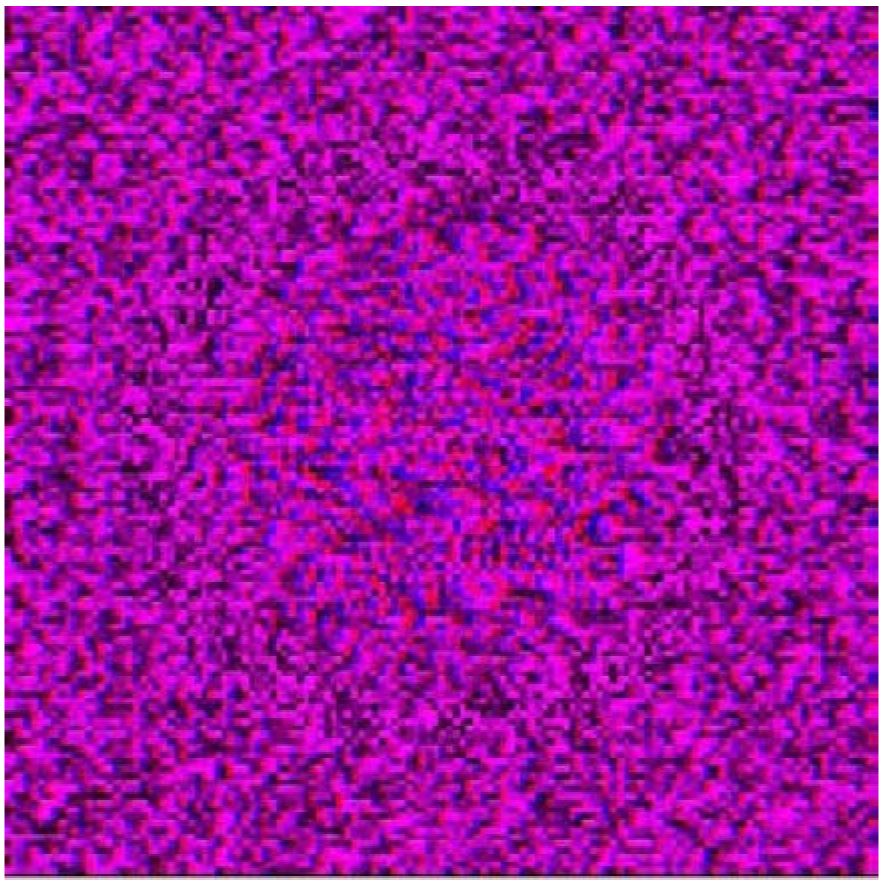

Figure 3.5. Anaglyph image. [16]

Another method is stereovision [15, 18], which produces two images and the distances for each pixel in the images. The main way to view this method is to use either stereo glasses or a stereo display. Both ways show the right image to only the right eye and the left image to only the left eye. Stereo glasses have two small monitors in front of each eye. Either two video streams are shown simultaneously, or the video frames are interleaved together and then split between the monitors. Stereo displays use crystals to shine only the correct images to each eye. The use of the two images allow for the human brain to combine the images, just as the brain does using the user's eyes. This is the most natural display.

3.3 Haptic Displays

Four main haptic interfaces have appeared in the literature: TENS, Sonic Torch, vOICe, and a brain implant. These interfaces are primarily are used by the blind and are not visual, except for the brain implant. The TENS (Transcutaneous Electro-Neural Stimulation) has the user wear a pair of gloves with electrodes mounted in them. The environment in front of the person is split into ten sections and each section's value is sent to the corresponding finger. To interpret the signals from the TENS, the user imagines that his/her hands are at waist height with palms horizontal with fingers extended [6]. The detected proximity is mapped to electrical impulses which can be sent to the fingers with each finger representing one section.

Kay's Sonic Torch has a sensor mounted on the cane and two tones are played to the user. The shorter the time between tones indicate the proximity to the closest object. Work is also being done in which sensors are mounted on the person's head. These methods are an attempt to give the blind person an additional sense of the proximity around them based on where he or she is looking, see Figure 3.6. These methods use sound to give the user the proximity map. Each depth value is given a certain tone. First a proximity map is created by the depth sensor worn by the users. Then each depth value's tone is played, in order to allow the person to hear the environment around him or her [6]. The main problem with this method is human memory. For any resolution, beyond coarse, of the environment the user will forget what the first part of the scene was by the time the end of the scene is played. This means that only a coarse proximity map can be used.

Figure 3.6. Left: The vOICe auditory substitute vision system, Right: A soundscape image. [6]

Another way proximity is displayed to blind people is to physically simulate the visual cortex of the brain. The blind person will be able to see blobs, see Figure 3.7. A television camera is placed within a pair of glasses for the person to wear. The video signal is then fed into a computer. The computer will then convert the video signal into electrical pulses that will stimulate the visual cortex via surgically embedded wires. The stimulation will cause the person to see flashes or blobs of light. These blobs can be connected to depth values that the user can interpret with training.

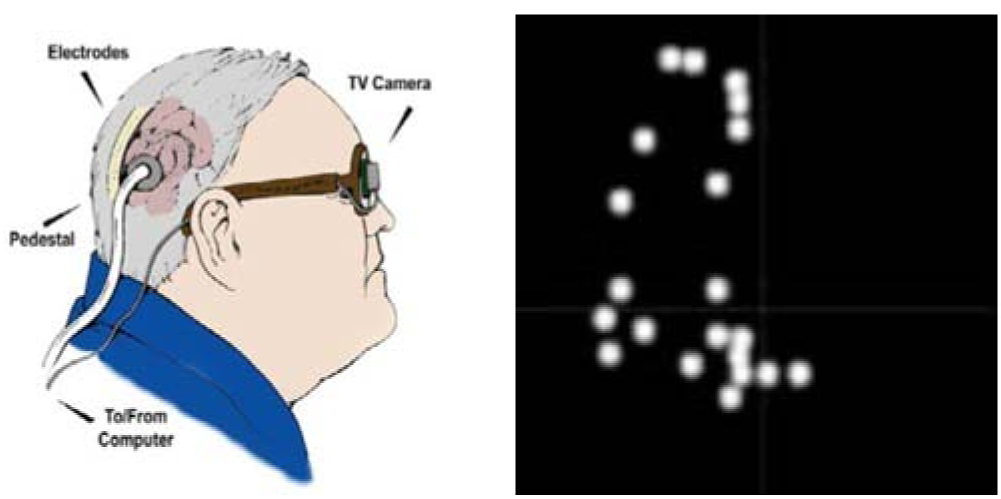

Figure 3.7. The Dobelle brain implant. (a) The visual cortex implant. (b) The resulting available vision. [6].

CHAPTER 4

CHARACTERISTICS OF THE HUMAN EYE

4.1 Characteristics Involving Depth Perception

4.1.1 Accommodation

Accommodation is how the eye focuses on objects and keeps them in focus as they move in the field of view [19, 20]. The muscles inside the eye control the shape of the lens, which focuses the light onto the retina. When the lens is thinner (muscles causing tension on the lens), farther objects become focused. However, when the muscles reduce tension the lens, it gets thicker in the middle allowing closer objects to come into focus.

4.1.2 Chromostereopsis

Because the focal distance for light passing through the lens is dependent on the wavelength of the light, colors at the ends of the visible spectrum will have different focal lengths. “Chromostereopsis is when two colours (generally red and blue or red and green) presented in the same depth plane are binocularly perceived as residing in separate depth planes” [21]. In lay terms, this means that when two blocks of different colors are equidistant from a viewer, they will not be perceived as equidistant by the viewer. Because parts of this research deal with depth and color perception, these phenomena need to be mentioned.

“The usual stimulus for the study of chromostereopsis consists of a target with red and blue regions where the red portion will typically be perceived as being in front of the blue, called positive chromostereopsis.” [21] This indicates that in situations involving objects of different colors, the objects whose color is a shorter wavelength will typically be perceived as being farther away than objects with colors of longer wavelengths. To take advantage of this, the mapping of the color spectrum to a proximity map would entail that the longer wavelength, red, would be mapped to the less distant proximity values while the shorter wave lengths should be mapped to the more distant proximity values. This allows the closer objects to naturally appear closer and the farther objects to naturally appear farther away. This mapping may be thought to produce a bias towards a proximity judgment exaggeration. An observer will receive both a proximity judgment based on the hue of the pixels and a proximity judgment based on a positive chromostereopsis. This may result in the observer perceiving the objects as farther apart than they are.

4.2 Sensitivity to Color

If color is used as an abstract representation of proximity, it will be necessary for the operator to be able to distinguish hues. Figure 4.1 shows the sensitivity of the human eye. As this shows, the eye is most sensitive to the hues between green and yellow. However, if false color were to be used in a teleoperation interface, the sensitivity to yellow, coupled with chromostereopsis, leads to a robotic interface having the grasper position i.e. zero point set to yellow, with red for negative values. This results in the greatest precision to be mapped to the greatest sensitivity.

When precision work is needed, the hue to distance mapping needs to be changed. When working with a robotic grasper, the need for long distance proximity information is not needed. So to switch from driving proximity information to grasper proximity information, the max distance parameter in the color mapping, is changed to a smaller distance value. This remapping makes a larger hue change for the same change in distance; causing small changes in distance to become more evident to the user.

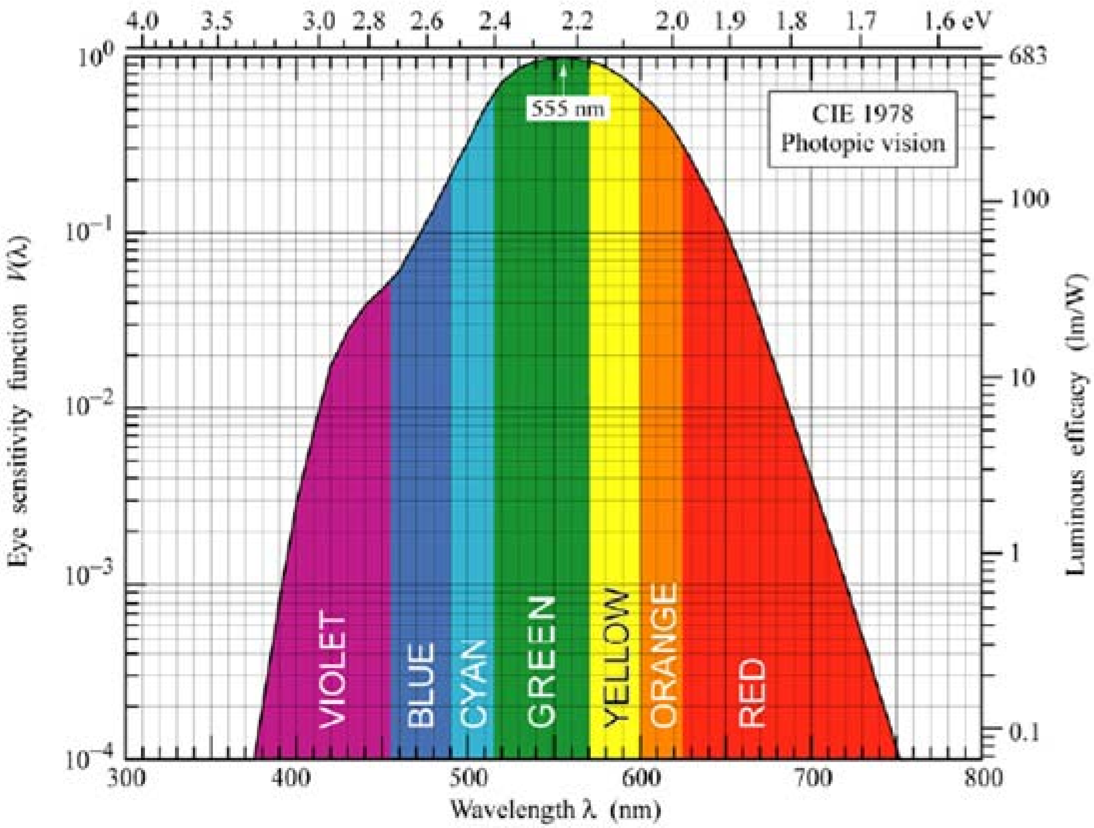

Figure 4.1. The average change in wavelength required for a perceivable difference in hue to appear. [22]

CHAPTER 5

HYPOTHESES AND PURPOSE

5.1 Initial Interests in Mobile Robotics

The initial idea for this research came from assisting with a bomb squad training/ demonstration day. All bomb squads must have and be proficient with EOD (Explosive ordinance disposal) robots. During that day the bomb squad's EOD robots were tested in simulated situations. These situations spanned navigating a simulated airline hallway (in order to reach a device), to picking up an egg without crushing it (testing gripper dexterity).

The robots used in the training day were brought in by the local bomb squads so they could train on/be familiar with their own robots. The robots did not have any additional sensors or cameras mounted to them; they were original from the manufacturer. This is important because with standard EOD robots only mono-cameras are mounted on the robots for proximity detection. Most of the robots had two to three cameras: a drive camera mounted on the body, a gripper camera mounted on the arm, and optionally a mast camera mounted above the robot. Even with these multiple cameras, the operators had trouble detecting proximity of objects to the robot, causing the robot's gripper to overshoot or run into the object.

The bomb squads expressed interest in a system that could provide proximity sensing and in an interface that could best provide that information. That started the process of finding different proximity displays and finding the lack of comparisons between them.

5.2 Lack of Comparisons

There are companies and researchers that are dealing with depth sensors and proximity sensors, as shown in Chapter 3. However there does not appear to be much, if any, research on how to present the data from the sensors. Each company seems to create an interface as they see fit. However, this may not produce the best results for the end user.

As a reminder, the focus of this paper is on stereoscopic view and a false color/intensity view. There is a need for comparing the interfaces. When robots are being controlled for precision work, such as with the bomb squad, the operator would need the best interface for presenting proximity information. With a lesser interface, the operator would have a harder time accurately determining proximity. He/she may even misjudge the proximity, causing an accident.

While the companies and researchers might do testing of their interfaces, there does not seem to be any constancy in type of interface. There does not appear to be any formal comparison of the visual interfaces or information on the precision of the interfaces.

5.3 Comparison Proposal

The proposal of this project is to find the optimal way to present proximity information using visual interfaces. Of the interfaces found thus far in related research, there appear to be two that are best candidates. The first is stereovision; it is the closest to normal human vision. In practice, this method would require two mounted cameras. The second is false color mapping, where the color spectrum is mapped onto the proximity distance which is used to color the objects in the interface. In practice this method would require a single mounted camera and one of the previously described methods for determining measurement. In addition, the combination of the two methods is to be compared. Finally a control, the standard 2D view without augmentation will be tested as well.

5.4 Hypotheses

H0: Augmented stereo will show a significant difference between itself and plain stereovision, with augmented stereo being better for the user.

H1: Augmented stereo will show a significant difference between itself and false color.

H2: For each additional layer of dimensionality, there will be less of an effect if augmentation is used.

CHAPTER 6

EXPERIMENTAL METHOD / DESIGN / ANALYSIS

6.1 Methods

Four methods are to be tested. There will be two with stereovision and two without. There will be two with false color and two without. These are shown in Table 6.1. The first method will be 2D; both eyes will be shown the same image so binocular disparity will not be available. The second method is stereovision. It will be used as the control method because that is similar to how we are used to seeing the world. The third method will be false color. This method will not use the stereovision and both eyes will be shown the same image but will have the additional depth cue of false color. The objects in the simulation will be colored with the color mapping based upon their proximity in the scene. The fourth method will incorporate stereo and false color in an augmented stereovision.

To control for the practice bias, the order of the methods shown are changed. There are twenty-four permutations of the four methods. The first user is shown the methods in the first permutation, the second in the second permutation and this will continue to the twenty-fifth person where the permutation list will restart.

Table 6.1. Simulation method and corresponding attributes.

Simulation	Stereovision	Color Augmentation
1) Non Stereovision	No	No
2) Stereovision	Yes	No
3) False color	No	Yes
4) Augmented Stereo	Yes	Yes

6.2 Experimental Simulation

All proximity judgment methods will involve presentation of a display depicting two cubes in a room-like setting. One cube will be stationary and one will move in the z axis, both away from and towards the user. There will be four random starting points per user, one per method, for the stationary cube. The task of the user is to press the spacebar when he/she perceives that the two cubes are equidistant from him/her. In the false color simulations the cubes are colored based on the proximity of the cube. The stationary cube's color will stay the same throughout the simulation while the moving cube's color will change as its proximity changes.

6.3 Hypotheses

H0: Augmented stereo will show a significant difference between itself and plain stereovision, with augmented stereo being better for the user.

H1: Augmented stereo will show a significant difference between itself and false color.

H2: For each additional layer of dimensionality, there will be less of an effect if augmentation is used.

6.4 Apparatus

This study was created using an Intel Core 2 Quad CPU running at 2.66 Ghz with 4 GB of RAM. The computer used a quad buffered Nvideo Quadro series video card which was needed to drive the shutter glasses, i-glasses SVGA Pro. All displays were computer generated images.

6.5 Subjects

Each of the 48 subjects were tested in each experimental condition. The subjects were college students, both male and female. Most subjects were in the 18-22 age range with some of older ages. The subjects were screened for epilepsy and colorblindness.

6.6 Data Collected

Two types of data are to be collected. The time required for the subject to make a decision and the proximity of the stationary cube to the moving cube were recorded for the experimental trials. Following each experimental trial a questionnaire was administered asking for feedback about what depth cues were used by the subject to make his/her decision for each of the experimental conditions.

6.7 Procedure

1) Subject Arrives. The subject is asked about epilepsy and colorblindness. If either condition is present the subject is thanked for his/her time and released.

2) The experiment is explained. The subject is told that they are going to see four scenes. Each scene will require proximity judgments. In each scene there will be two cubes. The subject's task is to press the spacebar when he/she thinks that the moving cube is at an equal distance to the stationary cube. The subject will be shown the stereo headset that he/she will wear for the duration of the test.

3) Consent. The subject will sign a consent form to participate in the experiment. A new hygiene cover will be placed on the headset. The subject will be allowed as much time as they need to read and understand the research. The investigator will be available to answer any questions the subject may have.

4) Trials. All four methods will be presented in a random order. After each method is presented the subject will be given a questionnaire to indicate which depth cues he/she used for judging distance. Multiple sets of the four methods will be shown to the subject. Each run should take less than five seconds and less than three minutes to fill out the questionnaire. Each subject participated in a minimum of sixteen sets of trials with some faster subjects completing as many as thirty-eight sets.

6.8 Analysis

The design was analyzed using a Repeated Measure-Two Factor analysis of variance. The data is prepared for the analysis tests by first computing the depth error then averaging the error per method per person. This way each person will produce one score per method. This accommodated the different number of experimental sets completed by various subjects.

The experiment controls for common biases. To control for practice bias among the four methods, each participant was presented the methods in a different order. The number of participants was decided on by the number of permutations of 4 methods. The twenty-four permutations were then doubled to get the final forty-eight participants. Therefore, each permutation was run twice. Each person ran through each method at least eighteen times. This allows for the participants to become used to the interfaces, evening the experience level of the participants. In addition, a bias due to distance of the stationary cube was controlled by each participant having a random distance for the stationary cube for each method.

CHAPTER 7

RESULTS

The results of running the Repeated Measure-Two factor analysis test were not what was expected. As the hypotheses stated, it was believed that the 2D white cubes would be the worst case followed by standard stereo and color proximity mapping, with the augmented stereo being the best. The results did not show the hypotheses to be correct. This section will go into the results, and a possible explanation of the results.

7.1 Error Data

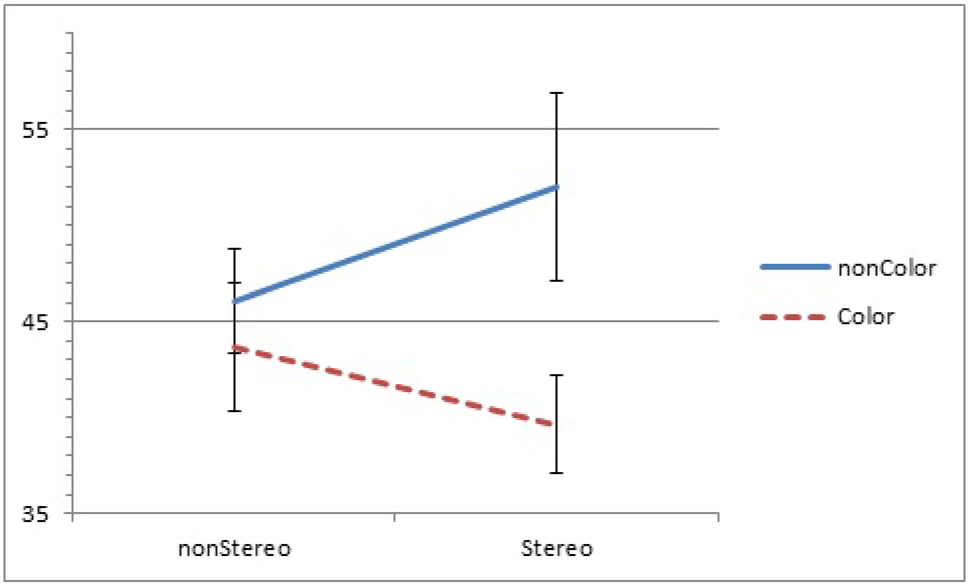

Figure 7.1. Graph with error bars of results, graphed data is in Table 7.1. The graph plots the average error in OpenGL units (Y) by Stereo (X) with the two lines representing Color vs nonColor

Table 7.1. Results of average error, standard deviation and standard error in OpenGL units. NSNC is nonStereo nonColor, SNC is Stereo nonColor, NSC is nonStereo Color, and SC is Stereo Color

	NSNC	SNC	NSC	SC
Mean	46.0625	52	43.66666667	39.64583333
STDDIV	18.6575	34	23.37969053	17.85579301
StdErr	2.69298	4.9	3.374567655	2.577261725

Table 7.2. Results of SAS two factor- repeated measure design for error. p<.05 if significant effect. Only Color has p<.05.

Source	SS	Df	ms	F	P
Total	116097.3	191			
Subjects	46046.31	47			
Color	2610.75	1	2610.75	5.869871	0.0193
Stereo	44.08333	1	44.08333	0.106789	0.7453
Stereo X Color	1190.021	1	1190.021	2.159499	0.1484
error Color	20904.25	47	444.7713		
error Stereo	19401.92	47	412.8067		
error Stereo X Color	25899.98	47	551.0634		

To understand the results, first the units must be explained. The units used are OpenGL units. The size of OpenGL units is based on a perspective ratio of how many units wide the object is, by how far the camera is from the object. For this experiment the cubes were 20 units wide and were between 550 units to 1200 units away from the object. The variable distance for the perspective was due to the cube moving from 550 to 1200 units as it moved in the screen. This means that table 7.1 is showing the mean, standard deviations and standard error of how many OpenGL units the subjects were off between the moving cube and the stationary cube. One OpenGL unit is equal to .92mm in this simulation. Therefore the cubes were 18.2mm wide moving from 506mm to 1104mm away.

The analysis was a Repeated Measure- Two factor test. This test checked for a main effect of the use of stereovision, a main effect of the use of false color, and an interaction between the two methods. There was not a significant main effect of the stereovision (p=0.7453) nor was there an interaction effect between color mapping and stereo (p=0.1484). Interestingly there was a significant main effect of color (p=0.0193). Please see Table 7.2. Figure 7.1 shows the mean error on the Y access, the X axis is whether stereo was used, and the blue solid line is no color, red dotted line is color. The results show the use of non-converging stereo cameras does not have an effect on the ability of someone to view proximity. Also, the results show that the use of non-converging stereo cameras along with color mapping does not significantly improve the use of color mapping alone. However, as seen in Table 7.2. the best result is actually stereo color, but not significantly better. The simplified results are that the best way to display proximity information to a user is to add a false proximity color mapping. Adding stereo, a non-converging stereo camera, does not significantly affect the error rates. The optimal way to present proximity information is to use both stereo and color but the use of color-only is not significantly worse.

7.2 Time Data

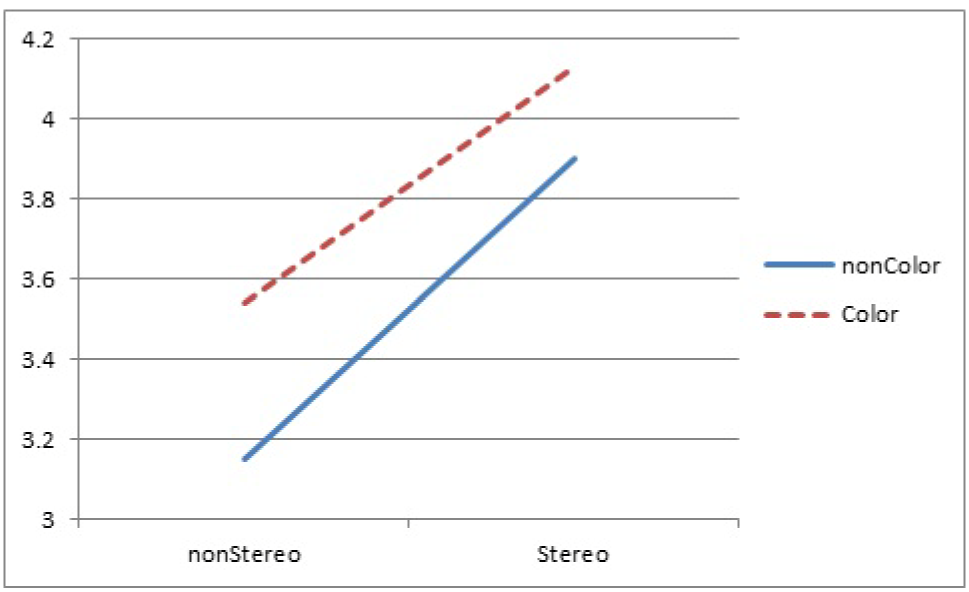

Figure 7.2. Graph of results of Time, graphed data is in Table 7.3. The graph plots the average Time in seconds (Y) by Stereo (X) with the two lines representing Color vs nonColor

Table 7.3. Results of SAS two factor- repeated measure design for time in seconds.

Source	SS	Df	ms	F	p
Stereo	21.33	1	21.33	19.99	<.0001
Color	4.69	1	4.69	2.35	1.1321
StereoXColor	0.333	1	0.333	0.2	0.6544

The analysis run was Repeated Measure-Two factor test on Time taken. There was no significant main effect of the color (p=1.13) nor was there an interaction effect between color mapping and stereo (p=0.65). The interesting result was that there was a significant effect on stereo (p=<.0001) which is the opposite of the main effect of the Error analysis. Please see Table 7.3. Figure 7.2 shows the mean time on the Y access, the X axis is whether stereo was used, the red dotted line is color and blue line is non-color. The results show the use of non-converging stereo cameras significantly increases the judging time, however, the use of false color also increases the time taken but not significantly. If the use of the interface is to have fast reaction time then do not use stereo and possibly don't use false color.

7.3 Analysis

The primary failures of the experimental hypothesis were related to the stereo result. Stereovision is used for proximity judgment; humans are hard wired to use it, therefore the claim that stereovision does nothing to help proximity judgment can not be made. The distinction between normal human stereovision and the stereovision provided in the study is important. First of all, humans use a converging stereo "camera" system. This means that the two "cameras", eyes, converge to one spot when focusing on an object. There are stereovision systems that replicate this action; however they are not common hardware and are not the type of system used in this experiment. The stereo system was of a fixed stereo camera system. This system also loses another important depth cue, motion parallax. Motion parallax refers to the effect that as a "camera" moves, closer objects move seemingly faster than farther objects. This depth cue is constantly used anytime a human's head is moved. Fixed cameras that do not move eliminate this depth cue. The static camera system was used because most commercial systems are fixed stereo camera systems. The experiment shows that the commonly employed use of these systems and display stereovision might not be the best way to use one of these systems. These commercial systems can still be used; they can gather the needed depth information but display it using a false color instead of stereovision.

The observation that color mapping improved performance was expected. The use of color distance mapping is good for relative distance but not absolute depth. This is likely due to the color matching, which could be used in this study's the color method. It is hard for the operator to come up with the perceived hue and translate that to an exact depth. It is the use of perceived hue and possibly not exact hue which distinguishes the two types of tasks.

Although the interaction of color and stereo was not statistically significant, the combination did produce somewhat better performance. If distance information is derived using stereo cameras, then producing a color enhanced stereo display to a user would probably be advisable.

The use of a converging stereo system is probably a dead end. There is technology to track the eye movement and move the cameras to converge as the eyes converge. However, the lag time between tracking the movement of the eyes, sending the movement command to move the cameras then finally returning the new video would probably be too hard on the users.

Before using a false color mapping system some decisions must be made. The big question is what color is to be used for what range. In this experiment the standard color spectrum was used due to the investigator's belief that the spectrum would be a natural and identifiable scale for the test subjects to understand. This is only a belief and is not the only way to set the scale; this was not the focus of this investigation.

Two attributes of color are worth repeating here due to their bearings on color choice; chromostereopsis and eye sensitivity. Chromostereopsis is the effect that two colors shown on the same plane appear to be on different planes. In the case of positive chromostereopsis, red is perceived as closer than blue, so it would make sense to have red be close and blue represent far. If the choice was to have blue be close and red be far, than the farther red objects could be perceived as closer than the blue closer objects. This could lead to a misunderstanding of the mapping or a misreading of the mapped data. Eye sensitivity is also an important attribute to consider. The eye is more sensitive to certain colors and can detect a smaller hue change of the colors. Yellow, and between blue and green, the eye is very sensitive, with yellow being most sensitive. This information should lead to using the yellow hues in a place where greater distinction is needed, such as around a robotic grasper. In this case the grasper is to interact with the environment so proximity judgment around the grasper is very important.

The explanation of the false color mapping proved difficult during the experiment. In the beginning of the study, participants frequently had the false color explanation re-explained to them or they waited until they saw the false color inaction to understand it. This is of concern if this method were to be used commercially. The investigator did stumble on a good explanation towards the end of the study. If the participant was reminded of a heat sensor with red being hot and blue being cold, then told that the false color is the same but with red as near and blue as far, this seemed easier to understand.

The possible reason for the slow down when using stereo could be that the subjects have a slightly harder time detecting proximity of the cubes, causing the subject to take more time to get the two cubes closer. The use of color may be easier to understand, so the subjects are able to get the cubes closer together but not take more time than if not used.

The questionnaires gave some interesting results. One questionnaire was given to the subjects after each of the four methods were run. There were 1384 questionnaires recorded. The questionnaire asked if the subject used size of objects, texture, object movement, perspective, color depth mapping, and stereovision depth cues. The reported most used depth cue was color depth mapping with 1180, followed by size of object with 1150, object movement with 966, perspective with 671, stereovision with 636, and finally the least used was texture with 91. This means that most subjects thought they primarily used color depth mapping and object movement as the main depth cues. The subjects may have checked color depth mapping because it was so easily identifiable as a depth cue. If there was more of a texture on the blocks then the texture depth cue may have been used more. It is interesting that so few people indicated that they used stereovision but the results indicate that there is an effect when stereovision is included.

CHAPTER 8

SUMMATION OF CONTRIBUTIONS AND FUTURE WORK

8.1 Mobile Robots

The main contribution of this research is to the field of robotics, specifically the area of teleoperation. It was observed that teleoperation of robots with arms, without proximity sensing, often causes the robotic arm to miss the object or move the object it is attempting to interact with. Both results suggest that the operators had difficulty judging the distance between the effector and the object it interacts with. The results of this study suggest that a color mapped display will help. If the operator knows what color to look for to be in proper position for the effector then the effector probably will not miss the object or run into it. The effectiveness is increased if the operator just has to match the color of the object to the effector. Matching hues is easier than attempting to judge hues for precise distance measurements.

Moving the robot is also aided by having a sense of relative proximity. With only a 2D robot view, it is hard to judge distances from objects. With the false color interface, the proximity to objects can be used to better avoid or interact with the objects. Driving the robot using the false color interface is fairly easy to understand. The operator can ignore the colors and just drive the robot similar to driving the robot with a grayscale monocamera. If the operator uses the colors then he/she can receive a relative proximity judgment to the objects allowing the operator to avoid them.

8.2 Proximity Judgments

What needs to be stressed about this research is the false color interface was seen to be better for relative proximity judgments and not for absolute precise proximity judgment. The ambiguity of hue differentiation and having to know what each hue directly maps to does not allow this interface to be used as a precise proximity judgment interface. Additional features must be added to allow this. With all of the presentations of proximity, depth cueing can be added that will provide a backup to the primary display of the interface [23]. Applications have incorporated object tracking and cursor location awareness to add precise proximity perception. For object tracking, the operator would select the object to be tracked so the proximity of the center of the object would be continuously displayed. Cursor location awareness allows the computer to know the location of the mouse cursor when over the GUI false color interface, allowing the computer to display the corresponding proximity measurement for the pixel the cursor is over. As the cursor is moved around the interface, then the displayed proximity would change depending on what the cursor was over. The same would occur as the robot moved; the proximity map would change underneath it, but the cursor would remain in the same location.

8.3 Future Work

This study can be taken in different directions. The focus of the stereo cameras was set to the stationary cube, however, an experiment where the focus was on the moving cube would see if there would be a difference in perception. Another experiment could compare fixed vs. converging stereo systems. This experiment would test to see of the stereo results found in this study, was due to an effect of the fixed stereo or for both fixed and converging stereo systems. The final experiment would be to double the subjects and to split the methods, showing half the group the stereo and have the false color. This will test if the subjects were not fully participating during the stereo methods in favor of the color methods due to color methods being easier.

REFERENCES

1. Jen LaPierre, J.L., Steph Egilo, and Darren Van Dam. *Depth Cues*. Available from: http://ahsmail.uwaterloo.ca/kin356/cues/cues.htm.

2. *Depth perception*. 2010; Available from: http://en.wikipedia.org/wiki/Depth_perception#Binocular_cues.

3. Teittinen, M. *Depth Cues in the Human Visual System* Available from: http://www.hitl.washington.edu/scivw/EVE/III.A.1.c.DepthCues.html.

4. Jain, R., Kasturi, R., Schunck, B., *Machine Vision*. 1995: McGraw Hill.

5. Murphy, R.R., *Introduction to AI Robotics*. 2000, Cambridge, Massachusetts: The MIT Press.

6. Meers, S. and K. Ward. *A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation*. in *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*. 2004.

7. Crane, C., et al. *Development of an Integrated Sensor System for Obstacle Detection and Terrain Evaluation for Application to Unmanned Ground Vehicles*. in *SPIE Defense and Security Symposium*. 2005.

8. Craighead, J., B. Day, and R. Murphy, *Evaluation of Canesta's Range Sensor Technology for Urban Search and Rescue and Robot Navigation.*, in *Proceedings of the 2006 International Workshop for Safety, Security, and Rescue Robotics (SSRR 2006).* 2006.

9. Kay, L., *Auditory perception of objects by blind persons, using a bioacoustic high resolution air sonar.* The Journal of the Acoustical Society of America, 2000. 107(6): p. 3266-3275.

10. Takeichi, A., et al. *New 3D display using lens array and depth division images*. 2007. Boston, MA, USA: SPIE - The International Society for Optical Engineering.

11. Bosse, M. and R. Zlot. *Continuous 3D scan-matching with a spinning 2D laser*. in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*. 2009.

12. L. Anat, F.R., Fr, D. Do, and T. F. William, *Image and Depth From a Conventional Camera with a Coded Aperture*, in *ACM SIGGRAPH 2007 papers*. 2007 ACM: San Diego, California.

13. Lawson L.S. Wong, A.S., Andrew Y., *Learning Grasp Strategies with Partial Shape Information,.* AAAI, 2008.

14. Qingxiong, Y., et al. *Spatial-Depth Super Resolution for Range Images*. in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*. 2007.

15. Liang, Z. and W.J. Tam, *Stereoscopic image generation based on depth images for 3D TV.* Broadcasting, IEEE Transactions on, 2005. 51(2): p. 191-199.

16. Ideses, I., et al. *Depth Map Quantization - How Much is Sufficient?* in *3DTV Conference, 2007*. 2007.

17. Ideses, I., L. Yaroslavsky, and B. Fishbain. *Depth Map Manipulation for 3D Visualization*. in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*. 2008.

18. Young-geun, K. and K. Hakil. *Dense 3D map building for autonomous mobile robots*. in *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*. 2003.

19. *Accommodation (eye)*. 2011; Available from: http://en.wikipedia.org/wiki/Accommodation_%28eye%29.

20. Redert, A. *Visualization of arbitrary-shaped 3D scenes on depth-limited 3D displays*. 2004. Thessaloniki, Greece: IEEE Comput. Soc.

21. Jocelyn, F., *Seeing depth in colour: More than just what meets the eyes.* Vision Research, 1994. 34(9): p. 1165-1186.

22. Schubert, E.F., *Light-Emitting Diodes*. Lite Emitting Diodes.

23. Welch, R.a.E., G., *Requirements for Robots Utilized at HAZMAT Incident Sites*, in *Space Programs and Technologies Conference*. 1994: Huntsville, AL.