SINGLE CAMERA 3D GAZE DETERMINATION

A Dissertation

by

JEFFERY LINN BECKMANN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2007

Major Subject: Computer Engineering

SINGLE CAMERA 3D GAZE DETERMINATION

A Dissertation

by

JEFFERY LINN BECKMANN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Richard A. Volz
Committee Members,	Robert J. Hall
	John Leggett
	Jennifer L. Welch
Head of Department,	Valerie E. Taylor

May 2007

Major Subject: Computer Engineering

ABSTRACT

Single Camera 3D Gaze Determination. (May 2007)

Jeffery Linn Beckmann, B.S., Texas A&M University;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Richard A. Volz

In this dissertation, a new approach for determining gaze direction is presented. This approach is based on the existence of a visual axes center for the human eye, the location of which is invariant with respect to the head. The vector from the visual axes center of an eye through the pupil center provides a reliable approximation for a gaze vector. Calibration camera images of human subjects looking at known points on a computer monitor are collected in a non-intrusive manner. Algorithms are applied to the images from two independent cameras whose spatial relationship is known with respect to the monitor. The calibration algorithms allow determination of physical distances between selected facial features visible in the images and the invariant location of the visual axes center for each eye (not visible) with respect to these features. Given these invariant relationships between a subject's facial features and eye visual axes centers, optimization techniques are applied to subsequent images collected from a single camera to obtain the three-dimensional locations of the visible facial features and the visual axes centers, and from these, the gaze direction.

The results of experiments conducted to determine the viability and accuracy of the visual axes center approach in determining the gaze direction are presented. The results show that the approach can provide acceptable gaze direction error values when high accuracy (< 1° angular error) is not required. Techniques to improve accuracy are discussed as well as potential limitations of the approach.

DEDICATION

In memory and on behalf of Tim, I dedicate this dissertation, and submit it to Dr. Way Johnston as "the paper that 'Townley' never quite got around to sending you!"

ACKNOWLEDGMENTS

I would like to acknowledge and thank my wife Judy, for her patience and her support and motivation expressed in her own unique way, and my parents, for giving me life and a solid foundation on which to live it.

I would also like to thank Carl McGrew, for providing me the opportunity of a lifetime in Kwajalein, and Tim Townley, for a periodic reminder of what it means to be an Aggie and that I am not crazy for giving up 'the big bucks' to work on my PhD.

In addition, I would like to thank Todd Beckmann, Clayton Coe, and Wes Gunter for their help, and my brother and sister, nieces, nephews, friends, and other family members who made the process just a little more bearable.

Finally, I would like to acknowledge and thank Dr. Volz and my Committee members, without whose guidance and help this dissertation would not have been possible.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

1. INTRODUCTION

In this dissertation, a novel approach for determining a person's direction of gaze in three dimensions, using images from a single camera is presented. The approach is based on determining the visual axes center of a subject's eye and relating it to other facial features during a one-time calibration session, and then, in any subsequent sessions, using this relation to determine the subject's gaze direction. The algorithms necessary to determine the visual axes center of a human eye are detailed. A technique for determining the three-dimensional (3D) location of a subject's various facial features (nostrils, pupils, etc.) is also presented.

1.1 Motivation

There is a wide range of applications that can benefit from the ability to determine someone's gaze: hands free interfaces for users with disabilities and driver awareness monitoring to name two. Duchowski [1] provides an overview of a broad array of eye tracking applications, many of which can benefit from gaze determination capabilities. These different classes of applications have diverse sets of requirements that have a great impact on the techniques required for a solution. The research in this dissertation is motivated by one particular application that is, itself, representative of a fairly broad class of applications. The planned application is intended to monitor students taking a university-type exam on a computer and then to determine if they are following the restrictions of the exam (closed-book, no notes, etc.). The desired application determines where an examinee is looking and for how long, and then, tries to gauge the examinee's adherence to the exam restrictions.

This particular class of application requires that the activities of the subject not be adversely impacted by the application and any peripheral hardware. In addition, the

This dissertation follows the style of Computer Vision and Image Understanding.

application must be able to operate in a fairly unrestricted, classroom-type environment. It must be economical in that at least some portion of the application system must be replicated for each student taking the exam. Also, the accuracy of determining where the subject is looking need only be determined on a short term average basis, not on an individual input (image) basis since adherence to the exam requirements will be based over a period of time during which several gaze inputs are available. Single, unclustered deviant results are not significant. Finally, the need for high accuracy (less than one degree of error between the reported and actual gaze direction) results is not mandatory, because the visual environment and location of information on the monitor can be adjusted so that larger angular errors still allow for detection of exam violations.

The original research objective of this dissertation was the development of this automated university-exam proctor application. The gaze determination capability required was to be purchased or a known technology was to be implemented. The innovation was to be the development of a model of a student's visual behavior during the conduct of an exam.

Based on a review of many of the published gaze determination systems, and telephone and e-mail contacts with representatives for many of the commercially available systems, it became evident that the monetary cost to procure a reasonably functional gaze determination system that would allow the collection of data with which to develop a proctoring model would be prohibitive. It is estimated that the most inexpensive system that would meet the perceived needs would be $20,000 [2]. For such a gaze determination system to be the basis of a fieldable proctoring application, a significant investment would be required in even the smallest, pseudo-realistic environment.

Because of the cost of commercial systems, an attempt was then made to locate a reasonably mature gaze determination system detailed in the literature that provided enough information for a reasonable likelihood of a successful implementation. One of the conditions established as an indicator for a successful implementation of a published but unavailable system is the availability of the code; either binary or source.

The only system where code was potentially available was for the system designed by Bakic and Stockman [3]. Correspondence with Dr. Stockman did not produce the code for the Bakic system [3], however, it did result in the acquisition of a copy of Dr. Bakic's dissertation from Michigan State University [4] and a compilation of C++ code modules from an attempt to migrate Dr. Bakic's system from a Unix-based system to a Windows-based system. Unfortunately, after experiencing reasonable success at cleaning and modifying the code to locate facial features (head, pupils, eye corners, nostrils, and mouth corners) in 320 x 240 webcam images of subjects, it became clear that it would provide no assistance in developing the capability to determine 3D gaze locations as needed for the proctoring application.

Based on this setback, the research focus was changed. Instead of an attempt to implement an exam proctoring application, the focus was changed to be on the development of a comparatively inexpensive technique to determine 3D gaze: still suitable of course for the exam proctoring application.

1.2 Vision and Gaze

Humans are visually-based creatures. We use our visual capabilities to safely walk across a crowded room, read a newspaper, or check the weather outside. Computers, on the other hand, have historically been very limited in their ability to make use of visual information. Developing the capability for computer systems to interpret their surroundings visually in a more human-like manner, will prove to be a significant step in creating a more ubiquitous computing environment. Creating the ability for a computer to interpret a user's facial expressions or determine where a user is looking would open up a whole new realm of possibilities for human-computer interaction.

1.2.1 Human Vision

In the simplest of terms, the human visual system is made up of the eyes, interconnecting nerves, and the brain. The eyes (see Fig. 1) collect and focus light from the environment and convert it into signals in a form that can be interpreted by the brain, the optic nerves carry these converted signals to the brain, and the brain interprets the signals and acts upon them accordingly.

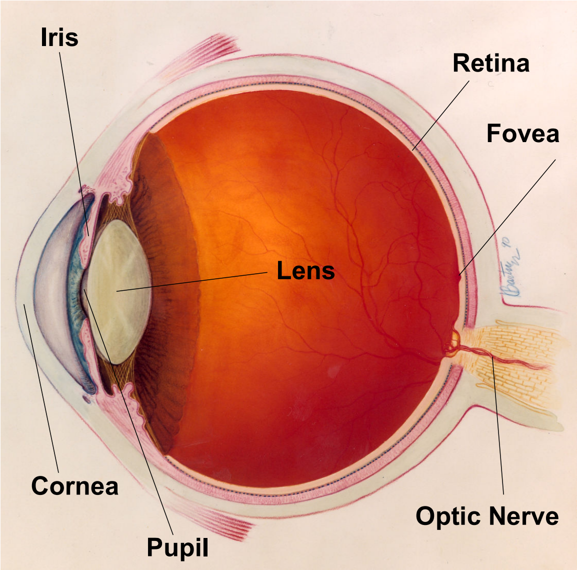


Fig. 1 Eye schematic (right eye, top view: modified) [5].

In order to create an image, light is reflected off, or transmitted from, an object and into the eye. Some of the light entering the eye (see Fig. 2) is refracted and focused as it passes through the cornea and the lens of the eye. The refraction/focusing process results in an inversion of the image the light represents. The 'inverted' light is focused on the retina where it activates nerve cells called rods and cones [6]. The signals from the monochromatic rods and the color-sensitive cones are then transmitted collectively down the optic nerves to the brain which effectively inverts the image and perceives an image of the original object.

If a clear, steady image is desired for an activity such as reading, the light coming into the eye must be focused onto a very small region (1.5 millimeter diameter creating an approximately 5.2 degree field of view) [7] of the retina called the fovea, made up of predominantly cones. The cone cells are capable of producing a high quality, color image signal to the brain. Because of the fovea's small size, the eye must be fairly still and have little movement with respect to the object being looked at in order for light to be focused on the fovea. This period of having a stable, detailed image and relatively no eye movement called a fixation is contrasted by periods of rapid eye movement called saccades. Saccades involve incoming light entering the eye and striking the retina: not necessarily in the foveal area. Periods of saccades are useful when motion detection is important and detail, especially color detail, is not. Activities such as traveling down a busy sidewalk would benefit from saccadic eye motion. For purposes of computer-based gaze determination, a fixation lasts approximately 350 milliseconds [8, 9].

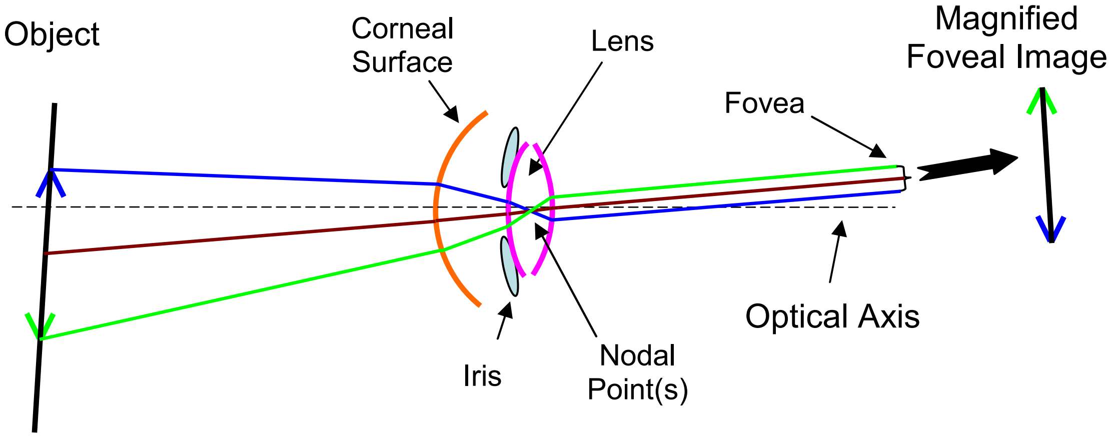


Fig. 2 Ocular focusing/inversion of light (right eye, top view).

The visual system of a computer can be thought of in terms similar to that of a human except a digital camera replaces the eye, some wiring or a bus replaces the optic nerve, and the computer and its software replaces the brain. The camera performs a function similar to that of the human eye in that it focuses light and converts it into a usable form that in this case, a computer, can interpret. However, the digital receptors in a camera are more similar to the cones of the fovea in that a period resembling an eye fixation is required in order to obtain a usable image.

While the organization and format of the foveal signals the brain receives are unknown, current digital technology results in an image being represented on a computer as a collection of distinct, individual units called pixels [7]. The pixels correspond to the individual sensor elements or phototransistors of the camera. For a given camera, each pixel is the same size. Each pixel is unique and, in general, its attributes are not dependant on neighboring pixels. The collection of pixels representing an image is organized in terms of rows and columns corresponding to the camera's phototransistor array arrangement. All of the information obtainable by the computer from an image is based on the computer's ability to manipulate (process) and interpret (analyze) the pixels [8]. A single image frame can provide a significant amount of information if the computer is able to process and analyze it correctly.

1.2.2 Gaze

In the vernacular of computer vision researchers, determining where someone is looking is usually referred to as gaze tracking [9]. However, Wang et al. [10] allude to a more appropriate terminology for the process of determining where someone is looking: gaze determination. 'Gaze tracking' implies a path or sequence of 'gaze' steps. 'Gaze determination' more appropriately describes the overriding challenge of determining where someone is looking at any single instant in time. Regardless of the nomenclature, significant effort is currently being expended to develop and improve the ability of computers to determine where someone is looking.

As a verb, gaze is often defined as the act of looking at something [11]. In terms of eye movement, gaze represents a period of fixation with little or no eye movement [12]. Gaze can also be thought of as the path reflected light would take from the object being viewed to the viewer. Using this notion allows gaze to be conveniently described or represented by a line or a vector.

In order to represent gaze as a line, one must be able to define the two ends of the line. The focus of much of the work described later in this dissertation is on the location of a point on a gaze line within the subject’s eye. The other end point of the gaze line is the point at which the subject is looking, often referred to as either the fixation point [13] or the point of regard [9]. The terminology fixation point will be used herein unless specifically stated otherwise to match the cited literature.

Unfortunately, the fixation point is usually an unknown. However, it is usually possible to initially determine a direction that someone is looking instead of a location. Because of the inherent directionality of a vector, it is more common and convenient to represent the notion of gaze as a vector rather than a line. This representative vector is commonly referred to as a gaze vector. With a direction, a location can subsequently be determined if the plane of intersection, such as a computer monitor, or the distance from the subject is known. This location can also be established by determining the intersection point of the gaze vectors from each eye. However, other effects such as the influence of eye dominance [14, 15] make this technique more difficult.

Various definitions of vectors exist that could serve as a gaze vector. All have the property of originating at some defined point on the human subject and extending to the fixation point. While the gaze vector origination point could be any point on the subject, in order to be representative, the originating location must be relatable to the eye. Therefore, the origination point of a gaze vector is usually a feature associated with the subject's eye, often making it lie on one of many of the eye's reference axes (the axes used to describe the many optical paths and relationships of the human visual system) [16]. By convention, the direction of a gaze vector is always pointed away from the subject and toward the object being looked at: opposite the direction the light travels when reflecting off of the object into the subject's eye.

There are several reference axes in the literature that are closely related to a gaze vector or might be a candidate for a gaze vector. Unfortunately, the terminology used in the literature is not entirely standardized, with different authors using different terminology, or in some cases, different definitions for the same axis. Using the terminology of the author who defined them, these axes include the following:

a. Morimoto [9]
    i. Line of gaze or optical axis
    ii. Line of sight
b. Carpenter [13]
    i. Fixation line
    ii. Optical axis
    iii. Visual axis
    iv. Pupillary axis
    v. Line of sight
c. Blaine [17]
    i. Pupillary axis or achromatic axis
    ii. Visual axis
d. Thibos [18]
    i. Visual axis
    ii. Achromatic axis

In the following paragraphs, each of these axes is discussed in terms of its usefulness as a gaze vector, and the differences among the various definitions are presented. In order to facilitate the discussion, Fig. 3 depicts each of the above lines graphically in the context of the fixation point and relevant eye features of a right eye viewed from the top of the head. The angles shown are not representative of their actual values. The angle differences have been increased to make it easier to distinguish between lines. It is important to realize that most of the definitions that follow define straight lines that can be related to what a subject sees, but are not lines that exactly represent the path of light waves through the eye. This is necessarily so because of refraction that occurs to the light waves as they pass through the eye to the retina. Thus, one should not necessarily

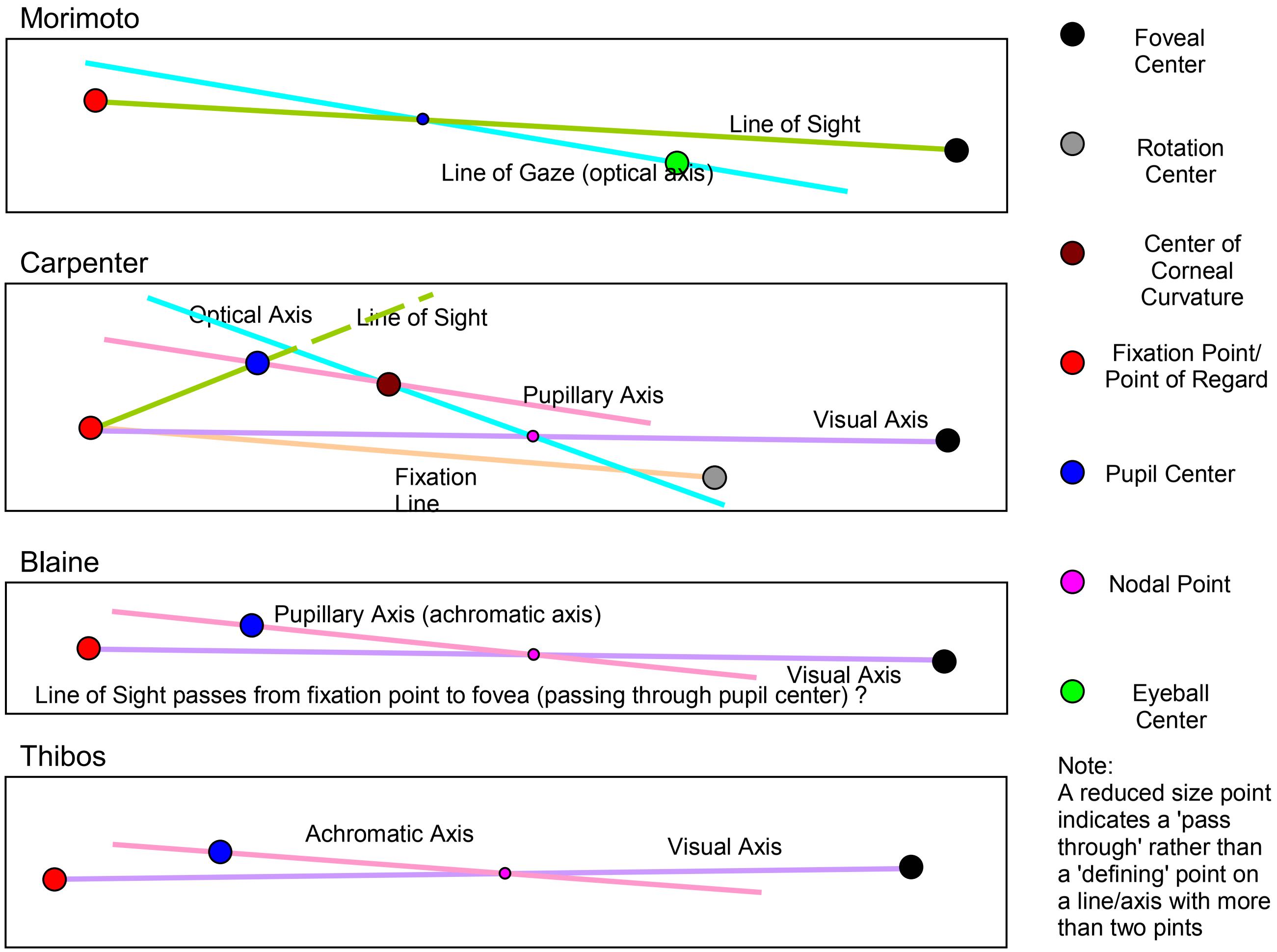

Morimoto
Line of Sight
Line of Gaze (optical axis)
Carpenter
Optical Axis
Line of Sight
Pupillary Axis
Visual Axis
Fixation
Line
Blaine
Pupillary Axis (achromatic axis)
Visual Axis
Line of Sight passes from fixation point to fovea (passing through pupil center) ?
Thibos
Achromatic Axis
Visual Axis
● Foveal Center
● Rotation Center
● Center of Corneal Curvature
● Fixation Point/ Point of Regard
● Pupil Center
● Nodal Point
● Eyeball Center
Note:
A reduced size point indicates a 'pass through' rather than a 'defining' point on a line/axis with more than two pints

Fig. 3 Eye feature axes (right eye, top view).

try to superimpose the notion of light paths on the various lines that are defined in the literature.

The presentation of these axes in the literature (and in Fig. 3) is from the perspective of the top of the subject/eye. The relationship of the various features (pupil center, nodal point, rotation center, etc.) projected onto a vertical plane (as observed from a side or temple/nasal view) is neglected. Bradley, however, has made a statement indicating that the lines all lie in virtually the same horizontal plane [16], and hence there is little useful to be gained by considering a vertical plane. As a result, the remainder of this dissertation assumes that these internal features of the eye lie in the same horizontal plane and all displacements occur in this horizontal plane.Morimoto [9] defines a vector called the line of gaze (LoG). He doesn't use this line in his method, but appears to include it to dispel the intuitive notion that the LoG represents a line passing through the fixation point. He further states that the line of gaze is collinear with the optical axis that originates at the center of the eyeball and passes through the center of the pupil. Unfortunately, Morimoto's definition of optical axis differs from that normally used, e.g. see Carpenter [13]. To understand Carpenter's definition of the optical axis, one must first consider the notion of the nodal points.

In general, nodal points are conceptual features most often depicted as being located inside or behind the lens of the eye. The nodal points have the property that a line from the posterior nodal point parallel to the line from an object point to the anterior nodal point will strike the image plane at the same point as the actual paraxial light ray defracted through the lens (see Fig. 4).

The axis connecting the two nodal points is widely accepted as the optical axis, though this differs from Morimoto's use of the term. Carpenter also indicates that this optical axis passes through the center of corneal curvature of the eye. The center of corneal curavture is defined as the point a radial distance *r* from all points on the surface of the cornea. However, because Carpenter's optical axis does not necessarily intersect the fixation point, it cannot be considered a gaze vector. Carpenter's book is a widely referenced authority on the eye and the various relevant concepts that are used for a variety of purposes. Consequently, when his definitions of the various lines differ from others, this dissertation will use Carpenter's definition.

Because the two nodal points are so close together they are often approximated by a single point: usually by a point between the two. Unless otherwise indicated, reference to a 'nodal point' in the remainder of this dissertation will be in reference to the single point approximation. This allows one to construct a straight line from the object point to the image point; this is what Carpenter calls the visual axis.

Morimoto also defines another vector, the line of sight (LoS), which originates at the foveal center and passes through the pupil center. Morimoto assumes that the LoS also passes through the fixation point and can thus be used to determine the gaze vector. Carpenter also defines a line of sight. Unlike Morimoto's, Carpenter's version of the LoS passes through the pupil center and the fixation point, but does not necessarily intersect the foveal center. Blaine [17] states that the fixation point, the pupil center and the fovea center all lie on the LoS, a contradiction with Carpenter, though the difference is small.

Blaine [17] and Thibos et al. [18] define an axis they call the achromatic axis. Blaine also calls his achromatic axis the pupillary axis. However, Carpenter again differs from Blaine in that he defines the pupillary axis as the line connecting the center of corneal curvature and the pupil center. Regardless of names, one of the lines that will be of use in the work described herein is the line between the nodal point and the pupil center.

Carpenter also defines a vector which originates at a notional center of eye rotation and intersects the fixation point and is called the fixation line. The originating point of this gaze vector is called 'notational' in that it is an average rotational center because the center of rotation is not fixed [19]. This rotational center appears to be closely related to Morimoto's line of gaze eyeball center, but the relationship is not made clear by either author. Since the rotation center is only notional and not actually fixed, it is not used further.

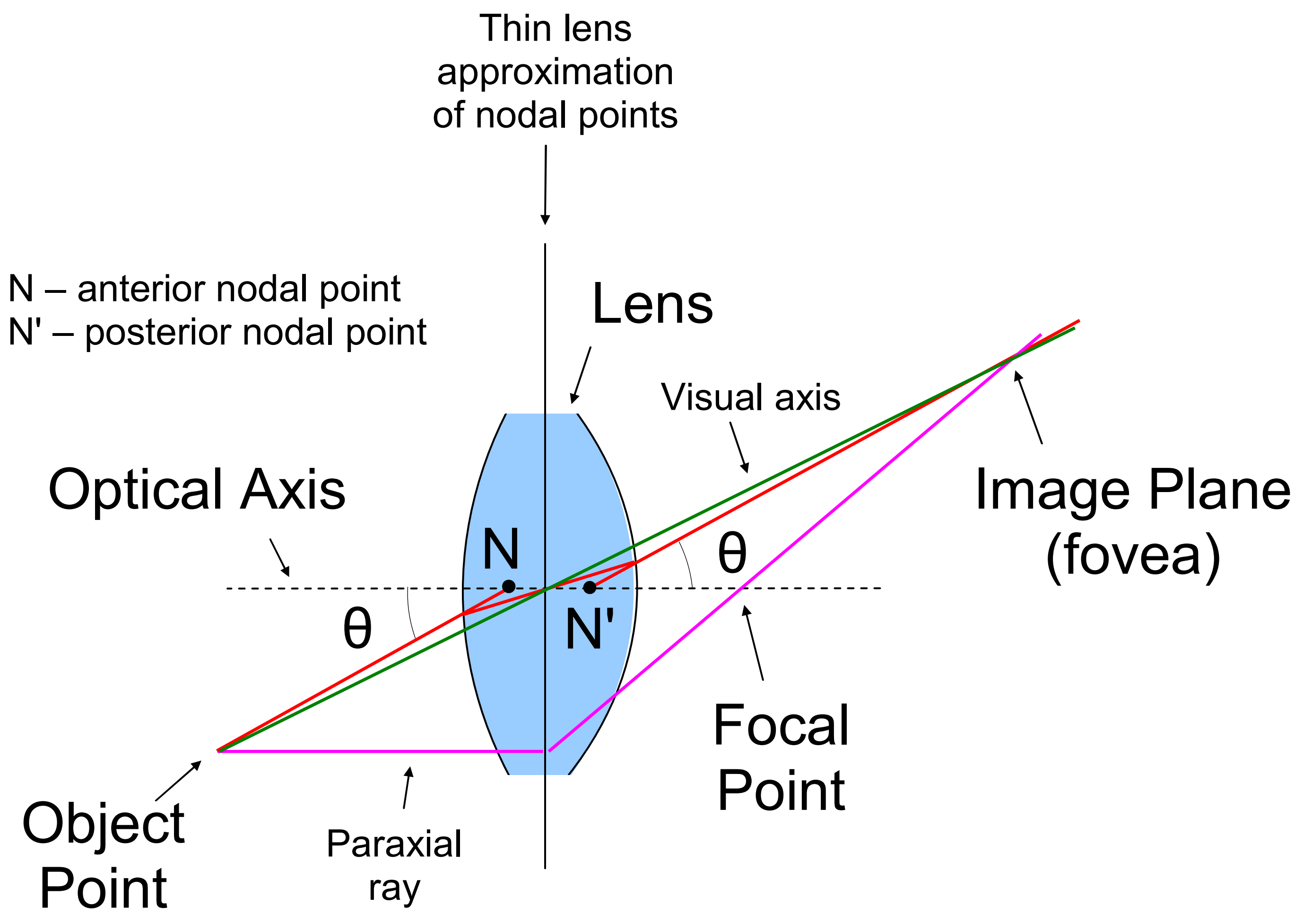


Fig. 4 Nodal points (modified) [20].

Another potential gaze vector that also uses the idea of 'nodal points' is referred to by Carpenter as the visual axis[13]. If one assumes two nodal points, Carpenter discusses the visual axis in terms of two parallel straight lines: the first line passing through the anterior nodal point and the fixation point and the second line passing through the posterior nodal point and a point on the fovea. However, Carpenter (as well as Blaine and Thibos) suggests that the nodal points can be assumed to be identical. His visual axis then becomes a line from the fixation point to a point on the fovea. This line also passes through the 'merged' nodal points.

One tends to think of gaze in terms of looking at a single point in space. However, for a given fixed position of the head and eye, there is actually a region in space the image of whose points hit the fovea. One can think then, not of one, but an infinite set of visual axes that instantaneously pass through the nodal point. Carpenter states that these visual axes are bounded by tangents to a sphere with fixed radius whose center is fixed in space with respect to the head. He refers to this center point as the visual axes center. This visual axes center is fixed with respect to the head regardless of eye position and what the subject is looking at, i.e., the object points whose corresponding image points are on the fovea (see Fig. 5).

Fig. 5 is a 2D representation of a subject looking at three spatially unique objects at three different instances in time. In the figure, the key features at the three different instances in time are superimposed on each other. Because the subject's head is fixed, only eye movement is observed and a single tangent sphere is formed. The tangent bounds of the sphere are determined primarily by the foveal edges, but are also affected by the refractive properties of the eye; including the cornea and the lens. Since there is no convenient way to determine the bounds of the portion of the object whose image is at the boundary of the fovea (i.e., the sphere itself cannot be determined), it is assumed, that for practical purposes, the line from the point of fixation through the nodal point actually passes through the center of the sphere, the visual axes center. A vector from this visual axes center along any of the visual axes would correctly represent a gaze vector.

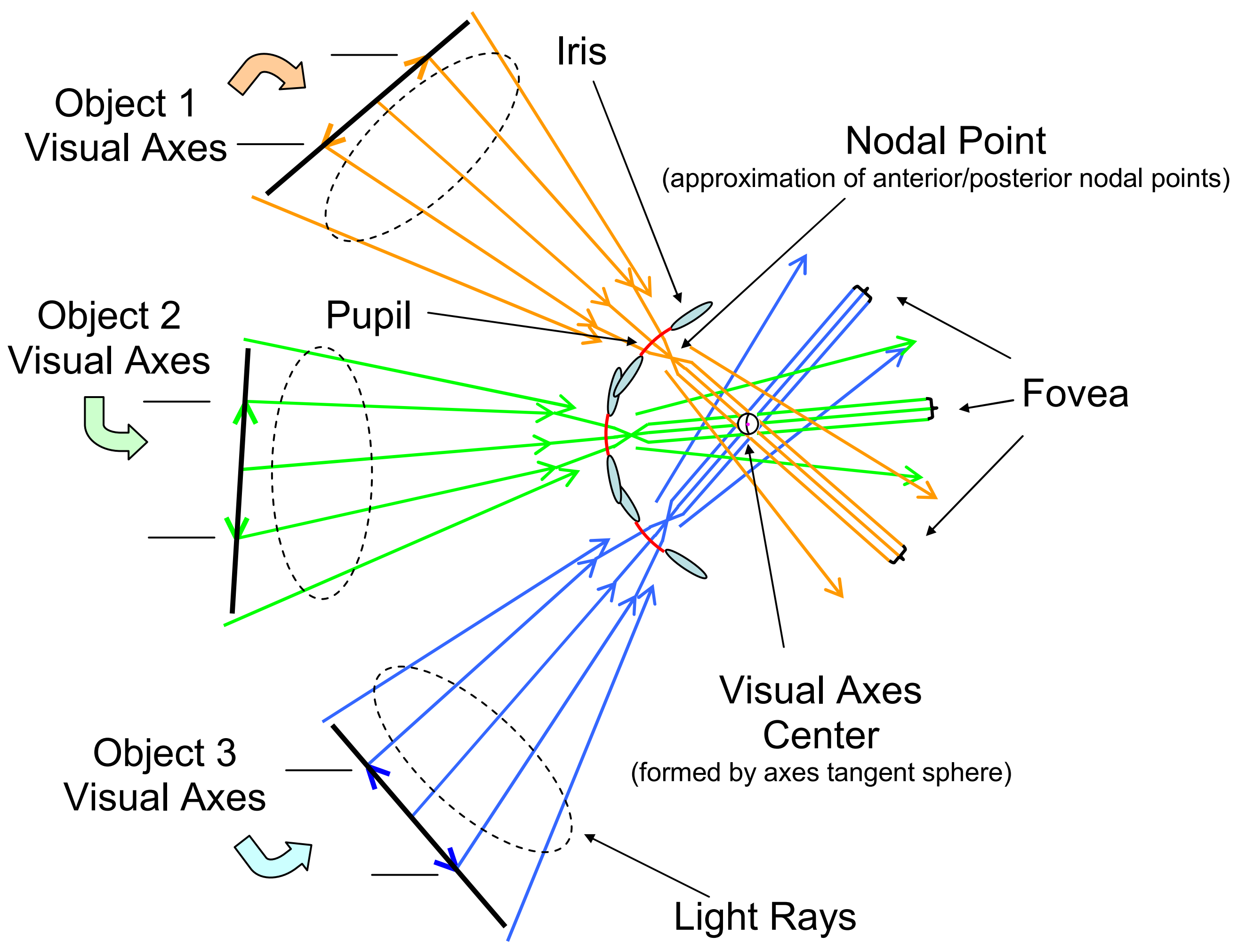


Fig. 5 Visual axes sphere.

Because the visual axes center remains fixed with respect to the head, once initially located with respect to the head of a given subject, the center is available whenever the location of the head can be determined. Locating the head can be directly accomplished using image processing. However, because neither the visual axes center nor the nodal point can be directly determined from image processing, none of the visual axes (candidate gaze vectors) can be determined unless these features can be related to or derived from more readily available features.

It is useful to also consider the quantitative relationships among certain of the defined axes. Carpenter [13] states that the angle between the optical axis and the visual axis can be as large as seven degrees and Martin [21] states it can be as large as 17 degrees with the visual axis offset toward the nose. According to Park [22], the pupillary axis and the line of sight vary by an angle that they called the physiological angle. Unfortunately, this angle varies considerably from subject to subject. It also has been reported to vary with time for the same subject. However, according to Thibos [18], the angle between the visual axis and the achromatic axis is approximately two degrees with the visual axis offset toward the nose. Even though Blaine differs in terms of the definition of the achromatic axis, he specifies a similar angle between the visual axis and the achromatic (pupillary) axis: the visual axis is less than three degrees toward the nose from the pupillary axis.

The method presented in this dissertation is based on developing an approximation to the visual axis that uses the pupil center as an approximation to the nodal point. Blaine's relationship between the visual and pupillary axes is used to bound the error resulting from this approximation.

1.3 Research Contribution

While there has been a considerable amount of research into the development of gaze determination technologies, these efforts have not produced systems capable and affordable enough to support everyday, real-world applications. The research presented in this dissertation provides several contributions that serve as preliminary steps toward the development of more affordable gaze determination systems. The primary achievement of this research is with respect to developing a methodology to determine 3D gaze locations without the continuous use of stereo cameras or the need for specialized illumination such as that needed to obtain consistent corneal reflections. Of equal importance, is developing the capability to estimate notional eye features such as the visual axes centers in relation to visible facial features and leverage this relationship on a subject by subject basis to provide gaze direction/location information. Finally, demonstrating the ability to simulate stereo image collection for calibration using commercial, off-the shelf (COTS) webcams and software, and the use of a single webcam for subsequent-usage image collection for 3D gaze determination provides encouragement for others trying to develop more economical gaze determination systems.

1.4 Dissertation Roadmap

Section 2 of this dissertation presents a review of the current technologies capable of providing gaze determination in comparison to the general requirements of the exam monitoring application, as well as providing background information needed in the remainder of the dissertation. A new gaze determination method, based on the visual axes center [13] of the human eye, is presented in Section 3 that more adequately meets the exam application requirements. Section 4 provides details of a technique to determine 3D locations of objects from images collected using a single camera. While not specific to gaze determination, this method will allow the 3D gaze determination from a single camera. Section 5 describes the conduct of a set of experiments designed to implement the methods described in Sections 3 and 4. Section 6 describes the image processing necessary to obtain the data from the experiments. Section 7 presents the pertinent data resulting from the experiments and an analysis of that data as it pertains to the gaze determination performance of the methods being examined. Finally, conclusions are presented and proposed efforts to enhance the new gaze determination method are discussed in Section 8.

2. BACKGROUND

In this section, several aspects of current gaze determination technology are discussed as they relate to the exam proctoring application. General requirements for gaze determination systems are presented, as well as representative methodologies for implementing gaze aware applications. Specific topics applicable to the new method to be presented in this dissertation are also delineated.

2.1 Gaze Determination System Features

Despite the wide range of applications that can benefit from gaze determination capability, there are several features that have been presented in the literature as being required of any gaze determination system. According to Scott and Findlay [23] and Hallett [24], an ideal image-based, gaze tracking or determination device must:

a. offer an unobstructed field of view with good access to the face and head,

b. make no contact with the subject,

c. meet the practical challenge of being capable of artificially stabilizing the retinal image if necessary

d. possess an accuracy of at least one percent or a few minutes of arc,

e. offer a resolution of 1 minute of arc sec^{-1}, and thus be capable of detecting the smallest changes in eye position,

f. Offer a wide dynamic range of 1 minute to 45° for eye position and 1 minute arc sec^{-1} to 800 sec^{-1} for eye velocity,

g. offer good temporal dynamics and speed of response,

h. possess a real-time response,

i. measure all 3 degrees of angular rotation and be insensitive to ocular translation,

j. be easily extended to binocular recording,

k. be compatible with head and body recordings, and

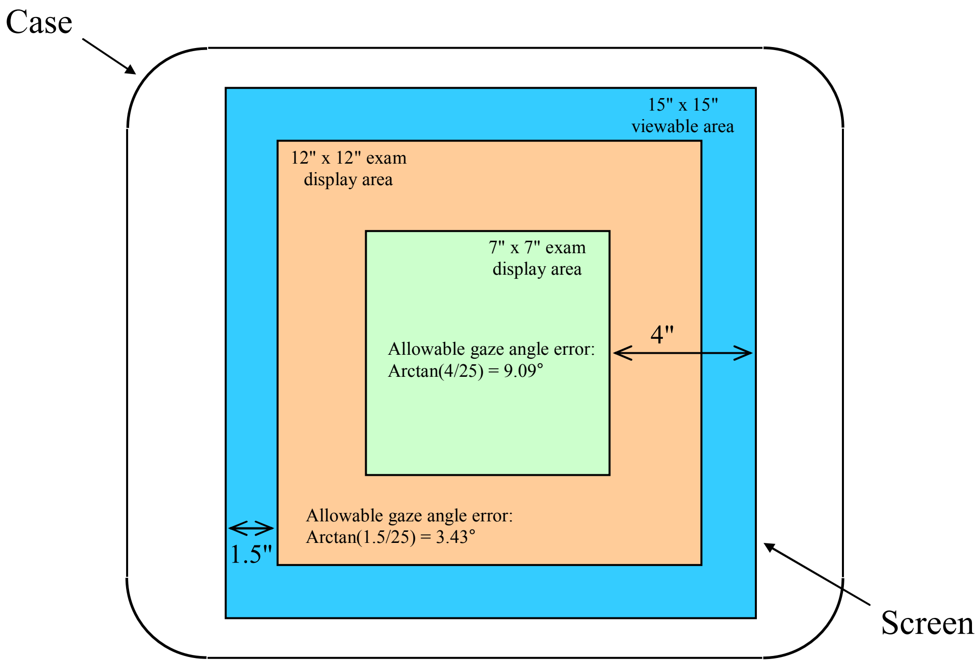
l. be easy to use on a variety of subjects.

However, many of these items are not considered mandatory, or appropriate, for the class of applications represented by the proposed proctoring application. In fact, adherence to many of these requirements could needlessly increase the cost/complexity of the gaze determination system without ensuring any additional useful capability. Therefore, the following (Subsections 2.1.1 and 2.1.2) attempts to segregate those items presented by Scott and Findlay [23] and Hallett [24] as being either mandatorially or optionally required based on the needs of the desired proctoring application. A brief discussion of the rationale leading to the segregation is also presented.

2.1.1 Mandatory Gaze System Capabilities

In support of the proctoring application, only items a, b, j, k, and l from Scott and Findlay [23] and Hallett's [24] list are considered mandatory for the gaze determination device. Systems that are image-based must maintain a view of the eyes (item a) in order to determine gaze. If the system routinely creates obstructions between the camera and the face or head of the user, the eyes will most likely be occluded as well. The performance of an examinee must not be affected by the operation of the proctoring system (item b). Physical contact with the user would create a distraction and affect the examinee's performance. In addition, because the human vision system is binocular in nature, it is important to facilitate the acquisition of binocular gaze information (item j). The ability to isolate the eyes in images is fundamental. The capability to do so must exist regardless of what other artifacts exist in the image (item k). Finally, the proctoring application must be flexible enough to accommodate a wide cross-section of students, and therefore, the gaze determination portion must be similarly flexible (item l).

The notion of accuracy (item d) is also required, but the criterion for acceptability is modified. It is assumed that the proctoring application would present only one question at a time. From experience, the entire question and response mechanism (answer choices, submit buttons, etc.) can be appropriately presented in a 7" x 7" area centered on a computer monitor with a viewing area of 15" x 15". If it is assumed that the user is viewing from approximately 25" away, then even a gaze direction error of nine degrees ensures that the ability to determine whether an examinee is looking at the question (viewable area) or not is maintained (see Fig. 6). If the presentation area is expanded to 12" x 12", a gaze direction error of three degrees still allows acceptable proctoring ability. Therefore, a somewhat arbitrary accuracy goal of maintaining an average gaze direction error (angle between the reported and actual gaze direction) of three degrees or less is proposed.



Objective: detect consistent gaze outside of viewable area with user 25" away from monitor

Fig. 6 Appropriate monitor viewing areas.

The issue of angular rotation and translation of the eye (item i) is not a driving requirement, but is implicit in being able to determine 3D gaze direction. If the gaze device is adversely affected by or does not account for all possible eye movements, it will be unable to maintain its required accuracy specification.

In addition, the requirements to not require special illumination and to use only a single camera (after calibration) for image collection are added to the list as mandatory, as is the rather subjective requirement for the device to be inexpensive. For purposes of supporting the proctoring application, an arbitrary hardware cost of $250 or less is considered inexpensive.

The following subsection (Subsection 2.1.2) discusses those list items (c, e, f, g, and h) that are not considered mandatory for the proctoring application.

2.1.2 Optional Gaze System Capabilities

Several of the ideal features listed by Scott and Findlay [23] and Hallett [24] are considered optional for the exam proctoring application. Item c is by definition optional ('if necessary'). In addition, 'artificially stabilizing the retinal image' implies restricting the movement of the image being viewed with respect to the viewer. Restricting movement seems to violate the practical intent of item b. Item e , in addition to item d, also addresses the accuracy of the system. However, item e appears to be more applicable to systems designed to study eye movement rather than gaze direction. Because the proctoring application is not interested in eye movement per se, the item e capability is optional. A similar assessment is applied to the requirements of item f that relate to eye movement. Finally, item g and h, and a portion of item f, address the issue of timing; particularly, the response time. This notion again applies more to eye movement studies. Although many applications may require gaze determination in real-time, the need to process high video frame rates (15 to 30 frames per second) will not ordinarily be required for gaze determination systems. Informal observations of students taking exams indicate that collecting and processing images at rate of two or three frames a second would be sufficient for the proctoring application.

2.2 Invasive vs. Non-invasive Methodologies

There are many systems discussed in the literature and/or that are commercially available that are capable or claim to be capable of providing gaze determination [2, 3, 25-33]. Many of these are discussed in a database of commercially available eye movement systems maintained by the Applied Vision Research Unit at the University of Derby [34]. Though not all of these systems are used for gaze determination, they do provide a good indication of the technologies available for studying movements of the human eye, of which the study of gaze determination methods is a subset.

Gaze determination systems are often grouped into one of two categories based on their physical interface with the subject. Invasive systems are those that require some physical contact with the user, while non-invasive systems do not. Invasive systems are sometimes referred to as intrusive [9] systems, and non-invasive systems are also referred to as non-contact [35] or remote [9] systems .

Although not used in common practice for gaze determination, the scleral search coil [36] as used by Robinson [37] is a dramatic example of an invasive technology. Small electrical wires are embedded in a device similar to a soft contact lens that the user wears. A surrounding magnetic field is used to detect eye movement of the subject while determining the 3D position of the eye. A similar technique employed by Kaufman [38] called an electro-oculogram (EOG) involves the use small electrodes to record eye movements. Small electrodes are placed on the skin around the eye. Small differences in skin potential caused by eye movements are detected. A more common invasive system used for gaze determination is a head mounted camera system similar to the one used on the iView X HED system [28]. Though often providing very accurate results, invasive systems tend to be more useful in a controlled, laboratory-type environment and not in real-world applications where the direct contact with the subject could affect the results.

Non-invasive, gaze determination systems tend to rely on the acquisition and processing of images that can be collected remotely and without subject contact to obtain their results. Other than the initial distraction based on the visibility of the image collection system, the modification of a subject's actions resulting from the use of a non-invasive system is usually minimal. While Tan, et al. [32] categorizes such non-invasive eye tracking categories as model-based, neural network-based, and appearance-based systems, these categorizations tend to indicate more about the process of identifying the various facial features than the invasiveness of the image collection or the process of tracking the eyes.

Model-based systems attempt to describe a model for some facial feature or portion of the face, and use this model to process the subject images. Such is the case with Daugman [39] who, in a user identification application, models the iris and the pupil as a circle with a specified variation around the contour. The model algorithm performs a course-to-fine search for a circular contour corresponding to the limbus (the border between the cornea or iris and the sclera). Once found, a similar localized search for the iris/pupil contour is then performed providing an image location for the pupil.. A more elaborate model is specified by Yuille et al.[40], who models the limbus as a circle, the eyelids as two parabolic sections, and the two visible portions of the sclera beside the iris and between the eyelids (above and below the iris) as two points or centroids. Gradient descent is then used to fit or deform the model template to images of the eye to locate the eye in the image and thus the pupil.

As the name implies, neural network systems such as the one presented by Baluja and Pomerleau [26], rely on artificial neural networks to interpret the images and provide location information. Both for training and usage, the image of a single eye (in their case the right eye) is extracted from the larger grayscale image and used as the input to the neural network. The training data consists of 2000 images of a user looking at known x, y points on a computer monitor. While the details of the neural network and its method for interpreting the images was not fully explained in the paper, the training apparently allows subsequent images to be interpreted and the user's fixation point on the monitor to be determined. When only limited head movement is allowed, gaze direction angle errors of approximately 1.5 degrees are obtainable [26]. However, it appears that stationary (IR) illumination is required to create specular reflections on the corneas and the training is required on a per subject basis.

Appearance- or view-based systems use techniques similar to those described by Murase and Nayar [41] to estimate pose in order to estimate gaze direction. Instead of using features like eyes to estimate gaze, a large set of images of a user looking at an object are collected with the user at varying positions with varying lighting conditions. Although not explicitly detailed, the head would be the mostly likely object modeled to provide pose from image appearance. In their experiment, Murase and Nayar [41] utilize an inanimate object, a motorized turntable to vary pose, and a robotic manipulator to vary the illumination. The images they collect are combined into a high-dimensional space called an appearance manifold. For a given image of the inanimate object, its pose parameters (gaze direction based strictly on pose) can be estimated by finding the closest point in the appearance manifold. Given that the manifold is not continuous, the resulting pose is only as accurate as how densely the manifold is populated or sampled. In addition, since only pose is determined, only a gaze estimation is determined.

However, Tan [32] reports an actual (not an estimate) average gaze angle error of 0.38 degrees with an appearance manifold derived from only 252 images and no restrictions being placed on user movement. It appears that this high degree of accuracy is as a result of the test images being part of the 252 images used to derive the training manifold. Unfortunately, the methodology is not described well enough to evaluate its usefulness or validity. The methodology also requires the use of IR illumination.

In addition to neural networks, Zhu [42] adds analytic approaches to his categorizations of non-invasive gaze determination systems. Analytic approaches are those that rely on the detection and location of facial or image features in some coordinate space to facilitate determining gaze direction. Analytical and model-based systems tend to share many similarities.

While the gaze determination method to be presented in this dissertation (see Section 3) falls into the non-invasive category and requires images, it does not directly address the issue or mechanism of feature finding; the basis on which many of the non-invasive categorizations are derived. However, the analytic approach most closely represents the principals on which this dissertation's method is based in that the method requires the locations of image features to be available in some coordinate space. The following subsections (Subsections 2.3, 2.4, and 2.5) describe techniques by which many analytical systems identify features (pupil centers, corneal reflections, etc.) in images and locate these features in a coordinate space other than that of the camera.

2.3 Stereo Images

In order to obtain image feature locations in 3D, some mechanism for converting the two-dimensional (2D) pixel locations to 3D is required. The most common method to obtain 3D locations from image features is to use stereo image processing. Most of the non-invasive, image-based gaze determination systems that provide 3D gaze results rely on stereo images to obtain the results in 3D. A stereo image is a pair of images that are taken of the same object at the same instant in time by a pair of cameras in different spatial locations. If the spatial relationship between the cameras is known or can be determined, stereo triangulation [43] techniques can be employed to determine the 3D location of any objects visible in both of the images. A technique approximating stereo image collection ('pseudo' stereo) will be used during hardware and subject calibration for this dissertation (see Subsection 6.4). Section 4 discuses a method intended to eliminate the need for stereo (or 'pseudo' stereo) images after calibration.

2.4 Illumination

Regardless of whether stereo images are used, all image-based gaze determination systems require that certain features be detected and located in the images. A significant number of the image-based, non-invasive analytical gaze determination systems rely on special or controlled illumination to accomplish feature detection. Many depend on the illumination of one or both of the subject's eyes with near-infrared (IR) light, usually with a wavelength around 880 nanometers [9]. Use of the 880 nanometer near-infrared source has the advantages of being virtually undetectable by the user and creating a reflection off of the retina that appears in images as a very well-defined bright spot inside the iris where a dark spot representing the pupil would normally be seen. This bright spot allows for easier pupil/pupil center location during image processing. Virtually all gaze determination systems require determination of a pupil center as part of the gaze determination process, so the use of IR illumination would benefit almost any image-based system.

Many gaze determinations systems, however, have a more fundamental requirement for illumination. Systems based on a technique known as the pupil center/corneal reflection (pccr) method first presented by Mason in 1969 [44, 45] mandate the use of specialized illumination. PCCR methods are based on the reflective properties of the cornea and rely on the use of IR light with a fixed location relative to the camera to produce reflections off of the surfaces of the cornea that appear in the images as bright spots on the iris. These reflections in the image are known as Purkinje images. Fig. 7 depicts the four Purkinje images.

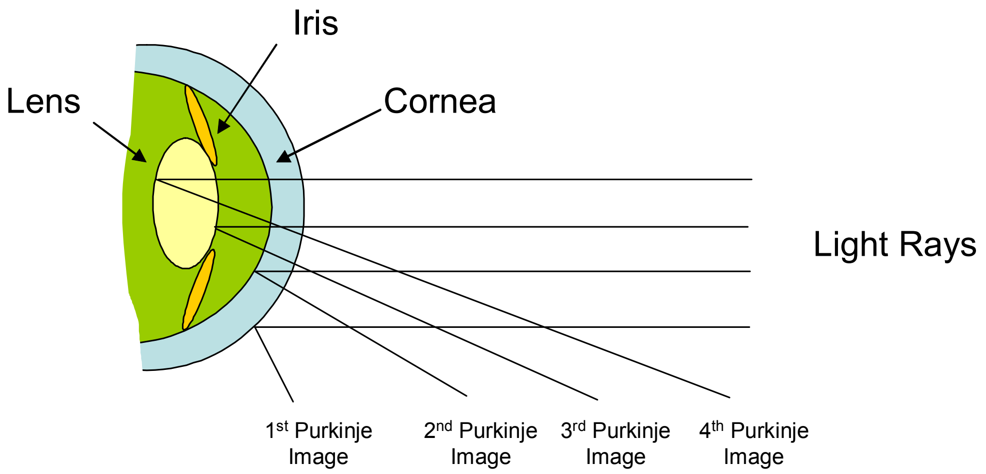


Fig. 7 Purkinje images [12].

The Purkinje images can be related during calibration to a fixation point and pupil center as a function of the curvature of the reflective surface of the cornea and the location/distance relative to the pupil center. These subject-specific calibrated parameters can then be used with subsequent images to determine a gaze direction.

Many pccr systems utilize the 1st Purkinje image, or the glint [12, 46], along with the pupil center for determining gaze direction and are capable of providing gaze vectors to within one degree of accuracy (neglecting the potential error associated with the one degree field of view with a stationary pupil [47]). The glint is the brightest and easiest reflection to detect and track [9]. In general, the use of a pccr method does not provide 3D location information. However, taking such actions as using stereo images can provide 3D locations. One of the seemingly more usable systems based on the pccr method, produced by Tobii [33], provides 3D gaze information relative to the axes of the stereo cameras. In addition, for ~$30,000, the Tobii 1750 claims to have an accuracy of 0.5 degrees for the gaze angle error. The relationship of this accuracy value compared to the one degree gaze angle error associated with the human eye is not discussed. According to Jacob [47], the same pupil position provides the subject with a one degree field of view on the fovea. Therefore, a subject can look clearly at any object within a one degree field of view while maintaining the same eye position. This would indicate that a system utilizing eye (pupil) position to determine gaze could not, on average, achieve gaze determination angular errors of less than one degree. The Tobii system also claims to have a drift of less than one degree and a compensation error for head translations in three dimensions and rotations across the entire head movement space of less than one degree. Unfortunately, the 3D location methodology is not adequately described. In addition, the exact relationship between the deviations and the gaze angle error over a prolonged sequence of images is not discussed.

Several gaze systems obtain high accuracy gaze results by utilizing either additional Purkinje images in their algorithms or additional illuminators producing multiple glints [25, 48]. Use of more than one Purkinje image usually requires specialized hardware for Purkinje image detection. However, multiple Purkinje image use does provide the capability to determine gaze direction/location in 3D without the use of stereo cameras by decoupling eye movement due to eye rotation and eye movement due to head translation. For example, Crane and Steele [48] use the 3rd and 4th Purkinje images to obtain a 3D gaze estimations without the need for stereo images.

In addition to Purkinje reflections and pupil centers, various other image features, most often facial features such as the nostrils and eye corners, can be used to determine gaze direction. Newman et al [49] locates the 3D position of the eye corners using stereo cameras and then computes the LoG (line of gaze) [9] using the orientation of the eyeball and an 'offset vector.' Park et al. [50] use the nostrils and lip corners along with the eyes to obtain a vector normal to the feature plane for estimating gaze. Their average reported gaze detection error with users 50 to 70 centimeters away from a 19 inch monitor was 5.11 centimeters.

2.5 Normal Lighting

If special illumination is not used, only facial features visible under normal lighting conditions or parameters derivable from normally visible facial features are available for gaze determination. Most often facial features such as the nostrils and eye corners are used to determine gaze. Newman [49] and Matsumoto [51] utilize eye corners in similar techniques for determining gaze. Matsumoto et al. [51] locates the 3D position of the eye corners using stereo cameras and then computes a gaze vector (gaze line) using the orientation of the eyeball and an 'offset vector.' Gaze angle errors averaging three degrees or less are reported. Park et al. [50] use the nostrils and lip corners along with the eyes to obtain a normal vector to the feature plane for estimating gaze. They report average gaze angle errors of less than five degrees. However, they restrict the amount of allowable head movement.

Regardless of the actual features involved, the process of normally visible facial feature finding is usually divided in to two logical steps: face localization and then the actual feature finding [4, 52].

2.5.1 Face Localization

The process of face localization involves finding a human face or faces in an image and identifying the boundaries of the face(s). After localization, a 'face' or face area is often defined as a rectangular sub-image (see Fig. 8) that contains those areas meeting the definition of a face [53]. The face area can then be used for further processing, and the remainder of the image can be discarded. While this step is not mandatory for either finding facial features or determining gaze, it is often incorporated in attempt to minimize the number of pixels that must be processed to locate the facial features.

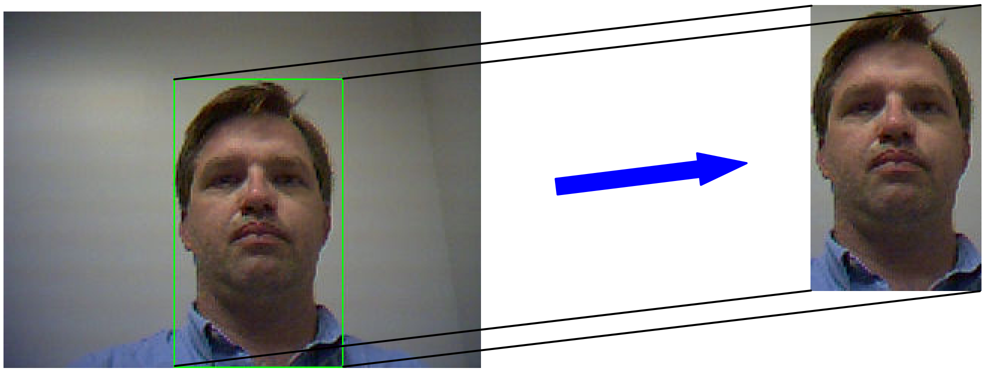


Fig. 8 Face localization using skin color based approach.

Bakic [4] outlines several approaches and provides numerous references for performing face localization:

a. clustering [54] (facial classification based on distance metrics from templates),
b. principal component analysis [55], (PCA, based on edge line extraction and matching with predefined templates),
c. layered rule matching [56] (manually coded rules of varying levels of complexity used with high-resolution, low-resolution, and edge-based versions of the images),
d. artificial neural networks [57, 58] (equalized pixel intensity values from sub-images are input to a neural network trained with face and non-face images),

e. support vector machine [59] (supervised learning function input/output vector (data point) pairs are created from training images and used to evaluate subsequent image outputs based on image function input), and
f. skin color based approaches [60] (cluster image based Gaussian distributions of image colors).

In her system, Bakic [4] implemented a skin color based approach. Image pixels in the red, green, and blue (RGB) color space are classified and clustered according to thresholds derived from combinations of the normalized red and green components. Pixels from those clusters defined as closely representing faces are grouped together into potential face objects using a connected component algorithm [61]. After eliminating objects deemed too small to be faces, the largest connected object is identified and then merged with objects that border it.

As mentioned previously, an attempt was made to implement Bakic's code. An example of the results using a slightly modified approach to her skin based method can be seen in Fig. 9.

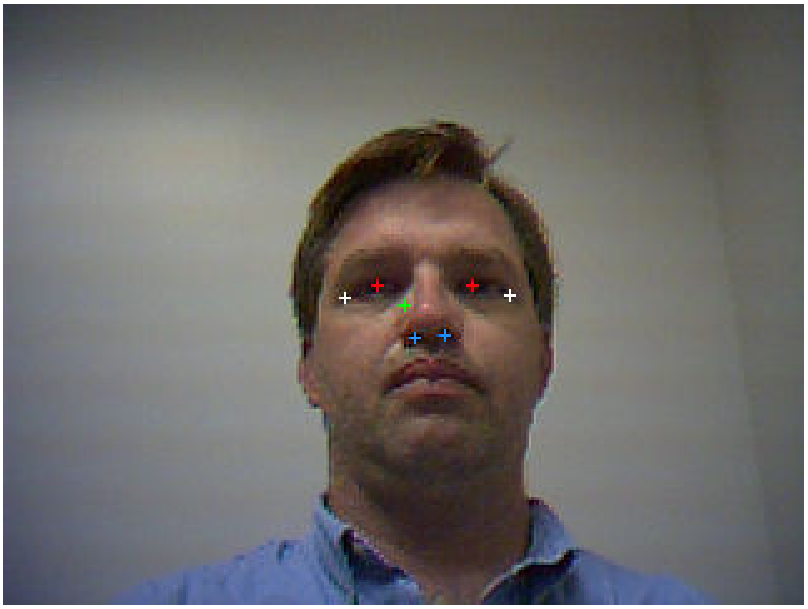


Fig. 9 Features located using skin color and geometric constraints.

2.5.2 Visible Facial Feature Location

Once the face area sub-image is extracted, the desired facial features can be located in the sub-image. With respect to gaze determination, facial features are usually thought of as those features of the human face that can be readily observed and consistently located in an image, and whose location can be meaningfully represented by a single pixel. They are often those features that have clearly identifiable boundaries or contours for which endpoints or centroids can be determined. The most common are eye corners (horizontal edge intersection point), mouth corners (horizontal edge intersection point), nostrils (centroid), and pupils (centroid). Because other visible anatomical features such as the cheeks, the chin, the nose, hair, etc. are not easily located, are difficult to represent with a single pixel, or deform significantly with respect to the head as a result of head movement and facial expression changes, they are usually excluded from gaze determination discussions.

As evidenced by the widespread use of illumination to highlight the pupil in current gaze tracking and eye movement systems, the identification of facial features without specialized illumination is not a trivial task. Unfortunately, locating facial features, particularly the pupils, is necessary for virtually all image-based gaze determination systems.

Bakic [4] lists several approaches found in the literature for locating various facial and non-facial features in images:

a. deformable templates [62] (image peaks and valleys are located and then feature templates are deformed to match peaks and valleys while minimizing template energy function),
b. eigen-feature template matching [63] (eigen vectors and eigenvalues are computed on the covariance matrix of training images for facial areas (eyes, nose, and mouth) keeping only the highest eigenvectors for matching of features in subsequent images),

c. snakelets [64] (curved shapes in facial images such as wrinkles , eyebrows, etc. are matched against preselected curves called snaklets and using the distance ratios between the snakelets for recognition),
d. skin color [60] (similar to face localization technique only Gaussian distributions of color for features are used),
e. geometric constraints [65] (use the image and anthropometric dimensions, relative positioning, and thresholding to isolate facial features), and
f. dark symmetry transformation [66] (the detection of significant edge configurations in an annular sampling region (eye positions) by first using wave propagation to compute a dark axial symmetry from an image phase and edge map and then computing a dark radial symmetry and using the strongest peaks as candidates for eye positions).

In her system, Bakic [4] uses a skin color model based approach to find various features. Assumptions that the pupils are the darkest objects in an image of the face, that the largest adsorption of light is represented by the red component of the image, and that skin is brighter in the red component than in the green component are leveraged to detect the pupils. Because all eyes are different, various threshold levels of the red component are used until pupil (eye) blobs appear as black regions in a white background. A connected components algorithm is run to create objects out of the black blobs. Objects that are too small, too big, or at the edge of the image are rejected. In addition to the pupils, nostrils, eye corners, mouth corners, and similar features are often detected. Bakic attempts to use eyebrows, and geometric and anthropometric relationships to identify those objects that indeed represent pupils.

During the attempt to implement Bakic's code, the eyebrows could not be reliably located. Therefore, geometric constraints were applied to the face image to designate the eye objects. An example of the results from adding geometric constraints is presented in Fig. 9 where the red crosses represent the pupils, the white crosses the outside eye corners, the blue crosses represent the nostrils, and the green cross represents the center of the image.

2.5.3 Hidden Feature Location

In addition to using features that are visible in an image for gaze determination, several approaches use non-visible, or hidden features as well. The most interesting of the hidden feature approaches is the approach used by Matsumoto [51] in the faceLab 4 system [31] by Seeing Machines and the one-circle approach used by Wang [10, 67]. The hidden feature used in these approaches, and the most commonly used for gaze determination, is a notional location in the interior of the eye called the eye center. Matsumoto [51] defines an offset vector from the eye center through the midpoint of a line between the corners of an eye that is fixed with respect to the head pose. The offset vector, determined in an unspecified fashion during a manual training session, is used to locate the eye center when the head pose and corners of an eye are known. Matsumoto then uses a hough transform [68] to locate the center of the iris. His gaze vector starts from the eye center and passes through the iris center.

Wang's method [10, 67] determines a user's iris radius during calibration. Then using the iris radius, a circular eye model, and the image of the eye, an iris plane is formed by the image-derived iris circle bisecting the eye model circle. The eye center is then found by projecting along the normal to the iris plane a distance derived from the iris radius and a generic eye radius. As with Matsumoto, Wang then projects from the eye center through the iris center to determine the user's gaze. Unfortunately, the technique degrades if the iris contour is symmetric about the Y-Z plane of the camera and the optical axis of the camera passes through the iris center.

Both Matsumoto and Wang seem to infer that the eye center remains fixed with respect to the head. Listing’s Law, as documented by Helmholtz [69, 70], states that "when the line of sight is moved from the primary position to an another position, the amount of torsion in this second position is such as if the eye had rotated about a fixed axis, which is perpendicular to the line of sight in the two positions." This implies that changes in gaze position due to eye movement can be modeled as pure rotation, and, that all axes of rotation and lines of sight share a common intersection point relative to the eye [71]. In addition, given that the eye is positioned within the head by the six muscles that facilitate eye movement [47, 72, 73], it seems to follow that the single point of rotation of the eye is in a fixed location with respect to the head [74, 75]. This location, lying along Morimoto's optical axis [9] and often referred to as the center of eye rotation, appears to be consistent with Masumoto and Wang's eye center. Although it turns out that there is no fixed center of rotation with respect to the head [13, 19] (nor would it lie on Carpenter's version of the optical axis [13]), an average center of eye rotation can be successfully used because of the amount of movement from the average is relatively small (0.4 mm [76]). Unfortunately, an assessment of the methodology for estimating and/or calculating the eye center cannot be made because neither Wang [67] nor Matsumoto [51] provide details of their eye center determination.

Once the desired image features (either visible and/or hidden) have been determined, they can then be used to determine the gaze (gaze vector) in whatever manner the gaze determination system being used allows.

2.6 Camera Calibration

Whether one uses visible features or a combination of hidden and visible features, the locations of these features are initially determined in the coordinate system of the camera being used to collect the images because the images are merely projections of physical objects onto the camera's image plane. For these features to be useful, they (or the resultant gaze vector) must be related to the outside world. Image-based systems must establish some relationship between the camera's image plane and the surrounding environment in order to be able to extract any environmental information (most often spatial information) from an image. Creating a relationship between the camera and its environment is the objective of camera calibration.

Seitz outlines four methods of camera calibration:

a. geometric [77, 78] (linear or non-linear method relying on multiple images collected of an object with a known geometry in different spatial orientations to produce the camera's intrinsic parameters),
b. radiometric [79] (multiple images of the same scene at different exposures are collected to produce a radiance response function for the camera),
c. structure-from motion [80] (tracks corresponding points over a sequence of images to solve for 2D location relative to camera position), and
d. self-calibration [81] (using sequences of corresponding images and an assumption of no skew to retrieve a metric reconstruction of varying intrinsic parameters due to zooming and focusing changes).

Because it provides spatial relationships, the most popular calibration method with respect to computer vision and gaze determination [82] is the geometric method. The geometric method consists of two phases. The first phase involves the determination of a camera's intrinsic parameters. The second phase, known as pose estimation [82], involves determination of extrinsic parameters relative to the desired real-world coordinate system.

2.6.1 Intrinsic Parameters

The intrinsic properties of a camera [83] consist of:
a. the focal length (in pixels),
b. the principal point or image center,
c. the skew coefficient or aspect ratio, and
d. the radial and tangential image distortion coefficients.

Heikkila [84] and Zhang [85] describe methods to determine the intrinsic parameters of a camera. Using these methods, a Matlab toolbox [86] is available that determines the intrinsic parameters from images of a checkerboard pattern (see Fig. 10) in varying spatial orientations. In addition to the intrinsic parameters for each image and the average parameters over all the images, the toolbox provides an average pixel error. The average pixel error provides an indication of the distance between where a point appeared on the actual image and where it was predicted to have appeared based on the intrinsic parameters.

2.6.2 Extrinsic Parameters

After the determination of a camera's intrinsic parameters, a geometric calibration then attempts to detail that camera's coordinate system spatial relationship to some alternate coordinate system. An alternate coordinate system may be a monitor coordinate system, the coordinate system of another camera, the coordinate system of a subject's head, or any other relevant system. The parameters that specify this relationship between the camera and the alternate coordinate system are known as extrinsic parameters. If a camera is being related to multiple alternate coordinate systems, there will be a set of extrinsic parameters for each camera/alternate coordinate system pair.

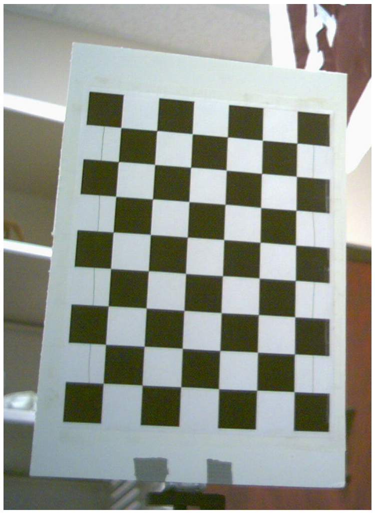


Fig. 10 Camera calibration checkerboard.

The extrinsic parameters for a given camera and a single alternate coordinate system pair consist of a rotation vector (*RV*) and a translation vector (*TV*). The vector RV gives the axis about which the rotation takes place. In the case of the Matlab routines used for this dissertation, RV is scaled so that its magnitude represents the angle of rotation. The translation vector represents the location of the alternate coordinate system relative to the camera. The rotation vector (*RV*) is often represented by a more familiar coordinate transformation structure, a 3x3 rotation matrix. A rotation matrix (*R*) can be derived from a rotation vector (*RV*) using the Rodrigues formula [87, 88]. The Rodrigues formula is derived/expressed using the following:

$$\theta = \|RV\|, \text{ where } \theta \neq 0 \qquad (1)$$

$$\alpha = \cos(\theta) \qquad (2)$$

$$\beta = \sin(\theta) \qquad (3)$$

$$\gamma = 1 - \cos(\theta) \qquad (4)$$

$$R = (I * \alpha) + \left(\begin{bmatrix} 0 & \frac{-Z_{RV}}{\theta} & \frac{-Y_{RV}}{\theta} \\ \frac{Z_{RV}}{\theta} & 0 & \frac{-X_{RV}}{\theta} \\ \frac{-Y_{RV}}{\theta} & \frac{X_{RV}}{\theta} & 0 \end{bmatrix} * \beta \right) + \left(\frac{(RV)^T}{\|RV\|} * \frac{RV}{\|RV\|} * \gamma \right) \qquad (5)$$

For the purposes of this dissertation, variables such as RV or α may represent either real scalar, vector, or matrix quantities depending on the context. However, the variable labels *X*. *Y*, or *Z*, unless specifically noted otherwise, will always represent vector component locations. Therefore, X_{RV} would represent the *X*-component of the vector RV. In addition, the coordinate system that values of a variable are represented in, if applicable, will be specified by a superscript to the left of the variable.

Tsai [77] uses a 3D grid with a defined location in the desired 3D world coordinate system to determine the extrinsic parameters of a camera with respect to the coordinate system of that grid. Since the grid point locations are known in the grid coordinate system, the coordinate transformation is readily determined. However, this is only useful for transformations between the camera system and the 3D grid. The method does not readily extend to arbitrary 3D objects.

However, because the Matlab routines [86] used for this dissertation utilize the 2D checkerboard in unknown orientations, additional information must be provided in order to relate the camera coordinate system to an alternate 3D coordinate system. Matlab provides a mechanism to relate one camera to another coordinate system, if the alternate coordinate system is another camera coordinate system, and images of identical calibration objects in identical orientations are available from both cameras, e.g., if both cameras image the objects at the same time.

The information Matlab requires to determine the extrinsic parameters relating one camera to another can be obtained by collecting stereo images of the checkerboard. This can be done by utilizing, as stereo pairs, the images obtained for intrinsic parameter determination for each of the stereo cameras. Then, using additional Matlab routines, the extrinsic parameters relating the two stereo cameras together can be found in conjunction with finding the intrinsic parameters. The Matlab routines output a single rotation vector and translation vector for each camera in the two camera pair. The resulting extrinsic parameters of the first camera (*Camera 1*) would relate that camera's coordinate system to the coordinate system of the second camera (*Camera 2*). The extrinsic parameters of *Camera 2* would relate the coordinate system of *Camera 2* to *Camera 1*. Having these relationships between the two cameras facilitates the determination of image features in 3D using stereo triangulation and images collected from both cameras.

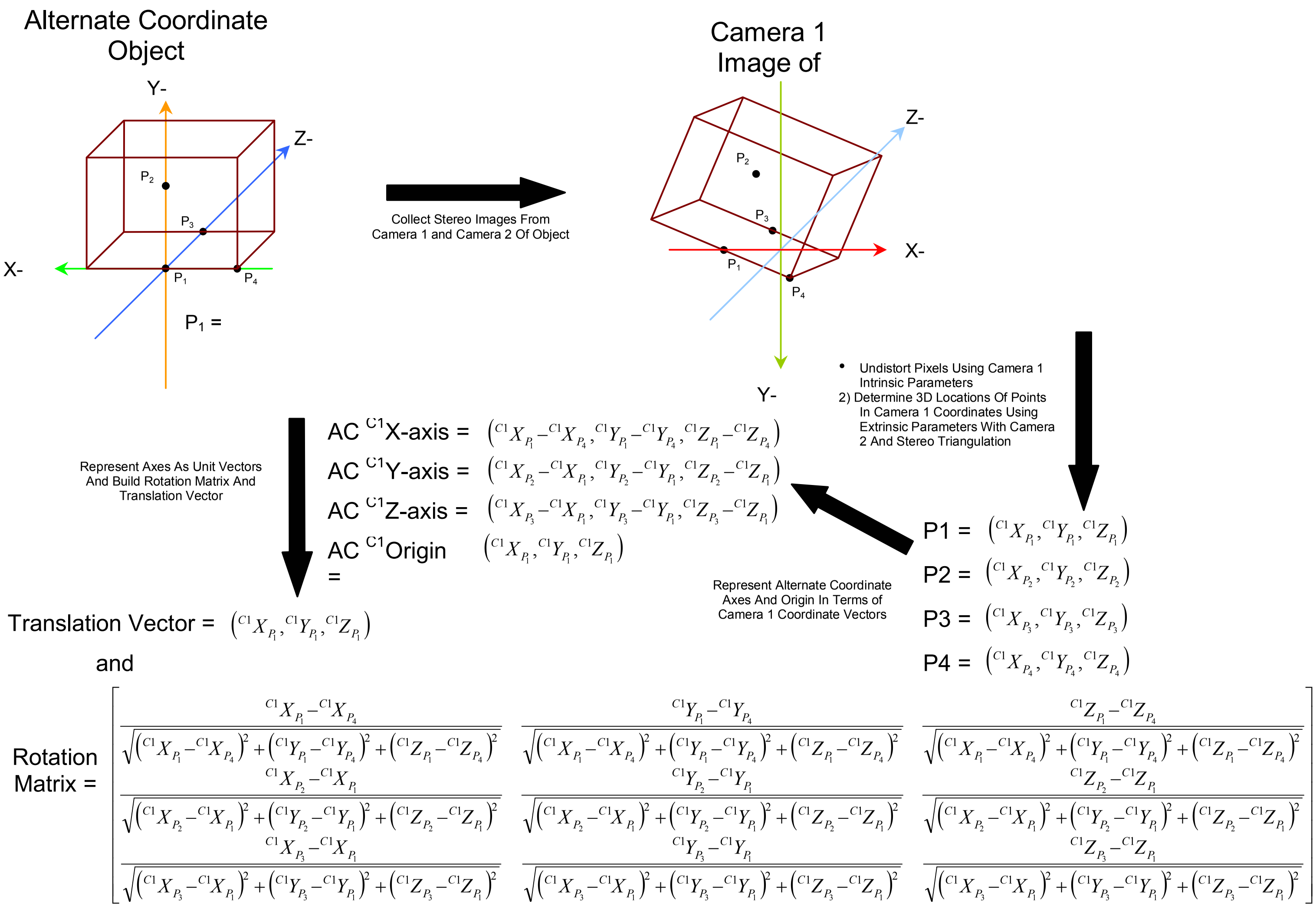
Because a relationship is desired in a reference frame or coordinate system other than that of one of the cameras, additional efforts must be made to obtain a camera's extrinsic parameters with respect to some non-camera coordinate system. In order to accomplish this, stereo images can be collected of objects with known locations in the desired, non-camera coordinate system. Points on these objects with known locations in the non-camera coordinate system can be identified in their stereo images and their 3D locations in either or both camera coordinate systems can be determined using stereo triangulation. With the 3D locations of points also known in one or both of the camera coordinate systems, the rotation matrix and translation vector relating the camera coordinate system to the alternate coordinate system can be determined.

A rotation matrix and translation vector can be readily generated by using known object points in the non-camera coordinate system that lie along the three axes of the non-camera coordinate system and identifying these points in the stereo images. After determining the 3D locations of these points in one of the camera coordinate systems (it doesn't matter which), unit vectors that represent the axes of the non-camera coordinate system in camera coordinates can be developed. The *X*, *Y*, and *Z* components of each of these unit vectors become the coefficients of the rotation matrix [89]. The components of the alternate coordinate system origin in camera coordinates define the translation vector (see Fig. 11).

Having the rotation matrix (*R*) and a translation vector (*TV*) relating the camera coordinates to an alternate coordinate system, allows any 3D location in the camera coordinate system to be represented as, or transformed, into a 3D location in the alternate coordinate system. The transformation of a location (*L)* in any coordinate system (*A*) to any other coordinate system (*B*) can be represented as:

$$ {}^{B}L = {}^{A}L * {}^{B}_{A}R + {}^{B}_{A}TV \tag{6} $$

This equality will be utilized throughout the remainder of this dissertation to transform not only between camera coordinate systems, but also between the rig/monitor (see Subsection 5.3), head (see Subsection 3.3), and various camera coordinate systems.



Collect Stereo Images From Camera 1 and Camera 2 Of Object

- Undistort Pixels Using Camera 1 Intrinsic Parameters

2) Determine 3D Locations Of Points In Camera 1 Coordinates Using Extrinsic Parameters With Camera 2 And Stereo Triangulation

P1 = $\left({}^{C1}X_{P_1}, {}^{C1}Y_{P_1}, {}^{C1}Z_{P_1}\right)$

P2 = $\left({}^{C1}X_{P_2}, {}^{C1}Y_{P_2}, {}^{C1}Z_{P_2}\right)$

P3 = $\left({}^{C1}X_{P_3}, {}^{C1}Y_{P_3}, {}^{C1}Z_{P_3}\right)$

P4 = $\left({}^{C1}X_{P_4}, {}^{C1}Y_{P_4}, {}^{C1}Z_{P_4}\right)$

Represent Alternate Coordinate Axes And Origin In Terms of Camera 1 Coordinate Vectors

AC ^{C1}X-axis = $\left({}^{C1}X_{P_1} - {}^{C1}X_{P_4}, {}^{C1}Y_{P_1} - {}^{C1}Y_{P_4}, {}^{C1}Z_{P_1} - {}^{C1}Z_{P_4}\right)$

AC ^{C1}Y-axis = $\left({}^{C1}X_{P_2} - {}^{C1}X_{P_1}, {}^{C1}Y_{P_2} - {}^{C1}Y_{P_1}, {}^{C1}Z_{P_2} - {}^{C1}Z_{P_1}\right)$

AC ^{C1}Z-axis = $\left({}^{C1}X_{P_3} - {}^{C1}X_{P_1}, {}^{C1}Y_{P_3} - {}^{C1}Y_{P_1}, {}^{C1}Z_{P_3} - {}^{C1}Z_{P_1}\right)$

AC ^{C1}Origin = $\left({}^{C1}X_{P_1}, {}^{C1}Y_{P_1}, {}^{C1}Z_{P_1}\right)$

Represent Axes As Unit Vectors And Build Rotation Matrix And Translation Vector

Translation Vector = $\left({}^{C1}X_{P_1}, {}^{C1}Y_{P_1}, {}^{C1}Z_{P_1}\right)$

and

Rotation Matrix =
$$\begin{bmatrix}
\frac{{}^{C1}X_{P_1} - {}^{C1}X_{P_4}}{\sqrt{\left({}^{C1}X_{P_1} - {}^{C1}X_{P_4}\right)^2 + \left({}^{C1}Y_{P_1} - {}^{C1}Y_{P_4}\right)^2 + \left({}^{C1}Z_{P_1} - {}^{C1}Z_{P_4}\right)^2}} & \frac{{}^{C1}Y_{P_1} - {}^{C1}Y_{P_4}}{\sqrt{\left({}^{C1}X_{P_1} - {}^{C1}X_{P_4}\right)^2 + \left({}^{C1}Y_{P_1} - {}^{C1}Y_{P_4}\right)^2 + \left({}^{C1}Z_{P_1} - {}^{C1}Z_{P_4}\right)^2}} & \frac{{}^{C1}Z_{P_1} - {}^{C1}Z_{P_4}}{\sqrt{\left({}^{C1}X_{P_1} - {}^{C1}X_{P_4}\right)^2 + \left({}^{C1}Y_{P_1} - {}^{C1}Y_{P_4}\right)^2 + \left({}^{C1}Z_{P_1} - {}^{C1}Z_{P_4}\right)^2}} \\
\frac{{}^{C1}X_{P_2} - {}^{C1}X_{P_1}}{\sqrt{\left({}^{C1}X_{P_2} - {}^{C1}X_{P_1}\right)^2 + \left({}^{C1}Y_{P_2} - {}^{C1}Y_{P_1}\right)^2 + \left({}^{C1}Z_{P_2} - {}^{C1}Z_{P_1}\right)^2}} & \frac{{}^{C1}Y_{P_2} - {}^{C1}Y_{P_1}}{\sqrt{\left({}^{C1}X_{P_2} - {}^{C1}X_{P_1}\right)^2 + \left({}^{C1}Y_{P_2} - {}^{C1}Y_{P_1}\right)^2 + \left({}^{C1}Z_{P_2} - {}^{C1}Z_{P_1}\right)^2}} & \frac{{}^{C1}Z_{P_2} - {}^{C1}Z_{P_1}}{\sqrt{\left({}^{C1}X_{P_2} - {}^{C1}X_{P_1}\right)^2 + \left({}^{C1}Y_{P_2} - {}^{C1}Y_{P_1}\right)^2 + \left({}^{C1}Z_{P_2} - {}^{C1}Z_{P_1}\right)^2}} \\
\frac{{}^{C1}X_{P_3} - {}^{C1}X_{P_1}}{\sqrt{\left({}^{C1}X_{P_3} - {}^{C1}X_{P_1}\right)^2 + \left({}^{C1}Y_{P_3} - {}^{C1}Y_{P_1}\right)^2 + \left({}^{C1}Z_{P_3} - {}^{C1}Z_{P_1}\right)^2}} & \frac{{}^{C1}Y_{P_3} - {}^{C1}Y_{P_1}}{\sqrt{\left({}^{C1}X_{P_3} - {}^{C1}X_{P_1}\right)^2 + \left({}^{C1}Y_{P_3} - {}^{C1}Y_{P_1}\right)^2 + \left({}^{C1}Z_{P_3} - {}^{C1}Z_{P_1}\right)^2}} & \frac{{}^{C1}Z_{P_3} - {}^{C1}Z_{P_1}}{\sqrt{\left({}^{C1}X_{P_3} - {}^{C1}X_{P_1}\right)^2 + \left({}^{C1}Y_{P_3} - {}^{C1}Y_{P_1}\right)^2 + \left({}^{C1}Z_{P_3} - {}^{C1}Z_{P_1}\right)^2}}
\end{bmatrix}$$

Fig. 11 Extrinsic parameter determination.

3. VISUAL AXES CENTER METHOD

This section of the dissertation discusses a non-invasive, image-based analytic approach for gaze determination. The method proposed utilizes the notion of a visual axes center as described by Carpenter [13]. Carpenter's notion of a visual axes center is discussed along with an approximation that uses the pupil center instead of the nodal points to estimate the visual axes center. The mechanics of the proposed method are then discussed along with the addition of an optimization technique designed to reduce possible errors. In addition, the determination of a head coordinate system with respect to which the visual axes center is fixed is discussed. Finally, the use of the proposed method in a gaze determination system is also discussed.

The method being proposed to facilitate gaze determination is based on Carpenter's [13] definition of a head-fixed visual axes center (see Subsection 1.2.2). With the visual axes center, a gaze vector can be derived by projecting from the visual axes center through the nodal point (see Subsection 1.2.2). Unfortunately, there is no mechanism to directly locate either the visual axes center or the nodal point using image processing. The next subsection discusses an approximation that can be used for the nodal point. The remaining problem, then, is the determination of the visual axes center. If the visual axes center can initially be spatially located by other means with respect to the head (which can be located directly from images), image processing could be used to locate the head and subsequently locate the visual axes center. Subsection 3.2 will discuss a technique for such a determination.

3.1 Nodal Point Approximation

Unfortunately, the nodal points are notional points and cannot be located directly from images. However, the data that Blaine [17] and Thibos et al. [16, 18] provide on the angle between the achromatic (pupillary) axis and the visual axis can be used to show that it is reasonable to approximate the nodal point location by the pupil location.

Given that the angle between the pupillary axis and the visual axis is in the range of two to three degrees [17] and the entrance pupil (pupil center) is no more than two millimeters in front of the nodal point [13] (assumed to be along the pupillary axis), the distance between the pupil center (*PC*) and the visual axis (*VA*) is estimated using the following relationship (see Fig. 12):

PC to VA distance (*dPCVA*) = 2 mm * sin(3°) ≈ 0.105 millimeters. (7)

Carpenter reports that the visual axes center is near the average center of rotation of the eye [13] and that the average center of rotation of the eye can be approximated by the center of the eyeball [13, 49]. Since it is known that the average radial distance of the eyeball is approximately 12.5 millimeters [90], an error estimation of using the pupil for the nodal point is made using a distance between the visual axes center and the pupil center of 12.5 millimeters. Using the following relationship for the angular error between the actual visual axis and the 'pseudo' visual axis found using the pupil center as a substitution for the nodal point the error is found to be:

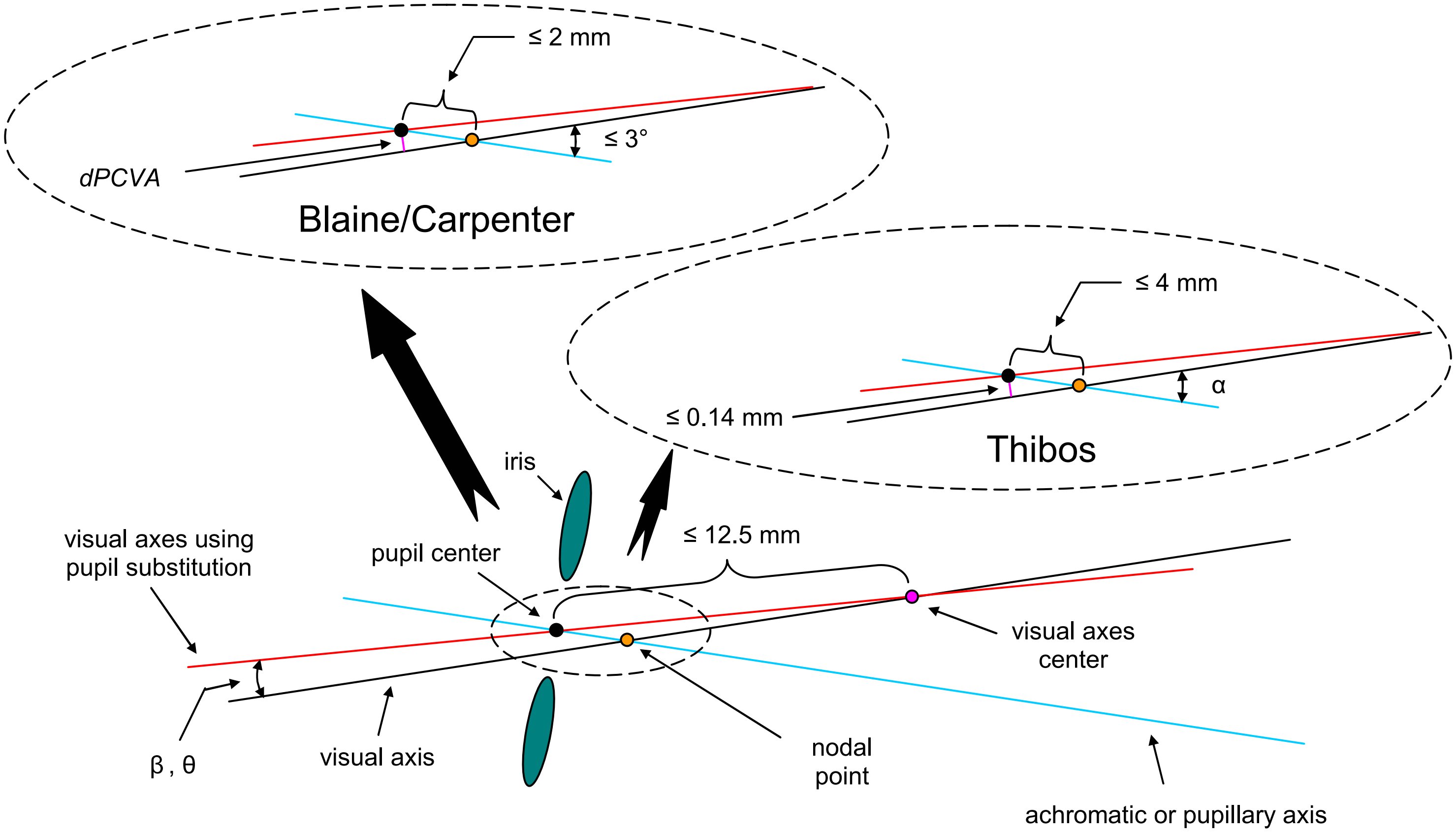
angular error = $\sin^{-1}$(0.105 mm / 12.5 mm) ≈ 0.49 degrees. (8)

Another error estimate can be made based on Thibos et al. [16, 18]. They estimate the displacement between the pupil center and the nodal point (the pupil center is closer to the temple than the nodal point) to be 0.14 millimeters perpendicular to the visual axis. Using Thibos' estimate of the pupil center/nodal point distance (4 millimeters) the following relationship for the angle between the visual axis and the achromatic axis is:

angle = $\tan^{-1}$(0.14 mm / 4.0 mm) ≈ 2.01 degrees (9)

Also, the angle between the actual visual axis and the 'pseudo' visual axis found using the pupil center as a substitution for the nodal point can be determined using the following:

angular error = $\sin^{-1}$(0.14 mm / 12.5 mm) ≈ 0.65 degrees. (10)



Calculations-

dPCVA = 2 mm * sin(3°) = 0.1047 mm

α = arcsin(0.14 mm / 4 mm) = 2.005°

β = arcsin(0.105 mm / 12.5 mm) = 0.481°

θ = arcsin(0.14 mm / 12.5 mm) = 0.642°

Fig. 12 Pupil center for nodal point substitution error (right eye, top view).

Assuming no other error sources, any angular error in the visual axis would result in an identical angular error in a gaze direction determined using the alternate visual axis. The overall errors measured during the conduct of several experiments (see Section 5) are deemed acceptable, and hence the approximation is also deemed to be acceptable.

The previous pupil substitution error estimates assume that the distance between the pupil center (*PC*) and the visual axes center (*VAC*) remains constant for a given subject. In the strictest sense, this is known to not be the case [13]. In an attempt to estimate the impact of this potential for movement, the value of *dVP* is adjusted such that the distance between the pupil center and visual axes center varies by no more than 0.4 millimeters , an estimate of the maximum translation of the eye with respect to the head [13, 76]. Since the angle between the pupillary axis and the visual axis is no more than 3 degrees, the worst case error occurs for the movement along the worst case pupillary axis shown in Fig. 13. Therefore, one can assume that, for a given subject, the pupil may move toward the visual axes center, or it may move away from the visual axes center. The movement that creates the maximum angular difference from either the original visual axis or the visual axis using pupil substitution is away from the visual axes center (see Fig. 13). The following derivation leads to a potential angular error estimate associated with the distance variation of less than 0.08 degrees:

a. using the result of Subsection 3.1, the angle between the pupil substitution visual axis and the pupillary axis (α) is

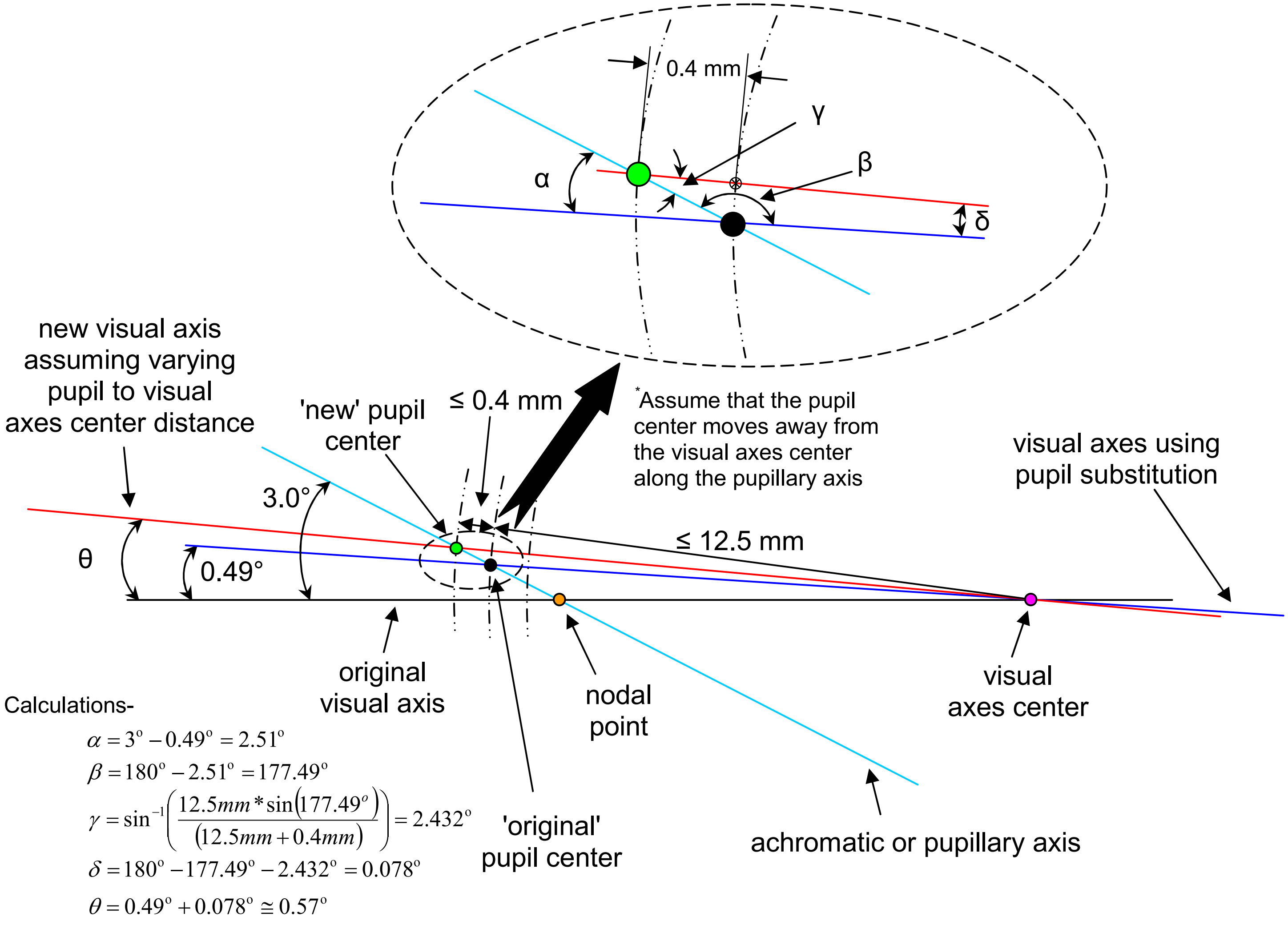
$$\alpha = 3^o - 0.49^o = 2.51^o \tag{11}$$

b. the angle between the pupil substitution visual axis and the pupillary axis (β) is

$$\beta = 180^o - 2.51^o = 177.49^o \tag{12}$$

c. using the law of sines, the angle between the pupillary axis and the new visual axis after accounting for pupil movement (γ) is

$$\gamma = \sin^{-1}\left(\frac{12.5mm * \sin\left(177.49^o\right)}{\left(12.5mm + 0.4mm\right)}\right) = 2.432^{\circ} \tag{13}$$


0.4 mm
γ
β
α
δ
new visual axis assuming varying pupil to visual axes center distance
'new' pupil center
≤ 0.4 mm
*Assume that the pupil center moves away from the visual axes center along the pupillary axis
visual axes using pupil substitution
3.0°
≤ 12.5 mm
θ
0.49°
original visual axis
nodal point
visual axes center
'original' pupil center
achromatic or pupillary axis

Calculations-

$$\alpha = 3^{o} - 0.49^{o} = 2.51^{o}$$

$$\beta = 180^{o} - 2.51^{o} = 177.49^{o}$$

$$\gamma = \sin^{-1}\left(\frac{12.5mm * \sin(177.49^{o})}{(12.5mm + 0.4mm)}\right) = 2.432^{o}$$

$$\delta = 180^{o} - 177.49^{o} - 2.432^{o} = 0.078^{o}$$

$$\theta = 0.49^{o} + 0.078^{o} \cong 0.57^{o}$$

Fig. 13 Angular error due to varying *dVP* (right eye, top view).

d. the angle between the pupil substitution visual axis and the new visual axis after accounting for pupil movement (δ) is

$$\delta = 180^{\circ} - 177.49^{\circ} - 2.432^{\circ} \;\; = 0.078^{\circ} \leq 0.08^{o} \tag{14}$$

Therefore, the error introduced by assuming that the pupil center to visual axes center distance remains constant appears to be insignificant.

3.2 Visual Axes Center Determination

In addition to the nodal point (approximated by the pupil center), another point is needed that intersects the visual axis in order to determine the path (direction) of a visual axis. The only other points that are defined to exist on a visual axis are the visual axes center and the fixation point. As mentioned previously, in gaze determination applications, the fixation point is usually an unknown. Therefore, the visual axes center becomes the only other possible determinable point with which to specify a subject's visual axis. However, once determined for a particular subject, the location of the visual axes center remains fixed with respect to the head. Therefore, determination of a subject's visual axes center is similar to the subject calibration efforts required by many other gaze determination systems in that it is required only once for a given subject.

Given a collection of stereo images of a subject looking at a variety of known 3D locations with sufficiently differing eye movement, an estimation of the visual axes center can be made during a calibration phase. Each visual axis can be determined by projecting from the known fixation point through the nodal point (approximated by the pupil center) found using image processing and stereo triangulation. Then, the visual axes center can be determined by taking the intersection point of any two visual axis pairs. Since eye motion was assumed between each of the calibration images, each of the visual axes has some angular displacement with all the other visual axes (no two visual axes will be parallel) when represented in a coordinate system fixed with respect to the head. In addition, all the visual axes for a particular subject should intersect.

However, because of the inherent errors associated with determining 3D locations using image processing (pixel location errors, intrinsic/extrinsic camera parameter estimation errors, rounding errors, etc.) and the fact that the pupil center was used as an approximation for the nodal point, there is a significant likelihood that none of the visual axes will actually intersect. Therefore, a more appropriate center determination method that doesn't rely on the visual axes actually intersecting is needed. The proposed alternative is to average the midpoints of the lines between the closest approach points for each available visual axis pair (see Fig. 14). This average represents a reasonable approximation that can be used as a substitute for the actual visual axes center.

3.2.1 Closest Approach Midpoint Averaging

In order to simplify the implementation of the closest approach midpoint averaging technique, it is assumed that the visual axis vectors VA_i (where i runs from 1 to the number of stereo image pairs being used in the calibration) are represented in a 3D head coordinate system similar to the one described in Subsection 3.3. The ability to represent the visual axes in a fixed, head coordinate system not only simplifies the required averaging calculations, it also virtually eliminates the need to restrict the subject's head movement during the actual collection of the stereo images. It also streamlines the effort required to use the approximated visual axes center for subsequent gaze determination.

Once each of the visual axis vectors (VA_i) is represented in the same head coordinate system, the closest approach midpoint averaging technique is initiated by finding all of the lines defined by the points of closest approach between all of the pairs of visual axes. After all of the lines representing the closest approach points are determined, the midpoint of each of these lines is determined. Then, all of the midpoints are averaged. It is this average midpoint that can be used to approximate the visual axes center for a particular eye of a particular subject.

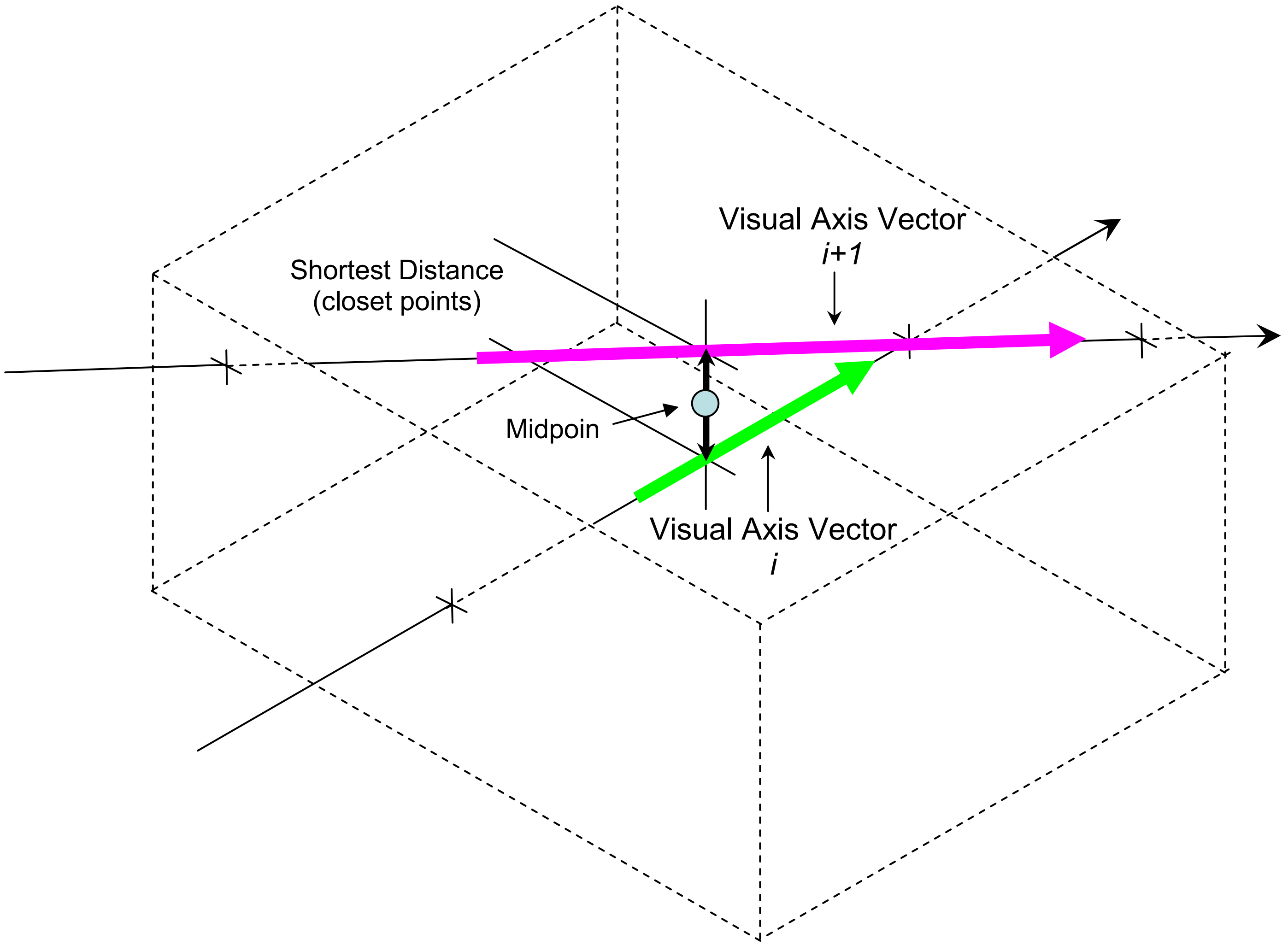


Fig. 14 Closest approach midpoint.

The determination of the closest approach points (CAP_i and CAP_j) between VA_i *and* VA_j, and thus, the closest approach line between the two points, is accomplished using the following equalities for each visual axis pair VA_i *and* VA_j ($i \neq j$ and $j>i$) represented as unit vectors $\hat{VA}_i$ and $\hat{VA}_j$:

$$a = \hat{VA}_i \bullet \hat{VA}_i \tag{15}$$

$$b = \hat{VA}_i \bullet \hat{VA}_j \tag{16}$$

$$c = \hat{VA}_j \bullet \hat{VA}_j \tag{17}$$

$$d = \hat{VA}_i \bullet \left(FP_i - FP_j\right) \tag{18}$$

$$e = \hat{VA}_j \bullet \left(FP_i - FP_j\right) \tag{19}$$

$$sc = \left(\frac{(b*e)-(c*d)}{a*c}\right) - b^2 \tag{20}$$

$$tc = \left(\frac{(a*e)-(b*d)}{a*c}\right) - b^2 \tag{21}$$

$$CAP_i = \left(FP_i + VA_i\right) * sc \tag{22}$$

$$CAP_j = \left(FP_j + VA_j\right) * tc \tag{23}$$

where FP_i and FP_j represent the fixation points associated with each of the visual axes VA_i *and* VA_j. An approximate visual axes center (*VAC*) can then determined by averaging all of the midpoints of each of the closest approach lines found by subtracting CAP_i and CAP_j for $j>i$:

$$\textit{average midpoint} = VAC = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{CAP_i + CAP_j}{2}\right)}{\sum_{k=1}^{n-1} (n-k)} \tag{24}$$

The average midpoint is then used to determine and average pupil center (*PC*) to visual axes center (*VAC*) distance (*dVP*). The purpose of determining the average *dVP* will be explained in subsequent subsections.

3.2.2 Closest Approach Midpoint Averaging Adjustment

The *VAC* will be used for gaze determination as described in Subsection 3.4. That is, a gaze vector will be determined by projecting from the *VAC* through the pupil center. The location of the pupil center during gaze determination will be found using not only image processing, but also the calibration values of *VAC* and *dVP*. During calibration, this provides a way to check the validity of the method used to determine *VAC* and *dVP*, and then to refine the estimates for *VAC* and *dVP*. Using the *VAC*, *dVP*, and a known fixation point, one can project from the *VAC* toward the fixation point a distance of *dVP*. This yields the effective pupil center point for the calibration. This effective pupil center point can be compared with the pupil center point obtained from the stereo image processing to determine an error metric. One can then modify the values of *VAC* and *dVP* to minimize the cumulative distance between the effective pupil center points and their corresponding actual locations from image processing in an attempt to improve the *VAC* and *dVP* estimates.

In order to obtain the refinement, an iterative adjustment of the closest approach midpoint averaging values of *VAC* and *dVP* is performed. The goal of this iterative adjustment is to minimize the total distance between the pupil centers determined from image processing and those that would be determined during gaze determination using the closest approach midpoint averaging estimates for *VAC* and *dVP*. The following pseudo-code details this iterative adjustment technique:

```
Execute 1st phase
1st phase
{
    Set $VAC = VAC_0$ ($VAC_0$ is found using the closest approach midpoint averaging technique)
    Set $dVP = dVP_0$ ($dVP_0$ is found using the closest approach midpoint averaging technique)
    Set $increment = 0.1$
    Set $lBound = -1.0$
    Set $uBound = 1.0$
    Skip to ★
}
2nd phase
{
    Set $VAC = VAC_{min}$
    Set $dVP = dVP_{min}$
    Set $increment = 0.01$
    Set $lBound = -0.1$
    Set $uBound = 0.1$
}
★   Set $VAC_{min} = VAC$
    Set $dVP_{min} = dVP$
    Set $min = \infty$
    Set $minJ = 0$
    Set $minK = 0$
    Set $minM = 0$
    Set $minN = 0$
    Vary $j$ from $lBound$ to $uBound$ in increments of $increment$
    {   Set $dVP_j = dVP + j$
    Vary $k$ from lBound to uBound in increments of increment
    {   Set $X_k = VAC_X + k$
        Vary $m$ from lBound to uBound in increments of increment
        {   Set $Y_m = VAC_Y + m$
            Vary $n$ from lBound to uBound in increments of increment
            {   Set $Z_n = VAC_Z + n$
                Set $sum = 0$
                For each image $i$
                {   Determine $PC'$ by projecting from the current value of $<X_k, Y_m, Z_n>$ a distance of the current value of $dVP_j$ toward $FP_i$ (where $FP_i$ is the fixation point for image $i$)
                    $sum = sum + |PC'-PC_i|$ where $PC_i$ is the pupil center from image processing for image $i$
                }
                if $sum < min$
                {   Set $min = sum$
                    $VAC_{min} = <X_k, Y_m, Z_n>$
                    $dVP_{min} = dVP_j$
                    $minJ = j$
                    $minK = k$
                    $minM = m$
                    $minN = n$
                }
            }
        }
    }
}
```

If *minJ*, *minK*, *minM*, or *minN* equals either *lBound* or uBound, repeat the current phase starting at ★ after setting $VAC = VAC_{min}$ and $dVP = dVP_{min}$. Otherwise, if on the 1st phase, go to the 2nd phase, or if already on the 2nd phase, stop.

The values of VAC_{min} and dVP_{min} at the completion of the last iteration of the 2nd phase will be used in all subsequent processing for *VAC* and *dVP*. Note that the pseuso-code as presented is applied during calibration to a single eye for a particular subject. Because both eyes will be used for subsequent gaze determination, the iterative adjustment must also be performed for the other eye of each subject.

Fig. 15 graphically summarizes the general notion behind the *VAC*/*dVP* adjustment for a single eye of a subject. The potential impact of the pupil center location, and therefore, the potential effect of this adjustment on the accuracy of gaze direction estimation for the calibration images will be discussed in Subsection 7.10.

3.2.3 Estimation of Visual Axes Center

The last consideration in the visual axes center determination results from the fact that the visual axes center determined using the pupil center as a substitution for the nodal point creates an error due to the difference between the estimated visual axes center and the true visual axes center. Unfortunately, the magnitude of this error is dependant on the location of the actual visual axes center and the amount the estimated visual axes center moves with respect to the head, both of which are unknowns. In addition, no definitive estimates of the distance between the actual visual axes center and other eye features (pupil center, rotation center, fovea, etc.) have been found in the literature that would allow reasonable approximations of these values to be derived. Therefore, an estimate of the error introduced by using the estimated visual axes center is not determinable at this time. Rather, the results of conducting the experiments described in Section 5 provide reasonable insight as to the acceptability of using the estimated visual axes center instead of the actual visual axes center.

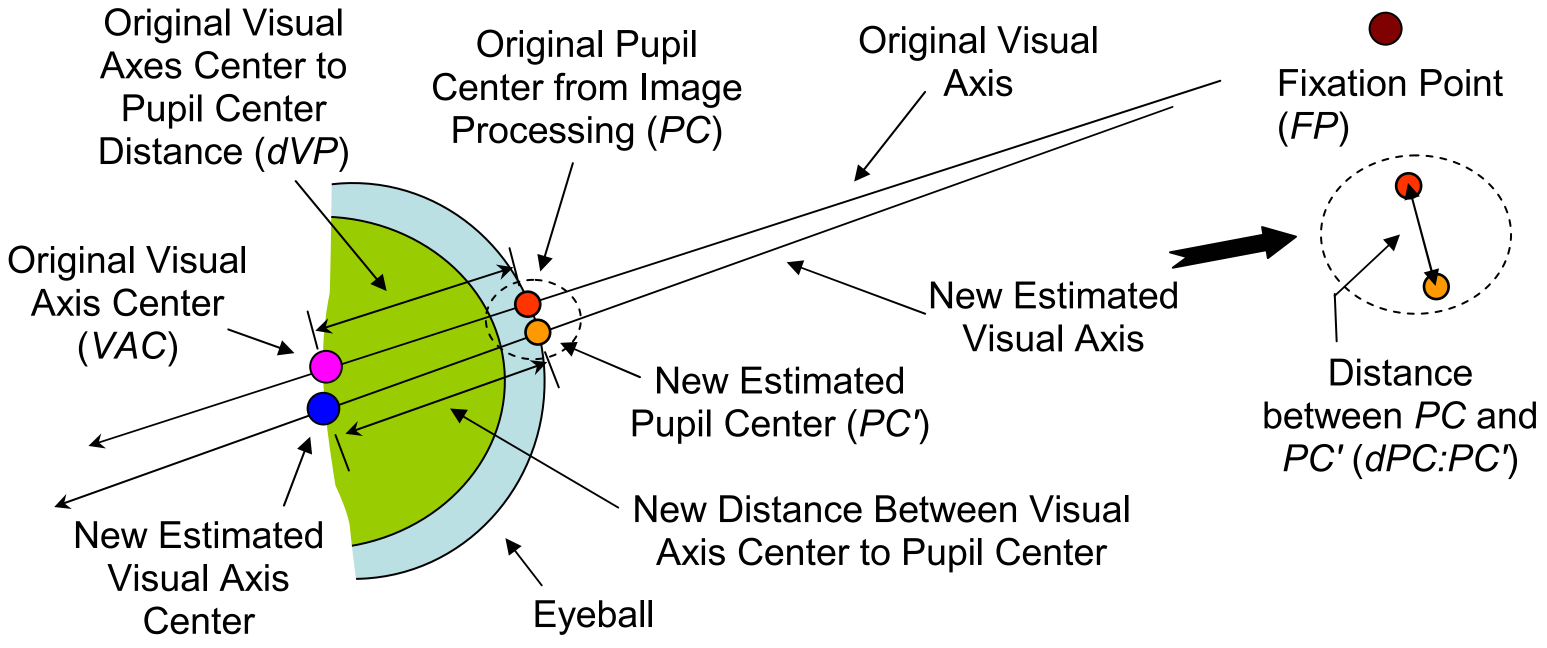

Original Visual Axes Center to Pupil Center Distance (*dVP*)
Original Pupil Center from Image Processing (*PC*)
Original Visual Axis
Fixation Point (*FP*)
Original Visual Axis Center (*VAC*)
New Estimated Visual Axis
New Estimated Pupil Center (*PC'*)
Distance between *PC* and *PC'* (*dPC:PC'*)
New Distance Between Visual Axis Center to Pupil Center
New Estimated Visual Axis Center
Eyeball

Fig. 15 *VAC*/*dVP* adjustment.

3.3 Head Coordinates

While not unique to the visual axes center method, or any other gaze determination method, the ability to determine and to transform to and from a consistent head coordinate system is so central to the visual axes center method, that at least a brief discussion is warranted at this time.

A simple process by which a head coordinate system can be developed and related to the camera coordinate system is based on identifying facial features that remain fixed with respect to the head, collecting images that contain these features, and then consistently defining the coordinate axes of the head coordinate system in terms of these selected facial features located in camera coordinates. Common features obtainable from images and used to define a head coordinate system are the nostrils and the eye corners.

The method proposed for actually defining the head coordinate system requires that at least three facial features (FF_1, FF_2, and FF_3: the same features in each image) be identified so that a plane and a normal to that plane can be consistently defined. This plane will be referred to as the face plane, and is taken to be the X-Y plane of the head coordinate system. The Z axis is taken to be a normal to the face plane.

$$\textit{face plane normal} = Z\ \textit{axis} = \left(FF_a - FF_c\right) \otimes \left(FF_b - FF_c\right) \tag{25}$$

when FF_a, FF_b, and FF_c are chosen from FF_1, FF_2, and FF_3 such that the inner product of the Z axis in face plane coordinates and the Z axis of the camera coordinate system is positive.

The Y axis is defined to be along the line from the centroid of the three facial feature points defining the face plane to one of the facial feature points. The centroid (CD) is specified as:

$$CD = \frac{FF_1 + FF_2 + FF_3}{3} \tag{26}$$

The Y axis is specified by selecting a facial feature point (either FF_1, FF_2, or FF_3) and constructing a vector from the centroid (CD) to the selected feature (FF_2 in this case) as follows:

$$Y\,axis = FF_2 - CD \tag{27}$$

With the Z and Y axes defined, the X axis is becomes the vector resulting from the cross product between the Z and Y axes vectors.

$$X\,axis = (Z\,axis) \otimes (Y\,axis) \tag{28}$$

Because the determination of gaze direction will be partially accomplished in head coordinates and all measured variables are in camera coordinates, it is necessary to determine a transformation between the camera and head coordinate systems. This transformation is constructed using the normalized X, Y, and Z axes of the head coordinate system (see Subsection 2.6.2). Each row of the rotation matrix R is merely the coefficients of the unit vectors representing the X, Y, or Z axis.

$$R = \begin{bmatrix} \dfrac{Xaxis - CD}{\|Xaxis - CD\|} \\ \dfrac{Yaxis - CD}{\|Yaxis - CD\|} \\ \dfrac{Zaxis - CD}{\|Zaxis - CD\|} \end{bmatrix} \tag{29}$$

The translation vector TV is simply the X, Y, and Z components of the face plane centroid.

$$TV = CD \tag{30}$$

The rotation matrix and translation vector can then be applied to the visual axes expressed in 3D camera coordinates (see Eq. 6) to transform them into the head coordinate system representation.

3.4 Visual Axes Center Method Summary

For a particular definition of a head coordinate system and for a particular eye for a given subject, the value of *dVP* and the location of *VAC* remain approximately constant in the head coordinate system as long as no physical changes to the subject occur. Once the head coordinate system is established during actual use, the *VAC* is transformed into camera coordinates. The pupil center (*PC*) is then determined in camera coordinates as described in Subsection 4.3. The resulting vector from the *VAC* through *PC* defines the estimated gaze direction in camera coordinates. One can then use the necessary coordinate system transformations to relate the gaze vector to any needed coordinate system. For purposes of the experiments conducted for this work, the gaze vector was transformed to a monitor coordinate system.

4. 3D FACIAL FEATURE LOCATION WITH A SINGLE CAMERA

The visual axes center method discussed in the previous section relies on the ability during calibration to determine a 3D head coordinate system and to locate the fixation point and pupil centers in this 3D coordinate system. Most often, image features are located in 3D by collecting stereo images of the desired features and using triangulation to determine the 3D feature locations. This will also be the case for the initial subject calibration. However, in actual use, it is desired to use only a single camera, hence eliminating the possibility of using stereo image processing to determine the 3D locations of the facial features and pupils. In this section, a technique for obtaining these 3D locations via a single camera is described.

The method is based upon matching the distances between pairs of facial features calculated during the operational mode with values obtained during the calibration phase.

For every facial feature pair there will be one distance that can be determined during calibration. Therefore, if there are 5 facial features there will be 10 unique pairs and 10 distances ('*m* choose *n*', distances, where *m* is the number of feature points and *n* is the number feature points in a pair: two). If the facial features remain constant with respect to the head, then each of these distances will also remain constant. One must therefore select facial features that remain constant with respect to the head.

4.1 3D Location Determination Assuming Perfect Measurements

For an image from a single camera, any location appearing in that image represents an object that is intersected by a 3D ray originating from the center or origin of the camera coordinate system [91]. Each pixel of an image represents a single, unique ray originating from the center of the camera (actually from the corresponding pixel) and extending out to infinity. Multiple objects that are intersected by a ray will appear on the image plane as a single object (see Fig. 16). The object that will actually appear on the image plane will be the object closest to the camera.

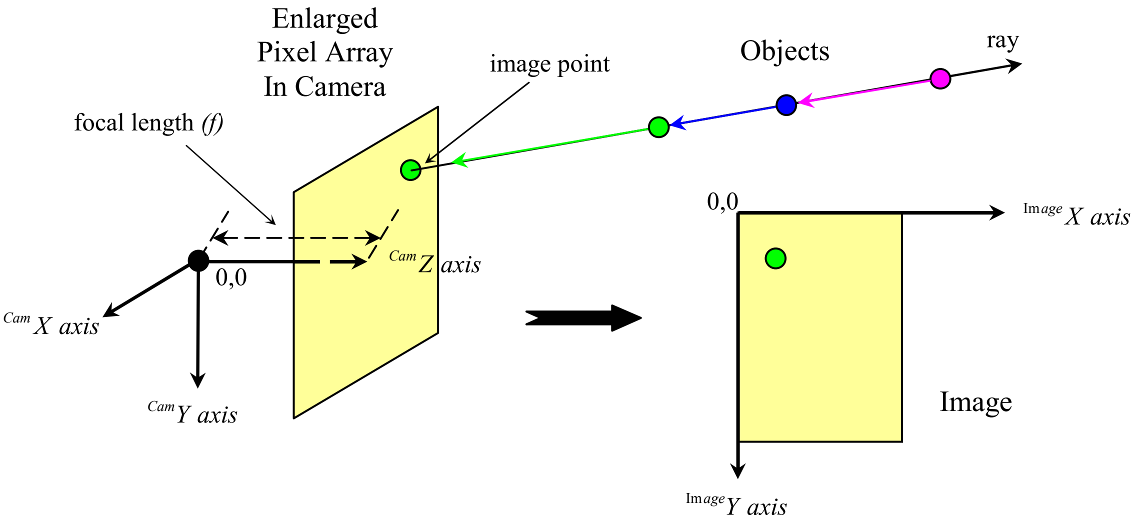


Fig. 16 Pixel ray.

If the camera used to capture the image were 'perfect' (no distortion, skew, etc.), the location of the object in the 3D camera coordinate system could be specified in terms of the *Z* displacement of the object:

$$^{Cam}X = {}^{Cam}Z * \frac{^{Image}X - X_C}{f} \tag{31}$$

$$^{Cam}Y = {}^{Cam}Z * \frac{^{Image}Y - Y_C}{f} \tag{32}$$

where *C* is the displacement between the camera origin and the image origin and *f* is the focal length.

4.2 3D Location Determination with Actual Camera Pixels

However, no camera is perfect. Therefore, to determine the actual camera system locations represented by the pixel $^{Image}X$, $^{Image}Y$, the camera errors inherent in the image must be removed. Matlab initially specifies functions, *g'(), h'(), g(), and h()*, that modify the image pixel locations to account for camera imperfections [86]:

$$u' = g'\left({}^{Image}X, {}^{Image}Y, C, fl, alpha, k_{1\,thru\,5}\right) \tag{33}$$

$$w' = h'\left({}^{Image}Y, C, fl, k_{1\,thru\,5}\right) \tag{34}$$

$$u = g\left(u', w', C, k_{1\,thru\,5}\right) \tag{35}$$

$$w = h\left(u', w', C, k_{1\,thru\,5}\right) \tag{36}$$

where *fl* (the 'focal length' according to Matlab), *alpha* (skew), and k_1 through k_5 (the coefficients of lens distortion) are additional intrinsic parameters output by the Matlab individual camera calibration routines. The variables *u* and *w* in Eqs. 35 and 36 are similar to the $\dfrac{{}^{Image}X - X_C}{f}$ and $\dfrac{{}^{Image}Y - Y_C}{f}$ components of an undistorted pixel in the 'perfect' camera equations and are determined by iterating on equations 35 and 36 a number of times, as follows:

```
Set $X = {}^{Image}X$
Set $Y = {}^{Image}Y$
Set $u' = g'(X, Y, C, fl, alpha, k_{1\,thru\,5})$
Set $w' = h'(Y, C, fl, k_{1\,thru\,5})$
Set $X = u'$
Set $Y = w'$
Set $lBound = 1$
Set $uBound = 20$
Set $increment = 1$
Vary $i$ from $lBound$ to $uBound$ in increments of $increment$
{   Set $X = g(u', w', X, Y, C, k_{1\,thru\,5})$
    Set $Y = h(u', w', X, Y, C, k_{1\,thru\,5})$
}
Set $u = X$
```

Set $w = Y$

Matlab recommends that a *uBound* of 20 is sufficient to yield convergence. The object location equations (Eqs. 31 and 32) then become:

$$^{Cam}X = {}^{Cam}Z * u \tag{37}$$

$$^{Cam}Y = {}^{Cam}Z * w \tag{38}$$

4.3 *Z* Location Determination Using A Single Camera

To determine the 3D location of a feature in an image, one needs to identify the 2D location of the desired feature in the image. Using Eqs. 37 and 38, the *X* and *Y* location of the object in camera coordinates can be determined in terms of its *Z* location in camera coordinates. However, the *Z* location of the object is still unknown.

Given the equations relating *X* and *Y* to *Z*, a set of equations relating the distances between the facial features and the *Z* locations can also be derived. Unfortunately, the equations are non-linear and not amenable to direct solution. To overcome this problem, an iterative optimization using the distances between facial feature pairs determined during calibration can be developed that allows for the determination of the *Z* locations of the feature points in an image.

Given a set of facial feature points (FF_k), where *k* is between 1 and *n* (the number of feature points), denote the distance between each pair of facial feature points FF_i and FF_j by $cdFF_{i,j}$. Assuming *i* is less than *j*,

$$cdFF_{i,j} = ||FF_i - FF_j|| \tag{39}$$

The values for the $cdFF_{i,j}$ determined during calibration are denoted $\overline{cdFF_{i,j}}$. Ideally, for measured points,

$$\overline{cdFF_{i,j}} - \sqrt{\left(X_{FF_j} - X_{FF_i}\right)^2 + \left(Y_{FF_j} - Y_{FF_i}\right)^2 + \left(Z_{FF_j} - Z_{FF_i}\right)^2} = 0 \tag{40}$$

In reality, the difference is unlikely to equal zero. One can create a composite metric of the deviation from zero by aggregating a non-negative measure of the difference across all pairs of facial features. Substituting for *X* and *Y* in terms of *Z* results in a function *J*, such that:

$$J = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\left(\overline{cdFF_{i,j}} \right)^2 - \left(\left[(Z_i * u_i) - (Z_j * u_j) \right]^2 + \left[(Z_i * w_i) - (Z_j * w_j) \right]^2 + \left[Z_i - Z_j \right]^2 \right) \right]^2 \quad (41)$$

where *i≠j* and *i<j*. Since the value of *J* is always greater than or equal to zero (optimally it should be zero), finding the values of Z_i (the unknowns) for *i* = 1 through *n* for a given image at which the minimum value of *J* occurs, provides an estimate of the *Z* locations of the facial features for that image.

In order to calculate the *Z* locations for which the minimum value of *J* occurs, an iterative optimization technique is used. The basic optimization method used is Newton’s method, summarized by:

$$Z_{updated} = Z_{initial} - \left(J''^{-1} * J' \right)_{Z_{initial}} \quad (42)$$

The method requires an initial estimate (guess) for the unknown vector of *Z* locations. Initially, a very rough approximation for a typical distance a person’s head would be in front of a monitor was used. Ordinarily, one would proceed with the initial guess and the optimization until a convergence criteria was reached. Clearly, if *J* = 0, the optimum has been reached. However, this will never occur in practice, and a reasonable criteria for convergence was not available. Moreover, there is a possibility of local minima. Therefore, the initial values for each unknown were simply varied through a range of values and the optimization run for each. The 'optimized' value was the minimum value of *J* over the range of iterations. The average minimum value of *J* was 8726.3. While this value of *J* may appear large when compared to it's optimum of zero, the construction of *J* (see Eq. 41) is such that small differences (even a few tenths of a millimeter) between the calibrated feature distances and those determined from the optimization, result in *J* values with this magnitude or greater.

The following pseudo-code provides a brief explanation of the overall technique, noting that the value of each $\overline{cdFF_{i,j}}$ is actually the average of the $cdFF_{i,j}$ s from the left camera perspective ($\overline{{}^{Left}cdFF_{i,j}}$) and the $cdFF_{i,j}$ s from the right camera perspective ($\overline{{}^{Right}cdFF_{i,j}}$):

```
Set $increment$ = 1.0
Set $threshold$ = 0.0001 (determined by trial and error during calibration)
Set $lBound$ = 400 (the lower iterative bound for $ZV$, estimated to be the smallest comfortable distance at which one's eyes would normally be from a monitor when operating a computer)
Set $uBound$ = 900 (the upper iterative bound for $ZV$, estimated to be the largest comfortable distance at which one's eyes would normally be from a monitor when operating a computer)
Set $J_{min} = \infty$
Set $ZV_{min} = <\infty, \infty, \infty, \infty, \infty>$
Vary $i$ from $lBound$ to $uBound$ in increments of $increment$
{   Set $ZV = <i, i, i, i, i>$
    $counter$ = 1
    while $counter$ is less than 1000
    {   Set $GR$ equal to the value of the first derivative of Eq. 41 with respect to $ZV$ ($GR$ is a 5x1 gradient matrix)
        Set $HS$ equal to the value of the second derivative of Eq. 41 with respect to $ZV$ (HS is a 5x5 Hessian matrix)
        Set $error = HS^{-1} * GR$
        Set $ZV = ZV - error$
        Set $J$ using Eq. 41 and $ZV$
        Set $norm = |GR|$
        if $norm <= threshold$ $exit$ $loop$
        $counter = counter + 1$
    }
    if $J_{min} > J$
    {   $J_{min} = J$
        $ZV_{min} = ZV$
    }
}
```

Upon completion of the optimization, ZV_{min} contains the estimated values of the facial feature Z locations that minimize J. Solving for the facial feature ${}^{Cam}X_i$ and ${}^{Cam}Y_i$ locations by substituting the ${}^{Cam}Z_i$ locations (ZV_{min} values) into Eqs. 37 and 38, results in the determination of all three of the 3D components of the facial feature locations.

Therefore, given an image of a subject's face from a single camera, the intrinsic properties of the camera, and the average facial feature distances determined during calibration, the 3D locations of the facial features (excluding the pupils) can be determined. Then, using the definition of dVP, the average value of dVP determined during subject calibration, and the image pixel location of the pupil center (PC), a quadratic equation relating the Z location of the pupil center to the visual axes center (VAC) and the dVP can be derived:

$$dVP = \sqrt{\left({}^{Cam}X_{VAC} - \left({}^{Cam}Z_{PC} * u\right)\right)^2 + \left({}^{Cam}Y_{VAC} - \left({}^{Cam}Z_{PC} * w\right)\right)^2 + \left({}^{Cam}Z_{VAC} - {}^{Cam}Z_{PC}\right)^2} \quad (43)$$

where all values are known except ${}^{Cam}Z_{PC}$.

Solving this quadratic for ${}^{Cam}Z_{PC}$ and substituting back into the equations relating X and Y to Z (Eqs. 37 and 38) results in the determination of the 3D location of PC. With the pupil center and the visual axes center located in 3D, a gaze vector can then be determined, having been accomplished using a single camera.

5. CONDUCT OF SUBJECT EXPERIMENTS

To test the viability of using the visual axes center methodology (discussed in Section 3) for gaze determination and the use of a single camera for determining 3D locations (discussed in Section 4), a series of experiments were conducted. This section will discuss the experiment protocol and the actual conduct of the experiments. The processing and analysis of the experimental data will be discussed in subsequent sections.

5.1 Experiment Overview

The experiments involved a pool of subjects being asked to look at known locations on a computer monitor while images were collected of them doing so. Based on the guidance provided by Ostle and Mensing [92], it was determined that a minimum of 30 subjects were needed in order to bestow statistical significance to the results. Therefore, it was decided to use a sample size of at least 30.

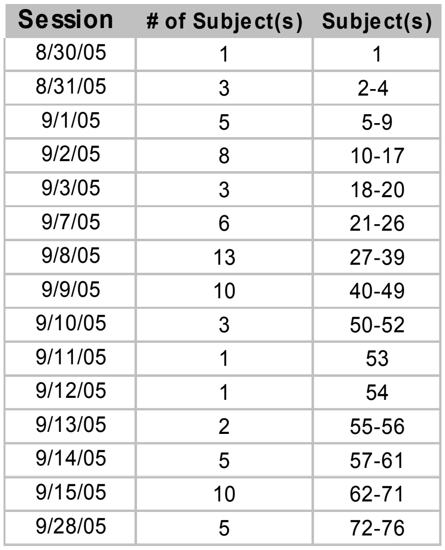
After image collection of both test images (single camera) and calibration images (multiple cameras), the subjects' gaze was then determined from the images and compared with that determined from the location on the monitor at which they reported they were looking. An assessment of the gaze determination accuracy of the visual axes center method, as well as a comparison of the results between finding 3D locations using pseudo-stereo triangulation and single camera 3D optimization was then made.

Because of the fact that human subjects were involved, an approval from the Institutional Review Board was obtained prior to conducting any experiments. This approval was granted on July 18, 2005 under protocol number 2005-0364. This protocol was amended on June 20, 2006 to allow an additional period before destruction of the subject images is required. The remainder of this section discusses the approved experiments, as they were conducted.

5.1.1 Subject Selection

Subjects for the experiments were drawn from Dr. Hall's Educational Psychology class (ESPY 435) at Texas A&M University during the Fall 2005 semester. Participation was voluntary, but extra credit for the course was offered for participation. Students were asked to sign-up for a 30-minute time slot by reviewing an on-line schedule and submitting an email request for the time they desired. The original duration of the conduct of the all the experiments was to have been three weeks. As mentioned previously, at least thirty subjects were desired to participate during this timeframe. A significantly greater number of subjects participated during the initial phase. However, due to the addition of hurricane Katrina refugees to the class, the experiments were continued through September. A total of 76 students participated during the entire experiment period from 8/30/05 through 9/28/05 (see Table 1). The experiments were conducted in Dr. Volz's storage room in the H.R. Bright Building (HRBB 311-A) on the Texas A&M University College Station campus.

Table 1 Experiment session summary.



Session	# of Subject(s)	Subject(s)
8/30/05	1	1
8/31/05	3	2-4
9/1/05	5	5-9
9/2/05	8	10-17
9/3/05	3	18-20
9/7/05	6	21-26
9/8/05	13	27-39
9/9/05	10	40-49
9/10/05	3	50-52
9/11/05	1	53
9/12/05	1	54
9/13/05	2	55-56
9/14/05	5	57-61
9/15/05	10	62-71
9/28/05	5	72-76

5.1.2 Physical Experiment Setup

The experimental setup included a single computer system connected to three Veo Velocity Connect universal serial bus (USB) webcams (1280x1024). An additional nineteen inch cathode ray tube (CRT) monitor simulating the monitor that would be used during an application was used as the object for subjects to view. The additional monitor remained powered off during the experiments. The unpowered monitor had eight, approximately 3/32" diameter green, adhesive dots placed around the monitor on the case near the edge with the screen and another dot placed on the center of the screen. These dots were the target points the subjects would look at.

The cameras were mounted under the bottom of the unpowered monitor, but were connected to the USB ports on the functional computer. A chair was placed in front of the table on which the unpowered monitor with the 'target' dots was placed so as to provide each subject with a view of the monitor similar to what would be expected if they were actually using the monitor to interface with a computer. The functioning computer was placed to the left of the unpowered monitor. It was clearly in the peripheral view of the subjects during the experiment, but was oriented so as to minimize subject distraction.

Because of the close quarters of the room in which the experiments were conducted and the non-adjustable fluorescent lighting present, several cloth 'drapes' were hung to prevent glare. One drape was hung from approximately eight feet above the floor to the floor on the right side of the subjects to prevent glare on the subject. A second drape was hung from the ceiling to about three feet below the ceiling between the computer and the inactive monitor to prevent glare into the cameras and onto the subject. These drapes were not adjustable and remained in a constant location relative to the room's lighting throughout the conduct of the experiments.

The computer was running under the Windows XP operating system, and the image collection routines were modifications of C++ capture routines described by Laganiere [93] that would allow full resolution, color images to be collected from each of the three cameras. The remainder of the experiment routines was written in Sun Java. The images were collected in individual three-image sequences from the left, then the right, then the middle camera when facing the monitor. It took between four and six seconds to capture a three-image sequence. The collection of each three-image sequence was manually initiated, and for the subject experiments, was initiated at the direction of the subject.

5.1.3 Experiment Protocol

This subsection contains an outline of the experiment protocol. The outline is presented in phases: Preliminary, Hardware Calibration, and Experiment. It is a slightly expanded representation of the information used by the test conductor (the author of this dissertation) to actually conduct the experiments. References to other portions of this dissertation contained in the outline were not present during the conduct of the experiments.

Preliminary

a. Upon arrival (of the experiment conductor), power up the computer if necessary
b. Review the schedule of prospective subjects and ensure that enough green dots for facial feature markers are available and marked with a black dot
c. Ensure that the monitor location diagram (see Fig. 21) is visible to the right and under the unpowered monitor on the table
d. Open a DOS window and execute the command: mkdir 'D:\Research\Current\???', where '???' is the current date (i.e. 8-31-05)
e. Execute the command: xcopy /s "D:\Research\Current\Data Template\" "D:\Research\Current\???" (copy the files/file structure necessary to conduct the experiment)

Hardware Calibration (see Subsection 5.2)

a. Move to the D:\Research\Current\???\CamCal folder

b. Place the camera calibration checkerboard tripod in front of the cameras and position the tripod feet to match the tape markers on the floor
c. Attach the checkerboard to the tripod and ensure that tripod feet are still correctly positioned
d. Open the Veo camera video application, and looking at the video of the checkerboard from the perspective of the middle camera, position the monitor (tilt/rotate, do not move the base) such that the checkerboard is approximately centered in the video
e. Close the Veo video application
f. Execute the command: java MultipleImageCollect "D:\VidCapture\VidCapture.exe" "CamCal" "20" "D:\Research\Current\???\CamCal" (capture multiple checkerboard images)
g. Execute the command: del *.ppm (delete the captured images after conversion to a format Matlab accepts)
h. Execute the command: cd ..
i. Check the images using Matlab to ensure all are appropriately focused and visible
j. Remove the checkerboard from the tripod and store for their next use
k. Affix the wooden calibration rig to the unpowered monitor
l. Execute the command: cd ToolCal
m. Execute the command: java SingleImageCollect "D:\VidCapture\VidCapture.exe" "ToolCal" "D:\Research\Current\???\ToolCal" (capture image of wooden calibration rig)
n. Execute the command: del *.ppm
o. Remove the wooden rig from the unpowered monitor taking care not to impact the cameras
p. Ensure that the 'Testing In Progress' sign is posted on the entrance to experiment room

Experiment

a. Prior to the arrival of each subject:
    i. Ensure that a Consent Form (see Appendix 1), a data sheet (see Appendix 2) with the date and subject number completed, and five marked, green dots are available.
    ii. Execute the command: mkdir 'D:\Research\Current\???\UserXX', where 'XX' is the subject number
    iii. Execute the command: xcopy /s "D:\Research\Current\Data Template\User" "D:\Research\Current\???\UserXX"
    iv. Execute the command: cd ..
    v. Execute the command: cd "D:\Research\Current\???\UserXX"
b. Upon arrival of each subject, ask them to be seated, provide them with a copy of the Consent Form, and ask them to read and sign the form
c. Determine if the subject is familiar with the notion of a dominant eye and whether they are left or right eye dominant. If they are not certain about their eye dominance, perform the 'thumb' test to determine it (closed eye with most movement is dominant, see Subsection 5.4.1 for explanation of determination method)
d. Record eye dominance on the data sheet
e. Determine if the subject is wearing eyewear (glasses or contacts). If the subject is wearing glasses, offer to perform the test with or without glasses. Explain that there may be some likelihood of poor results with glasses, but wearing glasses is preferred
f. Ensure that the subject can read the monitor point diagram and can see the green dots on the monitor
g. Record use of eyewear on the data sheet
h. Have the subject affix the marked, green dots to their face in the approximate locations corresponding to those described by the test conductor.

i. Ensure the subject's comfort with the dots and their seating location/position

j. Open the Veo camera video application, and looking at the video of the subject from the perspective of the middle camera, position the monitor (tilt/rotate, do not move the base) such that the subject's face is approximately centered in the video. Care should be taken not to impact the cameras

k. Ensure the subject has an opportunity to see their image with the green dots in place

l. Close the Veo video application

m. Ensure the subject understands their required actions during the collection of the images in the next step (see Subsection 5.4.1 for a discussion of the actual instructions) particularly the viewing requirements and the recording of viewing locations on the data sheet

n. Execute the command: java MultipleImageCollect "D:\VidCapture \VidCapture.exe" "UserXXStare" "27" "D:\Research\Current\??? \UserXX" (allow the individual capture of subject experiment images based on input from the experiment conductor)

o. Execute the command: del *.ppm

p. Ensure that all data points were recorded on the data sheet

q. Ensure the subject understands their required actions during the collection of the images in the next step (see Subsection 5.4.1 for a discussion of the actual instructions) particularly the viewing requirements and the recording of viewing locations on the data sheet

r. Execute the command: java MultipleImageCollect "D:\VidCapture

s. \VidCapture.exe" "UserXXGlance" "9" "D:\Research\Current\???

t. \UserXX" (allow the individual capture of subject calibration images based on input from the experiment conductor)

u. Execute the command: del *.ppm

v. Ensure that all data points were recorded on the data sheet

w. With the subject watching, ensure that the required images were acceptable

x. Ask the subject to remove the green dots on their face. The test conductor shall verify that the dots are removed

y. Ask the subject to access their discomfort level during the experiment, and record it on the data sheet

z. Ask the subject to provide any comments or observations they may have on the back of the data sheet

aa. Ensure that the subject's questions have been answered, and release them from the test area

The execution of the experiment protocol will be discussed in the remainder of this section.

5.2 Camera Calibration Image Collection

Camera calibration images were collected in order to facilitate the determination of the intrinsic and extrinsic parameters of each camera using a geometric technique similar to the one described in Subsection 2.6. The actual camera calibration image collection consisted of collecting multiple three-image sequences (left, right, and middle camera images) of the checkerboard depicted in Fig. 10 mounted to a tripod. The pitch and yaw of the checkerboard were varied using tripod adjustments to collect 20 unique-orientation, three-image sequences. These 20, three-image sequences represented a single set of camera calibration images. Although these 'checkerboard' images were used as part of the extrinsic camera parameter determination in that they facilitated stereo triangulation, their primary purpose was for intrinsic camera parameter determination.

With the exception of 9/5/05, each day of experiments included a collection of a set of camera calibration images: one set per day. For the noted exception, the camera calibration and 'tool' calibration (see Subsection 5.3) were not performed because of a schedule misunderstanding. A subject arrived at the test location, but had not scheduled a time. Despite the fact that other commitments had been made that precluded calibrations either before or after, it was decided to perform the experiment for this subject anyway, rather than making the subject re-schedule and return on another day. Because of the averaging of calibration sets discussed later (see Subsection 7.2), the impact of not having this calibration data was negligible.

The purpose of performing multiple camera calibrations or having multiple sets of camera calibration images was twofold. Although any set of calibration images was believed to be sufficient to calibrate the cameras, having multiple sets provided the ability to assess if problems with the camera focus or internal circuitry had occurred from one day to the next. In addition to the ability to detect a problem, having routinely re-calibrated all the hardware (including the cameras) would facilitate the usage of all images collected after the anomaly/change had occurred by simply using a calibration set conducted after the discrepancy occurred. At most, only the number of subject experiments collected since the previous calibration would be suspect.

Most often, camera calibration image collections occurred prior to subject experiments being conducted. However, on the first day of experiments (8/30/05), calibration image collection was performed both before and after subject experiments were conducted, and on 9/13/05 and 9/28/05 calibration image collection was conducted between Subjects 55/56 and Subjects 72/73 respectively. However, because of the calibration averaging technique that would ultimately be used to determine both intrinsic and extrinsic camera parameters (see Subsection 7.2), the ordering of the collection of camera calibration image sets was determined to be unimportant.

5.3 Camera/Monitor Calibration Image Collection

As with the camera calibration, the camera/monitor calibration images were collected routinely (with the same frequency as the camera calibration) so as to minimize the loss of experimental data should an anomaly occur. Each day, after camera calibration images were collected, a single three-image sequence (one image from each camera) of a monitor rig (see Fig. 17 and Fig. 18) attached to the unpowered monitor was also collected. The use of this rig (rig locations were known in the monitor coordinate system) was to allow for the extrinsic parameters of the camera to be determined in a non-camera coordinate system of interest (ultimately the monitor coordinate system: see Subsection 2.6.2) by creating a relationship between the camera and the monitor/rig. Any one of these daily three-image sequences, along with the camera calibration images, was thought to be sufficient to determine the transformations between the cameras and the monitor coordinate systems.

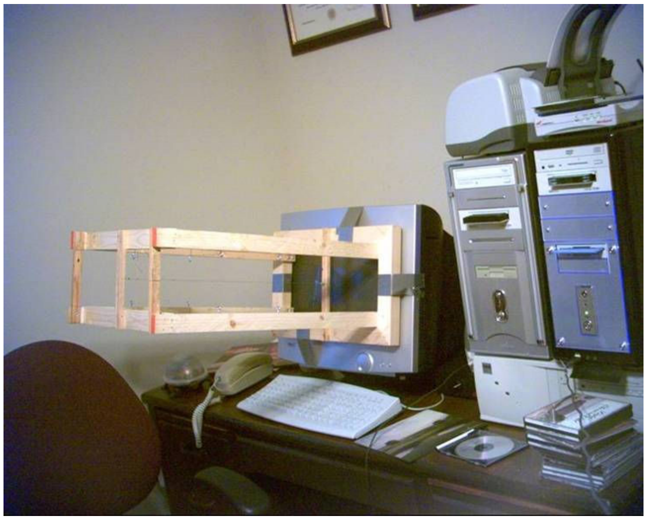


Fig. 17 Monitor with camera/monitor calibration rig attached.

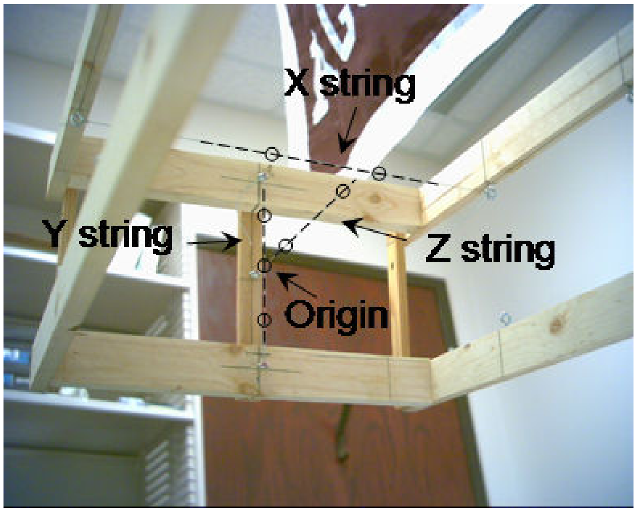


Fig. 18 Camera/monitor calibration rig from a webcam perspective.

The camera/monitor calibration (wooden rig) images were collected after the camera calibration images in an attempt to ensure that any change in camera position caused by attaching the rig to the monitor would be adequately reflected in the extrinsic parameters. It was assumed that the intrinsic parameters of the cameras (those parameters determined using only the 'checkerboard' images) would not be adversely impacted by installation and removal of the wooden rig (the focus ring of the camera was taped in a fixed position and the case of each camera was designed to protect the camera electronics). However, the potential adjustment of the monitor position so that the 'checkerboard' was entirely in the view of each of the cameras was thought to include some risk of moving one or more of the cameras in relation to the monitor. By doing the checkerboard test first, any camera movement would not matter as long as the cameras remained fixed thereafter. While this ordering did not address the possibility of invalidating the camera/monitor calibration due to monitor rig removal (it was assumed that removal of the rig was more likely to be accomplished without impacting the cameras than the installation), it was believed to provide the greatest likelihood of a successful camera and camera/monitor calibration.

The rig used was constructed so that it would attach in a pre-determined spatial relationship with the monitor in the same location with respect to the monitor each time it was attached. The *X*, *Y* , and *Z* axes of the rig coordinate system were intended to be parallel to the corresponding axes of the monitor coordinate system (the rotation matrices between the cameras and either the rig or the monitor were the same). The origins of the rig and monitor coordinate systems differed by approximately <6.03 mm, -139.76 mm, -539.38 mm> (translation vector *TV* from rig to monitor). This translation vector was determined by averaging the results of three measurements. Each measurement involved attaching the rig to the monitor and measuring the distance between the rig and monitor origins using a taught string and either a micrometer and/or a ruler, depending the distances involved. For the three measurement trials, no dimension varied by more than two millimeters. Unfortunately, there was no plausible method devised to determine the angular discrepancies between the rig and monitor whether the rig was attached to the monitor or not.

However, the possible errors resulting from rig construction and use manifest themselves in the monitor to camera transformation. These errors are further transmitted to any other coordinate system to which vectors are transformed using this transformation. There is insufficient data to quantitatively determine an error bound. Moreover, an exact error analysis is quite complex. However, a very crude rig error estimate and plausibility argument for a small impact of the errors can be given, and the end to end errors from the experiment described in Section 7 can be used to argue that the impact is acceptably small.

Assuming no errors other than those from the rig measurement/construction are present, let ${}^{Cam}TP$ be a valid target point in camera coordinates. The transformation to a valid target point in monitor coordinates is represented by:

$$ {}^{Mon}TP = \left[{}^{Cam}TP * {}^{Rig}_{Cam}R + {}^{Rig}_{Cam}TV \right] + {}^{Mon}_{Rig}TV \tag{44} $$

The transformation involving rig construction and use errors is represented by:

$$ {}^{Mon}TP_{err} = \left[{}^{Cam}TP * {}^{Rig}_{Cam}R + {}^{Rig}_{Cam}TV \right] * {}^{Mon}_{Rig}R_{err} + {}^{Mon}_{Rig}TV_{err} \tag{45} $$

One can easily show that for any point ${}^{Mon}TP$, there is an apparent error Δ_{err} such that ${}^{Mon}TP + \Delta_{err}$ is the point at which TP would appear to be:

$$ {}^{Mon}TP_{err} = {}^{Mon}TP + \Delta_{err} \tag{46} $$

Δ_{err} cannot be directly bounded because quantitative measurements on the rig rotational errors are not available. However, it can be argued that that a one degree rotational error bound is likely. The rig is constructed to exactly fit in the physical frame of the monitor, with the origin of the rig approximately at the center of the monitor. Thus, using approximate monitor dimensions of 15" wide by 11" tall, Eq. 47, and assuming the monitor screen is flat, it is approximately 13.31 inches from the origin (bottom, center) of the monitor to the top corners of the rig:

$$distance = \sqrt{(11")^2 + (7.5")^2} = 13.31" \tag{47}$$

A one degree rotational error at a distance of 13.31 inches would produce a translational shift of approximately:

$$shift = 13.31" * 1° * \pi \text{ radians} / 180° = 0.23" \tag{48}$$

This is certainly a large enough error to have been noticed, and no such deviations were noted. Thus, in the subsequent discussion, a conservative one degree rotational error will be used.

Now consider the pair of lines from *TP* and $TP + \Delta_{err}$, respectively through the pupil center (*PC*). Suppose the distance from *TP* to *PC* is 635 millimeters (comparable to what was observed in the experiment: see Fig. 19) and measurements of ${}^{Mon}_{Rig}TV_{err} - {}^{Mon}_{Rig}TV$ suggest that the translation error was no more than two millimeters in any one dimension. As noted above, the rotation error component will be taken to be one degree. Because rotations involving the displacement of the *Z* axis would only minimally affect resulting gaze angle errors, the worst case would be if the rotation were to occur about the *Z* axis.

As the most distant target point (*TP*) on the monitor (one of the corners) is about 13.31 inches, the distance error produced by a one degree error is 0.23 inches, as determined from Eq. 48. Converting 0.23 inches to millimeters results in a distance error of approximately 5.8 millimeters. Assuming ${}^{Mon}TP$ is the top right corner of the monitor, ${}^{Mon}TP$ is <190.5, 279.4, 0>. Given a positive rotation error of one degree about the *Z* axis, the apparent location of *TP* after accounting for the rotational error would be <185.7, 282.5, 0>:

$$X = \cos\left(1^o + \cos^{-1}\left(\frac{7.5"}{13.31"}\right)\right) * 13.31 = 7.31 \text{ inches} \approx 185.7 \text{ mm} \tag{49}$$

$$Y = \sin\left(1^o + \sin^{-1}\left(\frac{11"}{13.31}\right)\right) * 13.31 = 11.12 \text{ inches} \approx 282.5 \text{ mm} \tag{50}$$

If it is assumed that the rotation and translation are such that the errors are additive and that the translation error was a worst case value of <-2, 2, -2>, then the worst case ${}^{Mon}TP_{err}$ would be <183.7, 284.5, -2> and Δ_{err} is:

$$\Delta_{err} = <(183.7 - 190.5), (284.5 - 279.4), -2> = < -6.8, 5.1, -2 > \tag{51}$$

which represents a distance error of approximately 8.8 millimeters. Assuming that in the worst case ${}^{Mon}PC$ is located at <190.5, 279.4, 635>, the distance error translates into an angular error between ${}^{Mon}PC - {}^{Mon}TP$ and ${}^{Mon}PC - {}^{Mon}TP_{err}$ of 0.68 degrees:

$$\textit{angular error} = \cos^{-1}\left(\frac{\left({}^{Mon}TP - {}^{Mon}PC\right) \bullet \left({}^{Mon}TP_{err} - {}^{Mon}PC\right)}{\left\|{}^{Mon}TP - {}^{Mon}PC\right\| * \left\|{}^{Mon}TP_{err} - {}^{Mon}PC\right\|}\right) = 0.68° \tag{52}$$

From the experiments, the VAC was, on average, 10.9 millimeters from the pupil center. Therefore, the worst case distance between ${}^{Mon}VAC$ (on the line from ${}^{Mon}TP$ to ${}^{Mon}PC$) and the line through ${}^{Mon}TP_{err}$ and ${}^{Mon}PC$ would be 0.13 millimeters:

$$\textit{distance} = 10.9\text{mm} * 0.68° * \pi \text{ radians} / 180° = 0.13" \tag{53}$$

This distance error is negligible as long as the target points used for the calibration to determine VAC are sufficiently far apart.

While the previous discussion is hardly a conclusive error analysis, it is sufficiently plausible to justify the conduct of the experiment: the end to end results of which are consistent with these approximations and show an acceptable end to end error.

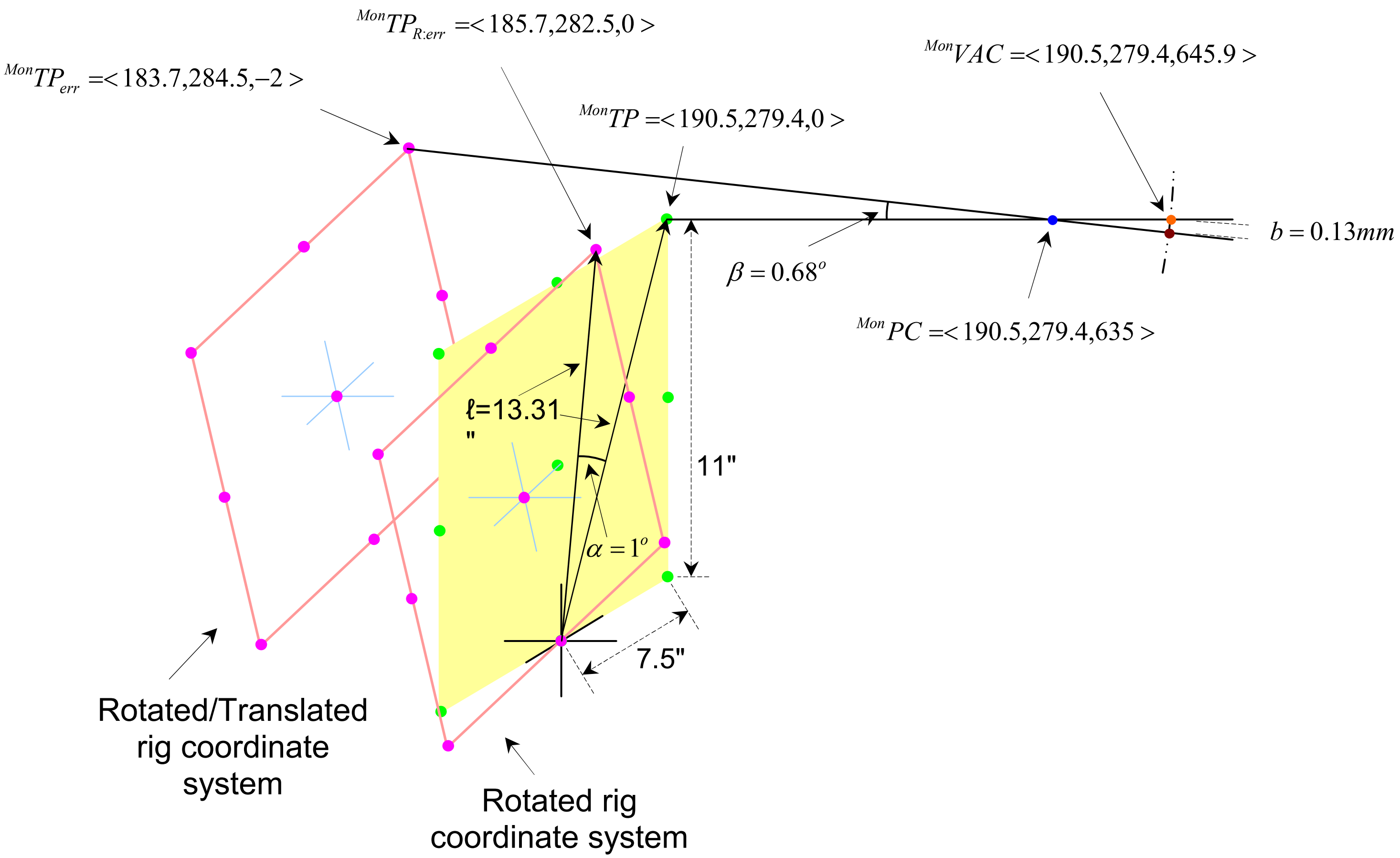

$^{Mon}TP_{R:err} =< 185.7,282.5,0 >$
$^{Mon}TP_{err} =< 183.7,284.5,-2 >$
$^{Mon}TP =< 190.5,279.4,0 >$
$^{Mon}VAC =< 190.5,279.4,645.9 >$
$b = 0.13mm$
$\beta = 0.68^o$
$^{Mon}PC =< 190.5,279.4,635 >$
ℓ=13.31
"
11"
$\alpha = 1^o$
7.5"
Rotated/Translated rig coordinate system
Rotated rig coordinate system

Fig. 19 Rig/monitor rotation and translation errors.

5.4 Individual Subject Experiment Image Collection

Prior to the arrival of the first subject each test day, the computer was powered up (if required) and the appropriate directory structure on the hard drive was created for that day of experiments. It remained powered up at least until the completion of all experiments for that day. Upon arrival of the first subject to the testing area, a sign indicating an experiment was in progress was posted outside of the entrance to the area. This sign was removed whenever there was no ongoing experiment and at the end of each day. The powering up/down, the in-progress posting, and the camera and camera/monitor calibrations were the only activities that were not repeated for each subject. The following was the general flow of activities associated with an individual subject experiment regardless of when the camera and camera/monitor calibration images were collected.

5.4.1 General Flow

Upon being seated in the testing area, each student was asked to read, sign, and date an Informed Consent form (see Appendix 1). Prior to signing the form, each student was given an opportunity to ask questions, was briefly told what the experiment would entail, and their willingness to participate was verbally verified. After signing the consent form, each student was assigned a unique identification number. The id was recorded on a data sheet (see Appendix 2), as was the fact as to whether the subject was wearing prescription glasses, contact lenses, or no corrective lenses. Those subjects who were wearing prescription glasses were informed that there was some possibility that their results would not be usable, but they were told the data collected would be valuable as a comparison and were encouraged to continue. Some offered to remove their glasses, but all agreed to continue with glasses even if their data might not be useable. After recording the prescription eyewear status on the data sheet, the sheet and a pen were given to each subject to record the remainder of the experiment data.

Next, because the opportunity presented itself and it was unknown what role the notion of eye dominance may play in determining gaze, each subject was asked whether they were left or right eye dominant. Unfortunately, most did not know. Therefore, a brief explanation of the concept of eye dominance was discussed, and then a simple dominance test was conducted. The test consisted of each subject being asked to place there hands together, interlocking their fingers while extending their thumbs upward, and extending their arms outward in front of them. They were then asked to align their thumbs with some target directly in front of them while keeping both eyes open. At this point, they were then asked to close alternate eyes and determine with which eye closed their thumbs moved most from the target. They were informed that the eye that was closed when their thumbs moved the most was their dominant eye. Most seemed to enjoy the activity and were surprised at the results.

After completing the dominance test, each subject was given a set of five, 1/4" green adhesive dots that had the approximate center marked with a permanent marker. They were asked to place the green dots on their face: one below the hair line, one on the bridge between the eyes, one on the tip of their nose, and one centered below each eye (see Fig. 20). A compact mirror was at their disposal to assist in placing the dots on their face. Surprisingly, most had little trouble completing the task. They were then asked to position themselves comfortably in their chair in front of the un-powered monitor as if they were going to be using it. Each subject was allowed to view a live video of themselves with the green markers attached while the camera/monitor assembly was adjusted so that their face was reasonably centered in the middle camera's view. Most seemed comforted by the opportunity to view what they would look like before images were actually collected.

The next phase was the actual collection of experiment images. Each subject was asked whether they could see the 3/32" dots affixed to the monitor and if they could see the numbering scheme assigned to the dots on the table to the right of the monitor (see Fig. 21). Given a positive response, the numbering of each monitor dot was reviewed.

Had subject's not been able to see the dots, the experiment would have been terminated. However, all subjects were able to clearly see the dots.

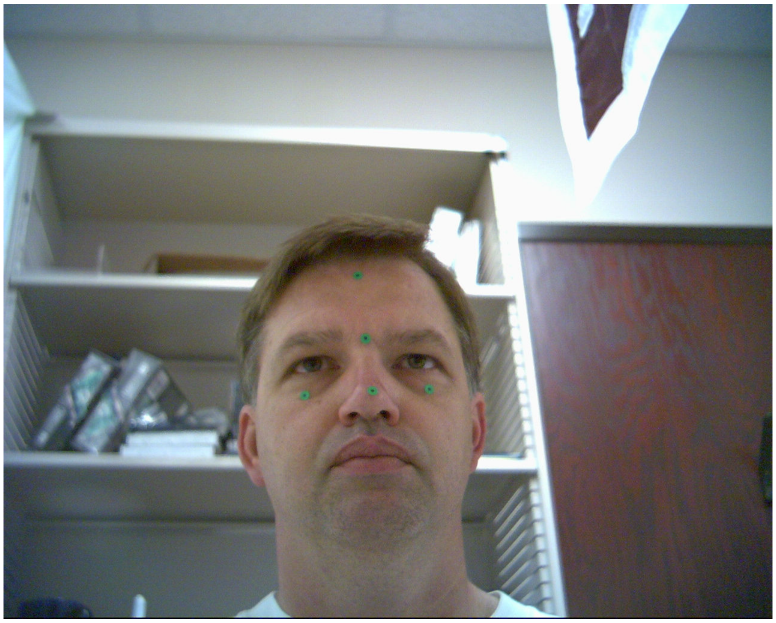


Fig. 20 Green markers.

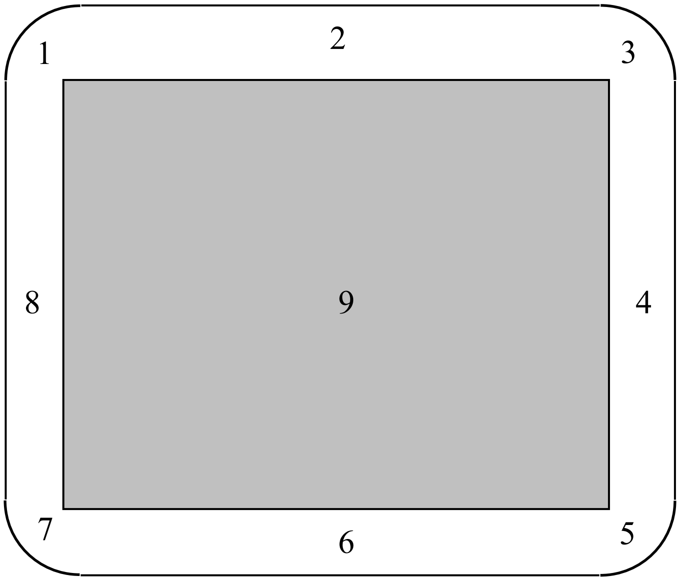


Fig. 21 Monitor dot numbering.

Each subject was informed that this phase would be the longest portion of the experiment. It was explained that during this phase they would be required to look at each monitor dot a total of three times: a total of 27 images being collected (the subjects were not told that 81 images were actually being collected). They were told that they could look at the dots in any order that they wished in any manner that was comfortable. There were no restrictions given with regard to body or head movement. However, they were instructed that once they had decided on a dot to look at, they should blink a few times and relax before focusing on that dot, because they would have to maintain their focus on that dot without blinking or moving for approximately five seconds.

After selecting and focusing on a dot, they were to issue a verbal 'go' or 'start' command to the test conductor. Upon receiving this command the test conductor would depress the 'Enter' key on the keyboard, resulting in the capturing of the subject's image. Once a 'go' command was issued, a subject was required to maintain their focus on the dot without moving their head or blinking until a verbal 'ok' or 'stop' command had been issued by the test conductor.

After receiving an 'ok' or 'stop' notice, the subject was to immediately write down on the data sheet the number of the dot they had just looked at. They were also instructed to evaluate how well they had remained focused, not moved, and not blinked and record this self-evaluation on the data sheet. After completing the data sheet for that focus/image, they were to immediately select another dot (it may be the same dot if they wanted) and issue a 'go' command when they were focused and ready.

Each student was informed that this 'focus, start, hold, and record' process would continue until 27 focuses/images were collected. Ensuring their comfort as a priority was stressed and they were reminded that the pace of this portion of the experiment was totally under their control. Upon completion of this phase, each subject was allowed to rest and given an opportunity to ask questions or make comments. During this rest period, the actual images were being reformatted and moved to another location on the hard disk. With the completion of the image file relocation (assuming all questions had been answered), the last image collection phase of the experiment was initiated.

For the final image collection phase of the experiment (the actual subject calibration portion of the experiment), each subject was informed that they would be looking at each of the same nine dots as before, but this time they would look at them in numerical order starting with dot 1. In addition, they were instructed that they would be looking at each dot as if they had positioned their head and body to look at the center dot (number 9), and then moved their eyes to focus on the appropriate dot, keeping their head and body still. Because of the additional effort required to hold their gaze when looking at any dot other than number 9, the need to relax and prepare before proceeding was stressed. Subjects were informed that the 'start' 'stop' command sequencing similar to that of the previous phase between themselves and the test conductor was still in effect. However, it was suggested that between looking at the different dots, the subject should return their focus to the center dot (number 9) each time and rest before moving their focus to the next dot and saying 'go.'

After each 'stop' command had been acknowledged by the subject, the subject was to circle on the data sheet the number of the dot of any focus/image for which they think they moved, blinked, or looked away during the 'start' 'stop' (image capture) periods. They were informed that the image for each number circled during this phase would be reviewed at the end and images re-captured for that particular focus dot if necessary. The importance of their adherence to the requirements during this phase was re-stressed. Upon completion of this phase, another opportunity for questions was provided while the images were transferred to the appropriate hard disk location.

The next to the last phase of the whole experiment involved a review of any items circled during the previous (last) image collection phase. If an item was circled, it was reviewed individually for image blurriness and subject eye blinking. Surprisingly, very few items were circled, and none of the images whose number had been circled showed visible evidence of problems. In addition to the individual review of circled items, each subject was allowed to view the nine images from the middle camera that were collected as part of the final image collection phase as a pseudo animation. This was accomplished by simply opening all of the nine stored images all at once in Microsoft's

Photo Editor (which overlays each open image window on top of the previous one) and rapidly closing the individual image windows: producing a simplistic animation. A similar animation of a 27 image sequence from the actual experiment phase was also presented to each subject. The animations not only provided a brief period of entertainment for the subject's, it more importantly provided the test conductor an opportunity to quickly review the images for gross errors. Had errors been detected and time permitted, the suspect portions of either of the image collection phases could have been repeated. Unfortunately, for the one case where significant errors were observed during this animation (Subject 59's top dot was out of the field of view of the middle camera for 16 of the 27 images during the first image collection phase), there was not enough time to repeat the necessary portions of the experiment. This subject's data that was out of the field of view was not considered for the analysis portions of this dissertation.

After the images were reviewed, each subject was asked to remove the adhesive markers from their face. Upon verifying the removal, they were asked to evaluate on the data sheet their level of discomfort throughout the entire experiment. They were then asked if they had any written or verbal questions or comments regarding their experience. Upon completing the documentation of their thoughts, each subject was thanked for their participation, assured they would receive credit, and allowed to leave the test area. The entire experimental process was usually completed in 15 to 20 minutes. One subject completed in 11 minutes and one subject took 28 minutes.

5.5 Protocol Modification

During the conduct of approximately five days worth of experiments (through Subject 20), it was noticed that a significant number of subjects were remaining unusually rigid (not moving their body or head) during the first image collection phase when changing their focus from one monitor location to another. While there were no restrictions placed on their movement during this phase, they appeared to believe that they must maintain the same body and head position throughout the entire phase, not just while the images were being captured. In an attempt to ensure that more natural movement was encouraged, a set of verbal instructions were given to a majority of the subjects (starting with Subject 21) as an initial step of the first image collection phase. Just as a control, some subjects after Subject 20 (Subjects 27 and 62 through 74) were not given the instructions. The issuance of these additional instructions was recorded on each subject's data sheet (see Appendix 2) as appropriate.

These additional instructions consisted of a demonstration of the difference between 'staring' at an object and 'glancing.' Staring was loosely defined as ensuring your head and body were positioned in front of or 'square' with the object being looked at. Establishing a 'stare' at a different object would normally require some head or body movement. This was contrasted with 'glancing,' which was described as moving the eyes to look at different objects and keeping the head and body relatively still. Subjects were encouraged to view the monitor locations in any manner they saw fit, but were also asked to remember that they were under no restrictions with respect to head or body movement when transferring their focus to a new location during the initial (test) image collection phase.

6. CALIBRATION AND EXPERIMENT IMAGE PROCESSING

For purposes of evaluating the performance of the visual axes center method and the single camera 3D optimization, it was desired to minimize the number of error sources not specifically related to either method. Therefore, it was decided not to attempt to implement any real-time image processing for the experiments beyond what was inherent in the Matlab camera calibration routines [86]. In addition, outside of those checkerboard features needed for individual camera calibration (the corners of the checkerboard squares), it was decided to locate all the image features manually. While there is some additional error potential associated with manual feature location specifically related to the performance of the individual locating the features (see Subsection 7.1), it was assumed that the incorporation of automated processing would not only require significant development effort and introduce its own error potential, but its incorporation was not pertinent to the evaluation of the methods themselves. It was also feared that errors resulting from automated feature extraction could not be identified and segregated from those errors specifically related to the methods being examined. The remainder of this section will discuss the image processing performed on the images collected during the experiment calibration and subject testing described in the previous section.

6.1 Camera Calibration Processing

The images collected of the camera calibration checkerboard from all three cameras were processed using Matlab camera calibration routines [86]. The intrinsic parameters were determined for each camera independently. A large rectangle was chosen that was visible in all three of the camera images. The bottom left corner of the chosen rectangle was taken as the origin. The corners were marked on the image. Matlab was told to search for the intersections initially within an 11 x 11 pixel region around the marked locations.

Matlab routines then determined an initial set of intrinsic parameters for each camera: an intrinsic parameter estimate if you will. Matlab suggests trying different sized regions around the marked points using a recalculation (suggesting some iterative estimation of the parameters). Thus, the recalculation was repeated several times with window sizes of 5 x 5, 3 x 3, and then 1 x 1. The results using the final 1 x 1 window yielded the best estimation of a camera's intrinsic parameters in that they resulted in the smallest average pixel error reported by Matlab.

There was a set of intrinsic calibration parameters determined for each camera for each day camera calibration images were collected, resulting in 15 different intrinsic parameter sets for each camera. A summary of the intrinsic parameter determinations for each camera is presented as Table 2, Table 3, and Table 4. It was anticipated that the results for a given camera across the 15 calibration images sets would have appeared very similar. This was not the case with some of the parameters (particularly skew and distortion). However, based on the discussions by Bouguet [87] that skew and several of the distortion coefficients can be assumed to be zero, the lower order terms were unimportant, comparatively speaking. Indeed, tests reported subsequently showed that the variations in these lower order coefficients did not lead to significant differences in the end results.

Once the intrinsic parameters were determined, the extrinsic parameters spatially relating one camera to the other were determined using additional Matlab routines. With the images marked as described above, the same images could be used for determining these extrinsic parameters.

The resulting extrinsic parameters were expressed as a pair of rotation vectors and a pair of translation vectors for the camera pair. As an example, for the left and right camera pair, execution of the Matlab code resulted in one rotation vector and translation vector representing a transformation from the left camera coordinate system to the right camera coordinate system. A second rotation vector and translation vector representing the transformation from the right camera coordinate system to the left camera coordinate system was also computed, though this is just the inverse of the first one.

Table 2 Intrinsic calibration results (left camera).

	Left Camera Calibration Intrinsic Parameters											
	Focal Length		Center		Skew	Distortion					Pixel Error	
	fcX	fcY	ccX	ccY	alpha_c	kc1	kc2	kc3	kc4	kc5	X	Y
8/30/05	1648.352	1646.565	753.139	565.137	1.35E-03	-0.351	2.986	5.35E-03	1.41E-02	-14.186	0.226	0.191
8/31/05	1642.568	1644.522	757.239	578.548	1.40E-03	-0.244	0.384	7.93E-03	1.65E-02	3.814	0.199	0.194
9/1/05	1616.909	1618.049	788.864	585.720	9.24E-04	-0.257	1.296	8.52E-03	2.02E-02	-5.641	0.206	0.199
9/2/05	1619.913	1619.072	801.027	591.881	1.58E-03	-0.262	1.788	9.49E-03	1.94E-02	-9.128	0.216	0.191
9/3/05	1638.220	1635.103	777.750	592.792	4.60E-03	-0.337	2.557	6.63E-03	1.53E-02	-10.573	0.202	0.189
9/7/05	1618.852	1621.604	794.108	590.629	6.41E-04	-0.294	2.134	9.33E-03	2.22E-02	-11.580	0.200	0.182
9/8/05	1617.688	1616.632	812.501	587.546	1.38E-03	-0.284	2.093	8.53E-03	2.21E-02	-11.642	0.224	0.207
9/9/05	1623.089	1624.405	806.108	568.913	-6.85E-04	-0.244	0.487	7.95E-03	2.51E-02	1.404	0.211	0.184
9/10/05	1625.934	1625.943	791.238	586.762	1.31E-03	-0.259	1.211	8.18E-03	2.06E-02	-5.188	0.215	0.197
9/11/05	1618.769	1618.738	797.938	583.409	1.20E-03	-0.250	0.974	8.40E-03	2.12E-02	-2.972	0.222	0.172
9/12/05	1617.057	1619.726	798.957	574.794	2.51E-04	-0.272	1.582	7.35E-03	2.32E-02	-7.180	0.211	0.179
9/13/05	1625.195	1631.415	787.850	576.973	-1.93E-03	-0.255	0.993	9.56E-03	2.46E-02	-4.231	0.214	0.192
9/14/05	1627.136	1624.574	807.351	575.738	-2.36E-04	-0.192	-0.439	9.08E-03	2.29E-02	5.164	0.219	0.182
9/15/05	1627.892	1628.871	807.330	575.472	2.57E-04	-0.333	2.862	7.56E-03	2.45E-02	-15.678	0.218	0.192
9/28/05	1633.264	1627.461	806.987	576.789	1.71E-03	-0.245	0.990	7.52E-03	2.02E-02	-3.790	0.220	0.187
Mean	1626.723	1626.845	792.559	580.740	9.17E-04	-0.272	1.460	8.09E-03	2.08E-02	-6.094	0.214	0.189
Standard Deviation	9.836	9.230	17.786	8.400	1.42E-03	0.042	0.962	1.13E-03	3.35E-03	6.264	8.54E-03	8.59E-03
Variance	96.750	85.201	316.333	70.552	2.03E-06	1.76E-03	0.925	1.28E-06	1.12E-05	39.238	7.29E-05	7.38E-05

Table 3 Intrinsic calibration results (right camera).

Right Camera Calibration Intrinsic Parameters												
	Focal Length		Center		Skew	Distortion					Pixel Error	
	fcX	fcY	ccX	ccY	alpha_c	kc1	kc2	kc3	kc4	kc5	X	Y
8/30/05	1643.986	1647.581	631.018	528.222	-2.21E-03	-0.213	0.967	-2.25E-03	-2.50E-03	-4.063	0.260	0.259
8/31/05	1626.902	1630.218	632.338	549.619	-7.80E-04	-0.180	0.642	1.17E-03	-1.65E-03	-3.634	0.253	0.265
9/1/05	1604.466	1607.791	607.094	552.903	-1.17E-03	-0.186	0.206	2.23E-03	-5.22E-03	-1.039	0.237	0.256
9/2/05	1612.863	1615.877	624.941	565.744	-6.19E-05	-0.158	-0.365	4.27E-03	-4.79E-03	2.682	0.284	0.259
9/3/05	1627.844	1630.525	636.136	552.640	-6.64E-04	-0.194	0.420	1.18E-03	-2.75E-03	-1.630	0.255	0.242
9/7/05	1613.115	1616.295	626.863	564.469	1.45E-04	-0.164	-0.509	4.77E-03	-5.36E-03	4.001	0.246	0.242
9/8/05	1612.545	1615.549	631.677	564.160	-3.46E-04	-0.201	-0.018	3.18E-03	-4.70E-03	2.028	0.257	0.256
9/9/05	1614.118	1616.689	630.603	559.991	-2.76E-04	-0.155	-0.798	4.05E-03	-4.18E-03	5.360	0.255	0.249
9/10/05	1615.760	1620.313	627.678	565.212	-3.72E-04	-0.163	-0.456	3.38E-03	-5.67E-03	3.399	0.240	0.219
9/11/05	1613.897	1615.976	617.764	558.710	-7.86E-05	-0.162	-0.600	3.98E-03	-5.10E-03	5.019	0.253	0.253
9/12/05	1616.632	1618.428	616.271	552.703	-5.21E-04	-0.192	0.010	3.14E-03	-4.88E-03	1.223	0.246	0.261
9/13/05	1612.554	1615.890	622.249	550.764	-2.07E-05	-0.137	-0.913	3.73E-03	-6.12E-03	5.364	0.257	0.247
9/14/05	1615.325	1620.385	632.667	565.918	6.20E-04	-0.164	-0.539	4.90E-03	-6.72E-03	3.435	0.256	0.237
9/15/05	1620.345	1623.327	637.318	566.164	-4.73E-04	-0.217	0.227	3.41E-03	-4.94E-03	0.400	0.260	0.257
9/28/05	1613.866	1617.656	637.522	567.361	-1.69E-04	-0.170	-0.313	3.57E-03	-5.06E-03	2.513	0.220	0.235
Mean	1617.615	1620.833	627.476	557.639	-4.26E-04	-0.177	-0.136	2.98E-03	-4.64E-03	1.671	0.252	0.249
Standard Deviation	9.299	9.367	8.589	10.363	6.48E-04	0.023	0.545	1.82E-03	1.37E-03	3.081	1.39E-02	1.23E-02
Variance	86.470	87.744	73.771	107.382	4.20E-07	5.26E-04	0.297	3.32E-06	1.88E-06	9.495	1.95E-04	1.50E-04

Table 4 Intrinsic calibration results (middle camera).

Middle Camera Calibration Intrinsic Parameters												
	Focal Length		Center		Skew	Distortion					Pixel Error	
	fcX	fcY	ccX	ccY	alpha_c	kc1	kc2	kc3	kc4	kc5	X	Y
8/30/05	1657.668	1657.885	659.702	534.451	8.21E-04	-0.131	-0.668	-1.35E-03	4.72E-03	5.027	0.250	0.263
8/31/05	1655.523	1655.884	668.911	548.786	6.36E-04	-0.152	-0.489	1.46E-03	4.70E-03	4.599	0.276	0.279
9/1/05	1634.785	1636.298	672.548	554.471	8.95E-04	-0.246	0.825	1.54E-03	6.40E-03	-1.303	0.275	0.257
9/2/05	1636.580	1638.060	677.107	561.365	1.10E-03	-0.247	0.912	2.35E-03	7.19E-03	-1.768	0.266	0.269
9/3/05	1651.430	1652.129	670.227	556.450	1.00E-03	-0.185	0.146	1.27E-03	5.67E-03	1.043	0.244	0.252
9/7/05	1638.812	1640.134	680.621	561.848	9.04E-04	-0.267	1.090	2.81E-03	8.34E-03	-1.893	0.276	0.258
9/8/05	1645.019	1646.340	684.223	558.622	1.20E-03	-0.281	1.413	2.25E-03	9.01E-03	-3.961	0.254	0.255
9/9/05	1644.039	1646.201	683.352	559.865	9.32E-04	-0.286	1.380	2.65E-03	9.70E-03	-3.412	0.265	0.264
9/10/05	1646.355	1646.634	673.810	558.147	1.33E-03	-0.272	1.297	2.68E-03	7.21E-03	-3.221	0.247	0.259
9/11/05	1642.816	1644.689	676.230	559.112	7.78E-04	-0.276	1.393	2.69E-03	7.09E-03	-4.492	0.279	0.267
9/12/05	1642.337	1644.535	675.672	555.249	6.02E-04	-0.270	1.307	1.76E-03	8.21E-03	-4.233	0.283	0.276
9/13/05	1644.612	1647.883	668.320	551.061	-8.85E-05	-0.269	1.037	2.42E-03	7.63E-03	-1.715	0.280	0.258
9/14/05	1641.002	1641.407	681.428	557.993	4.75E-04	-0.265	1.128	3.07E-03	7.51E-03	-2.908	0.282	0.261
9/15/05	1650.450	1651.222	684.751	552.204	1.10E-03	-0.231	0.328	4.99E-04	9.69E-03	1.498	0.272	0.259
9/28/05	1642.561	1642.658	686.225	552.187	7.37E-04	-0.266	1.205	1.53E-03	7.07E-03	-3.351	0.256	0.245
Mean	1644.933	1646.131	676.208	554.787	8.28E-04	-0.243	0.820	1.84E-03	7.34E-03	-1.339	0.267	0.262
Standard Deviation	6.505	6.184	7.461	6.834	3.44E-04	0.048	0.679	1.13E-03	1.55E-03	3.044	1.35E-02	8.73E-03
Variance	42.317	38.240	55.662	46.704	1.18E-07	2.32E-03	0.461	1.28E-06	2.39E-06	9.263	1.83E-04	7.62E-05

Table 5, Table 6, and Table 7 provide the resulting vectors for the relationship between the left/right, left/middle, and right/middle camera pairs.

Table 5 Extrinsic calibration results (left/right cameras).

Left/Right Extrinsic Parameters						
	Rotation Vector			Translation Vector		
	omX	omY	omZ	TX	TY	TZ
8/30/05	-0.0097	-0.6912	0.1860	338.0795	43.0844	130.1759
8/31/05	-0.0043	-0.6916	0.1807	339.0703	42.6411	124.4348
9/1/05	-0.0080	-0.6636	0.1770	337.0821	41.2095	120.7500
9/2/05	-0.0051	-0.6660	0.1742	334.6101	40.6066	128.6037
9/3/05	-0.0125	-0.6836	0.1759	336.2281	41.7201	131.2395
9/7/05	-0.0039	-0.6706	0.1745	334.5186	40.4190	127.5987
9/8/05	-0.0029	-0.6632	0.1753	333.0856	40.2127	130.5159
9/9/05	0.0058	-0.6652	0.1797	335.7693	40.6569	127.3157
9/10/05	-0.0010	-0.6712	0.1746	335.0533	39.9577	127.6551
9/11/05	-0.0020	-0.6629	0.1770	335.4929	41.0001	127.1256
9/12/05	-0.0009	-0.6603	0.1797	335.8162	41.1299	126.6055
9/13/05	-0.0022	-0.6685	0.1804	336.8774	41.9148	121.8422
9/14/05	0.0064	-0.6645	0.1774	334.7653	40.4813	129.5166
9/15/05	0.0057	-0.6660	0.1771	334.1111	39.7866	129.9557
9/28/05	0.0059	-0.6695	0.1759	335.7718	39.9404	129.5878
Mean	-0.0019	-0.6705	0.1777	335.7554	40.9841	127.5282
Standard Deviation	0.0058	0.0101	0.0032	1.5550	0.9867	3.0924
Variance	3.412E-05	1.014E-04	9.963E-06	2.4180	0.9736	9.5628

Table 6 Extrinsic calibration results (left/middle cameras).

Left/Middle Extrinsic Parameters						
	Rotation Vector			Translation Vector		
	omX	omY	omZ	TX	TY	TZ
8/30/05	0.0335	-0.3486	0.0270	177.3764	7.7852	-109.0949
8/31/05	0.0352	-0.3515	0.0244	178.3913	9.1298	-108.3191
9/1/05	0.0342	-0.3386	0.0233	177.0245	8.9764	-103.8198
9/2/05	0.0350	-0.3341	0.0224	177.4040	9.5574	-102.6574
9/3/05	0.0316	-0.3411	0.0209	177.8755	9.0379	-103.4697
9/7/05	0.0373	-0.3409	0.0222	177.4623	9.5921	-102.7869
9/8/05	0.0373	-0.3314	0.0230	176.8918	8.8545	-97.5782
9/9/05	0.0494	-0.3348	0.0245	178.0810	9.0838	-102.4115
9/10/05	0.0378	-0.3356	0.0224	176.9527	9.3357	-103.4492
9/11/05	0.0402	-0.3341	0.0225	177.0799	9.0536	-100.5106
9/12/05	0.0427	-0.3335	0.0241	177.3405	8.5795	-101.2444
9/13/05	0.0400	-0.3354	0.0243	177.6840	9.0108	-106.8422
9/14/05	0.0448	-0.3307	0.0234	177.9535	9.4060	-104.4814
9/15/05	0.0408	-0.3316	0.0245	177.6073	8.2229	-101.0740
9/28/05	0.0408	-0.3338	0.0229	178.5992	8.4762	-103.9913
Mean	0.0387	-0.3370	0.0234	177.5816	8.9401	-103.4487
Standard Deviation	0.0047	0.0061	0.0014	0.5179	0.4976	2.9803
Variance	2.196E-05	3.769E-05	2.017E-06	0.2683	0.2476	8.8820

Table 7 Extrinsic calibration results (right/middle cameras).

Right/Middle Extrinsic Parameters						
	Rotation Vector			Translation Vector		
	omX	omY	omZ	TX	TY	TZ
8/30/05	0.0633	0.3472	-0.1421	-187.5829	19.5871	-116.8910
8/31/05	0.0599	0.3448	-0.1395	-185.7416	19.8640	-111.3751
9/1/05	0.0618	0.3311	-0.1381	-184.7381	20.3048	-109.2262
9/2/05	0.0590	0.3380	-0.1365	-184.6267	20.9822	-113.7981
9/3/05	0.0639	0.3467	-0.1381	-185.6149	20.6412	-113.1175
9/7/05	0.0606	0.3362	-0.1363	-184.2604	21.1348	-113.7358
9/8/05	0.0588	0.3369	-0.1362	-183.5624	20.5855	-111.1813
9/9/05	0.0623	0.3371	-0.1365	-184.3112	20.7826	-112.2284
9/10/05	0.0578	0.3412	-0.1362	-184.7285	21.1993	-112.2647
9/11/05	0.0614	0.3343	-0.1375	-184.5826	20.6622	-111.1439
9/12/05	0.0627	0.3329	-0.1382	-184.6405	20.4796	-111.8344
9/13/05	0.0613	0.3403	-0.1397	-184.8735	20.0798	-110.2509
9/14/05	0.0570	0.3402	-0.1372	-184.3165	21.1203	-115.5440
9/15/05	0.0537	0.3398	-0.1369	-183.5593	20.3161	-112.6506
9/28/05	0.0537	0.3413	-0.1370	-184.1713	20.5675	-114.2087
Mean	0.0598	0.3392	-0.1377	-184.7540	20.5538	-112.6300
Standard Deviation	0.0032	0.0047	0.0016	0.9860	0.4698	2.0004
Variance	1.007E-05	2.221E-05	2.694E-06	0.9722	0.2208	4.0015

All subsequent efforts to provide 3D locations using triangulation rely on the availability of these extrinsic camera parameters. Unlike the differences perceived in the intrinsic parameters between each of the daily calibration sets, the camera/camera extrinsic parameters appeared fairly consistent from experiment day to experiment day. However, in an attempt to negate the impact of location differences that did occur, the averaging technique to be discussed in Subsection 7.2 was employed.

6.2 Camera/Monitor Calibration Processing

In addition to requiring relationships between the cameras, spatial relationships between the camera and the monitor were also required. As discussed in Subsection 5.3, determining the spatial relationship between the cameras and the monitor relied on developing a 'bridge' relationship between the cameras and the camera/monitor calibration rig (see Fig. 17 and Fig. 18). Toward this end, the images of the camera/monitor calibration rig (see Fig. 17 and Fig. 18) collected by the three cameras during calibration were processed by manually locating the pixels in the images of the colored areas on the strings (see Fig. 18). From these, a coordinate system for the rig was constructed. In addition to the two colored areas being identified along each string, another point representing the origin of the rig was identified. The distance from the rig origin to the monitor origin, a translation vector, was manually determined after the construction and mounting of the rig using a string, ruler, and micrometer. Three attachment/ measurement trials were attempted, and the average distances were found to be 6.03 mm in the *X* direction, -139.76 mm in the *Y* direction, and -539.38 mm in the *Z* direction (see Subsection 5.3).

Unfortunately, establishing the orientation of the rig coordinate system was not quite as simple as directly treating the vectors determined from the string points shown in Fig. 18 (obtained from image processing) as coordinate axes. Due to the construction of the rig, it was known that the rig strings were not perfectly parallel to the monitor coordinate system axes. Even if the rig had been constructed perfectly and the rig strings were indeed parallel to the monitor coordinate axes, the likelihood that string vectors derived from image processing and triangulation would be exactly the needed vectors was infinitesimal. Therefore, a method to estimate a rig coordinate system having the needed relation to the monitor coordinate system was created.

It was estimated from the construction methods and visual examination that the *Z* axis of the rig was the axis that most accurately approximated the corresponding monitor axis when the rig was attached to the monitor. Therefore, the 3D *Z* line defined by the *Z* string point locations determined using triangulation were taken to be the *Z* axis of the rig coordinate system. One of the points used was taken to be the origin (see 'Origin' in Fig. 18). Because it was estimated that the *Y* string was closer to being orthogonal to the *Z* axis than the *X* string, the *Y* string was used to define a *Y* axis for the rig coordinate system. In particular, the *Z* component of the upper most string location on the rig's *Y* axis string was set equal to the *Z* component of the rig's origin, thus defining a *Y* axis orthogonal to the *Z* axis. Finally, the *X* axis of the rig coordinate system was then established by taking the cross product of the *Z* axis and *Y* axis vectors. The use of *Z* cross *Y* rather than *Y* cross *Z* resulted from the coordinate system commonly used for images has the positive *Y* axis down rather than up.

As mentioned previously, the rig coordinate system provides a 'bridge' between the camera and monitor coordinate systems. Because of the construction of the rig and the rig coordinate system, the only difference between a camera to rig transformation and camera to monitor transformation is the difference in origins, which is given by the translation vector stated previously. Therefore, the transformation of a location *P* from camera (*Cam*) coordinates to monitor (*Mon*) can be expressed by the following:

$$^{Mon}P = {}^{Cam}P * {}^{Rig}_{Cam}R + {}^{Rig}_{Cam}TV + {}^{Mon}_{Rig}TV \tag{54}$$

where ${}^{Rig}_{Cam}R$ is the rotation matrix between the camera and the rig, ${}^{Rig}_{Cam}TV$ is the translation vector between the camera and the rig, and ${}^{Mon}_{Rig}TV$ is the translation vector between the rig and the monitor. Of course, the construction of the rig coordinate system introduces both rotational and translational errors in the transformation. These errors were discussed in Subsection 5.3, and the end-to-end error measurements reports indicate that the use of the rig for determining the transformation was acceptable.

Table 8,Table 9, and Table 10 present the pixel values used to derive the rig axes in camera coordinates for each of the 15 calibration image sets collected. Based on the assumption that the rig would be attached at the same location relative to the monitor each time and that the cameras were fixed with respect to the monitor, it was anticipated that the pixel values would be almost identical for corresponding points from the same camera perspective across all the calibration sets. While the corresponding pixel values were reasonably close, they were not as close as had been hoped (all within one pixel of each other). It was assumed that a majority of the discrepancies were due to errors attaching the rig to the monitor and errors in locating the string markers. However, the possibility that slight camera movements caused the location errors could not be ruled out. In an attempt to negate the impact of these location differences, the averaging technique discussed in Subsection 7.2 was employed to find the average relationship between the camera and the rig.

Table 8 Camera/monitor rig pixel points (left camera).

Left Camera Monitor/Camera Calibration Rig Points														
	X Axis				Y Axis				Z Axis				Origin	
	Point 1		Point 2		Point 3		Point 4		Point 5		Point 6		Point 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
8/30/05	915	291	683	318	813	638	818	428	688	368	780	485	816	530
8/31/05	919	293	687	321	817	641	822	431	692	371	784	488	821	533
9/1/05	918	293	685	320	815	640	820	430	690	370	783	487	819	532
9/2/05	920	294	687	321	817	641	822	432	692	372	785	489	821	533
9/3/05	918	293	686	319	815	641	821	429	690	370	783	487	819	532
9/7/05	915	292	683	319	814	640	819	429	688	369	781	487	817	531
9/8/05	915	292	683	319	812	640	817	429	687	369	779	486	816	532
9/9/05	915	289	683	316	813	637	819	427	688	366	781	484	817	529
9/10/05	918	291	686	319	816	640	821	429	691	369	783	486	819	531
9/11/05	916	292	684	319	813	641	819	429	688	369	781	487	816	532
9/12/05	914	290	681	317	811	638	816	427	686	368	778	484	815	529
9/13/05	917	287	684	314	814	635	819	424	689	364	781	481	817	526
9/14/05	916	288	684	315	814	637	819	425	689	366	781	483	817	528
9/15/05	915	291	682	318	812	639	817	428	687	368	779	485	815	530
9/28/05	918	291	686	318	816	639	821	428	691	368	783	485	819	531
Mean	916.60	291.13	684.27	318.20	814.13	639.13	819.33	428.33	689.07	368.47	781.47	485.60	817.60	530.60
Standard Deviation	1.80	1.96	1.83	2.01	1.85	1.81	1.84	2.06	1.87	2.03	2.00	2.06	1.96	1.96
Variance	3.26	3.84	3.35	4.03	3.41	3.27	3.38	4.24	3.50	4.12	3.98	4.26	3.83	3.83

Table 9 Camera/monitor rig pixel points (right camera).

Right Camera Monitor/Camera Calibration Rig Points														
	X Axis				Y Axis				Z Axis				Origin	
	Point 1		Point 2		Point 3		Point 4		Point 5		Point 6		Point 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
8/30/05	767	340	548	301	534	639	532	425	693	377	577	487	533	528
8/31/05	763	340	546	301	532	639	530	425	691	378	574	488	531	530
9/1/05	767	341	548	302	534	639	532	427	693	378	576	488	533	530
9/2/05	767	343	548	303	534	640	533	428	693	380	577	489	533	531
9/3/05	767	343	548	303	533	641	532	429	692	380	577	489	533	532
9/7/05	767	341	547	302	533	640	532	427	692	379	576	489	533	530
9/8/05	768	342	549	303	534	640	533	428	694	379	577	489	534	531
9/9/05	770	342	551	303	537	640	535	427	696	379	580	489	536	531
9/10/05	768	341	549	303	535	640	533	428	694	379	578	489	535	530
9/11/05	768	343	549	304	535	641	533	429	694	380	577	490	534	532
9/12/05	769	343	550	305	535	642	534	429	695	380	579	490	534	532
9/13/05	769	343	550	303	535	640	535	428	695	379	578	490	535	531
9/14/05	769	343	549	304	535	641	533	429	694	380	578	490	535	532
9/15/05	768	345	549	306	535	643	534	431	694	381	578	491	534	534
9/28/05	770	344	551	305	537	642	536	431	696	381	580	491	536	533
Mean	767.80	342.27	548.80	303.20	534.53	640.47	533.13	428.07	693.73	379.33	577.47	489.27	533.93	531.13
Standard Deviation	1.70	1.44	1.37	1.42	1.36	1.19	1.51	1.75	1.44	1.11	1.55	1.10	1.33	1.46
Variance	2.89	2.07	1.89	2.03	1.84	1.41	2.27	3.07	2.07	1.24	2.41	1.21	1.78	2.12

Table 10 Camera/monitor rig pixel points (middle camera).

Middle Camera Monitor/Camera Calibration Rig Points														
	X Axis				Y Axis				Z Axis				Origin	
	Point 1		Point 2		Point 3		Point 4		Point 5		Point 6		Point 7	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
8/30/05	837	118	513	127	625	569	612	296	620	188	618	367	619	426
8/31/05	838	124	512	133	626	575	613	301	621	193	619	372	619	433
9/1/05	837	123	512	131	625	573	613	298	621	191	619	370	619	431
9/2/05	839	123	511	132	625	574	612	299	621	191	618	371	619	431
9/3/05	838	123	512	132	625	574	613	299	621	192	619	371	619	431
9/7/05	838	121	512	131	625	572	612	298	621	189	619	369	619	430
9/8/05	839	122	512	130	625	571	612	298	621	190	619	369	619	430
9/9/05	837	121	511	129	625	572	612	297	621	189	619	368	619	429
9/10/05	837	121	511	130	625	573	612	298	621	190	618	369	619	429
9/11/05	837	124	511	133	624	574	611	300	621	191	618	372	618	431
9/12/05	837	123	512	132	625	574	613	299	621	191	619	371	619	431
9/13/05	837	120	511	129	624	571	612	296	620	188	618	367	618	428
9/14/05	837	123	512	131	625	573	613	298	620	190	619	370	619	431
9/15/05	837	128	511	136	624	578	612	304	621	195	617	374	618	436
9/28/05	836	129	511	137	624	578	612	304	621	196	618	375	618	436
Mean	837.40	122.87	511.60	131.53	624.80	573.40	612.27	299.00	620.80	190.93	618.47	370.33	618.73	430.87
Standard Deviation	0.83	2.80	0.63	2.59	0.56	2.41	0.59	2.42	0.41	2.31	0.64	2.32	0.46	2.64
Variance	0.69	7.84	0.40	6.70	0.31	5.83	0.35	5.86	0.17	5.35	0.41	5.38	0.21	6.98

6.3 Hardware Calibration Averaging

Because some of the results of the session (daily) hardware (camera and camera/monitor) calibrations appeared slightly inconsistent and the impact of these inconsistencies was unknown, it was decided to also calculate an average set of calibration parameters to use for further processing and determine the effect of using them. A comparison of results from processing using daily calibration values and processing using average calibration values could then be made to gain insight into the effect (importance) of the calibration inconsistencies.

The average intrinsic parameters were determined by simply averaging the intrinsic results for each individual camera. Instead of then using the average intrinsic parameters for each camera to re-compute the extrinsic parameters (the average intrinsic parameters would only be used during the determination of individual 'undistorted' image pixels as described in Subsection 4.2), it was decided to simply average the extrinsic results that were determined using the session intrinsic parameters. Because of the interrelationship between the camera calibration and the camera/monitor calibration, a method to decouple and isolate the error sources was elusive. Therefore, it was decided to use the individual daily camera calibration results to develop a set of camera/monitor calibration results, and then average these individual results to create a set of average camera/monitor calibration parameters. This resulted in a potential for four different calibration parameter combinations:

a. intrinsic parameters calculated based on the given day’s camera calibration images and the extrinsic parameters based on the given day’s rig/monitor calibration images (Day:Day)
b. intrinsic parameters calculated based on the given day’s camera calibration images and the average of the extrinsic parameters across all experiment days (Day:Avg)
c. the average of the intrinsic parameters calculated across all experiment days and the extrinsic parameters based on the given day’s rig/monitor calibration images (Avg:Day)

d. the average of the intrinsic parameters calculated across all experiment days and the average of the extrinsic parameters across all experiment days. (Avg:Avg)

All subsequent processing of subject images was then accomplished using each of the four hardware calibration combinations.

6.4 Subject Calibration Processing

Similar to the camera and camera/monitor calibrations that were performed using combinations of images from two cameras, subject calibration required two cameras and triangulation to locate the image features in 3D. However, for subject calibrations, only the left and right camera images were used. And unlike with the hardware calibrations, the fact that the stereo images were pseudo-stereo images became an issue. The term 'pseudo' is used because the images were actually collected approximately one second apart as there was no image multiplexing hardware. This means the spatial consistency between the images (a requirement for proper triangulation) was dependant on whether the subject moved during the capturing of each image. Unfortunately, there is no way to definitively identify the occurrence of subject movement, or to evaluate its magnitude or effect. However, in an attempt to estimate the likelihood of movement, the subject calibrations were carried out in two steps: a preliminary and then a final calibration.

The objective of the preliminary calibration was to identify images that should not be used for calibration either because of facial/head movement or because the subject was not gazing at the appropriate location. The objective of the final calibration step for each subject was to use the methods described in Section 3 to determine the average visual axes center in relation to the head, the average distance between the pupil and the visual axes center (similar to an eye radius), the average distances between each of the five non-pupillary facial features (resulting in 10 averages), and the average distance between the pupils.

In order facilitate that any discrepant or suspect images might be excluded from the subject calibration, the preliminary calibration was designed to identify images where either facial movement between the pseudo-stereo images for each gaze position occurred, head or eye movement between the acquisitions of each of the pseudo-stereo images occurred, or the subject's gaze point was different from that reported. While the likelihood of facial feature distortion was believed to have been minimized by the general placement of the green features dots, the variability of where subjects actually placed the markers introduced the possibility that feature distortion might be observed for some subjects. Head or eye movement during the collection of pseudo-stereo images would virtually ensure that the actual 3D locations determined using triangulation would be in error. Finally, if subjects reported erroneous gaze points, the calibration results would be in error due to the employed calibration methodology relying on the accuracy of the gaze pointed reported by the subject.

For detecting facial feature movement, the assumption was that significant differences in facial feature distances between images would indicate the likelihood that feature distortion had occurred. In addition, distance discrepancies might also indicate feature point 3D location errors that resulted from head or eye movement during pseudo-stereo image capture. For detecting an incorrect gaze location, it was assumed that the incorrect location would be one of the other monitor target locations, and therefore, there would be a smaller gaze direction angular error calculated using one of the other monitor locations as opposed to using the gaze location reported by the subject.

In an attempt to assess the movement potential, the mean of the individual feature pair distances across all nine of the subject calibration images for each subject for each camera was determined, $\overline{dFF_{p,s}}$ where p replaces the previous designation i, j for a particular facial feature pair. As the entire set of the following computations was repeated for each camera and none of the computations involved multiple camera values, no subscript for a particular camera is used. A normalized feature pair distance ($ndFF_{p,i,s}$) was determined for each feature pair (p) in each image (i: from 1 to n) for each subject (s) by taking the feature pair distance $dFF_{p,i,s}$ and dividing it by $\overline{dFF_{p,s}}$:

$$ndFF_{p,i,s} = \frac{dFF_{p,i,s}}{\left(\sum_{i=1}^{n} dFF_{p,i,s} / n\right)} = \frac{dFF_{p,i,s}}{\overline{dFF_{p,s}}} \tag{55}$$

Then, the differences between $ndFF_{p,i,s}$ and $\overline{ndFF_{p,s}}$ ($\overline{ndFF_{p,s}} = 1$) represented by $\Delta ndFF_{p,i,s}$ were determined for every feature pair for every image of every subject:

$$\Delta ndFF_{p,i,s} = \overline{ndFF_{p,s}} - ndFF_{p,i,s} = 1 - ndFF_{p,i,s} \tag{56}$$

The average difference, then, has to be zero (from the definitions). Computing this average verified that and was a useful cross check.

$$\overline{\Delta ndFF_{p}} = \frac{\sum_{s=1}^{k} \sum_{i=1}^{n} \Delta ndFF_{p,i,s}}{n * s} = 0 \tag{57}$$

Then, if $\Delta ndFF_{p,i,s}$ was more than 1.25[1] standard deviations away from the total mean ($\overline{\Delta ndFF_{p}} = 0$), that feature pair p in image i for subject s was deemed suspect. If a majority of the feature pairs in a given image (six or more out of the 10) were suspect, movement was assumed to have occurred, and that image was rejected. If less than six feature pairs were suspect, the image was deemed acceptable. Because both the left and right images are needed for triangulation, rejecting either resulted in both images being rejected.

To determine the existence of an erroneously reported gaze location, the preliminary calibration image results were used to determine a gaze direction angle error based on each of the nine possible monitor gaze target points being the actual gaze target point viewed by the subject. Each of these nine gaze angle errors was then compared to the gaze angle error determined using the monitor gaze target point reported by the subject. If the minimum gaze direction angle error occurred for a gaze location other than that

[1] The choice of 1.25 was made empirically by studying the characteristics of images rejected for several different thresholds.

reported, it was assumed that the subject had been looking at a location different from the one reported. For these cases, the pseudo-stereo image pair for that reported location was rejected.

It should be noted that the efforts to identify suspect images were conducted for all camera and camera/monitor calibration result categories (Avg:Avg, Day:Day, etc.) The rejected images represented a total compilation of images that were rejected over all calibration categories for a particular subject. No effort was made to determine for which calibration category a particular image was rejected. Therefore, there is some likelihood that a particular subject calibration image was rejected from subsequent usage even though the reason for rejection may not have occurred for all calibration result categories.

Upon completion of the preliminary subject calibration, a final calibration was performed. In addition to the images rejected during the preliminary calibration, two other groups of images were excluded from usage during the final calibration. The first group involved those images collected when the subject had expressed a concern about their adherence to the 'no blinking, no movement' requirements. Despite the fact that these images had been reviewed and no issues were observed, it was decided to remove these images from the final subject calibration to be safe. The second grouping involved all of those subject calibration images that, during a manual re-review, had appeared blurry, had one or more facial features and/or pupils that were not clearly visible, or had significant facial deformation due to smiling, laughing, etc.

Table 11 provides a summary of the total number of subject calibration images that were rejected for various reasons and not included as part of the final calibration. Because Subject 17 had fewer than four acceptable calibration images, they were excluded from further consideration. In an actual application, calibration results would be checked before continuing, and attempts would have been made to re-calibrate this subject.

Once the appropriate images for exclusion were identified, the final calibration was performed. Resulting from the final calibration for each subject was the average visual axes center expressed in a head coordinate system defined by using the method described in Subsection 3.3. The face plane required by the method was defined using the locations of the hairline marker and the left and right 'under eye' markers (see Fig. 20). In addition to the visual axes center, the mean, standard deviation, and variance of the distances between the facial feature points (green marker points) and the distance between the pupils were determined. Finally, the average visual axes center to pupil center distance was also recorded for each eye for each subject.

Table 11 Subject calibration rejected images.

	Rejection Cause			
	Problem Images	**Normalized Distance**	**Min Angle**	**Lack of Stability**
Total Images Identified	39	31	19	32
# of Different Subjects	23	23	15	22
Max for Any Subject	6	3	3	3
Subject 17	6	1	1	0
Total Image Pool = 684				

6.5 Experiment Image Processing

The images collected of the subjects looking at the monitor gaze target points in any order of the subjects' choosing (see Subsection 5.4) were the actual experiment images. To simulate the usage of a single camera, only the images collected from the middle camera were used. In order to avoid biasing subject behavior by the calibration processing, it was necessary to obtain the experiment data first. For an actual application, the calibration would be performed first. However, because the analyses were all performed after the fact, the order did not impact the analysis of the results.

The actual processing of the images involved manually determining the pixel locations of the pupil centers and the facial marker centers in the images collected using the middle camera. Then, using the intrinsic camera calibration parameters of the middle camera, each pupil center/feature pixel location was converted into an undistorted pixel location. The undistorted locations were then used along with the single camera 3D optimization methodology described in Section 4 to determine 3D locations in the middle camera coordinate system for each of the facial features (excluding the pupils). It should be noted that all locations and distances for the five facial markers were used for the optimization.

Once the 3D locations of the facial markers were determined, a head coordinate system was derived in a manner identical to that used during the final calibration. Given that no significant facial distortion occurred between experiment and calibration, the calibration head coordinate system and the newly derived one should be identical. Thus, the location of the visual axes center should be the same as the one determined during subject calibration.

Given the visual axes center and the visual axes center to pupil center distance also determined during subject calibration, the 3D location of the pupil centers can be determined by solving the quadratic equation (see Eq. 44) developed using the methodology described in Section 4.

With the 3D locations of the visual axes center (*VAC*) and the pupil center (*PC*), a 3D gaze vector (*GV*) was calculated for each eye (*e*), for each image (*i*):

$$GV_{e,i} = PC_{e,i} - VAC_{e,i} \tag{58}$$

Using the camera to monitor transformations determined during hardware calibration, the gaze vector (${}^{Mon}GV_{e,i}$), and the known gaze target point (${}^{Mon}TP_i$), a gaze intercept point on the monitor for each eye/image (${}^{Mon}GIP_{e,i}$) was determined using the following:

$${}^{Mon}unit_{e,i} = \frac{{}^{Mon}GV_{e,i}}{\left\| {}^{Mon}GV_{e,i} \right\|} \tag{59}$$

$$c_{e,i} = \frac{-\,{}^{Mon}Z_{PC_{e,i}}}{{}^{Mon}Z_{unit_{e,i}}} \tag{60}$$

$${}^{Mon}GIP_{e,i} = < {}^{Mon}X_{PC_e} + \left({}^{Mon}X_{unit_{e,i}} * c_{e,i}\right), {}^{Mon}Y_{PC_e} + \left({}^{Mon}Y_{unit_{e,i}} * c_{e,i}\right), {}^{Mon}Z_{TP_i} > \tag{61}$$

In addition to the two gaze vectors ($GV_{left,i}$ and $GV_{right,i}$), a third gaze vector ($GV_{avg,i}$: see Fig. 22) was then derived by finding the midpoint between the visual axes centers for each eye and each image ($midVAC_i$) and projecting from this point to the midpoint between the gaze intercept points for the individual eyes ($midGIP_i$):

$$^{Mon}midGIP_i = \frac{^{Mon}GIP_{left,i} + ^{Mon}GIP_{right,i}}{2} \tag{62}$$

$$^{Mon}midVAC_i = \frac{^{Mon}VAC_{left,i} + ^{Mon}VAC_{right,i}}{2} \tag{63}$$

$$^{Mon}GV_{avg,i} = ^{Mon}midGIP_i - ^{Mon}midVAC_i \tag{64}$$

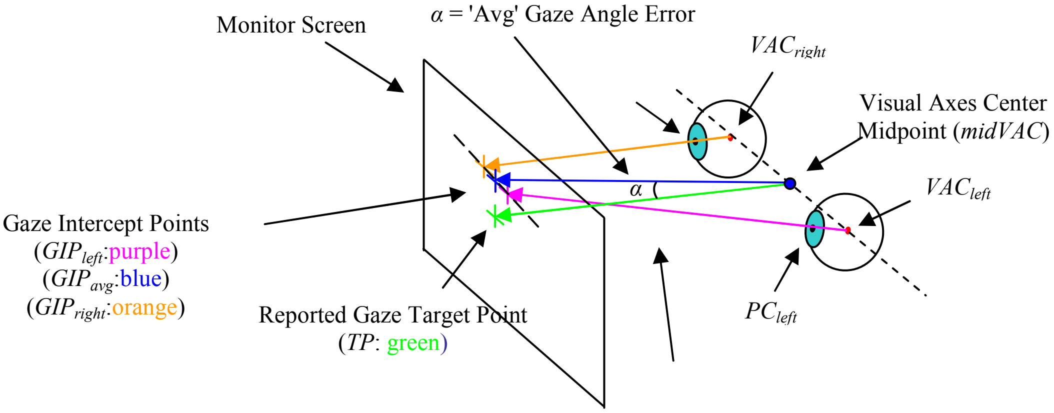


Fig. 22 Visual axes center midpoint.

The angular difference (gaze angle error) between either the left eye, right eye, or average gaze vectors and the vector from either the left eye, right eye, or midpoint to the reported gaze target point all provide an indication of the performance of the methods being examined and a metric by which to compare the performance. However, since it was not immediately obvious which gaze angle error to use in the analysis, gaze angle errors were calculated using all of the possible gaze vectors depicted in Fig. 22.

It had been anticipated that using the gaze vector for the dominant eye of the subject would result in the smallest angular errors. However, this was not the case: the average of the gaze angle errors determined using the dominant eye gaze vector for all subjects was 3.75 degrees, the average of the gaze angle errors determined using the non-dominant eye gaze vector for all subjects was 3.70 degrees, and the average of the gaze angle errors using the average gaze vector for all subjects was 3.15 degrees. Because of the smaller average of the gaze angle errors determined using the average gaze vector, the fact that incorporating eye dominance determination would significantly complicate the actual implementation of the visual axes center method, and the disagreement in the literature as to whether the notion of eye dominance even exists, it was decided to use the average (visual axes center midpoint to average gaze intercept point) gaze vector to determine the gaze angle errors throughout the remainder of this dissertation.

7. EXPERIMENT RESULTS/ANALYSES

The primary objective of the subject experiments was to provide the data necessary to evaluate the performance of the visual axes center and single camera 3D optimization methodologies in determining gaze. The primary measure of this performance is the magnitude of the gaze angle error: the angle between the subject-reported gaze vector and the one calculated using the developed methodologies. The remainder of this section details not only the results of the experiments in terms of the resulting gaze angle errors, but also discusses some of the experiment parameters, and how they affected the results.

7.1 Manual Feature Identification Analysis

For all of the subject image processing discussed in Section 6, the determination of the pixel locations of the various facial features and pupil centers was accomplished manually. For this dissertation, over 23,940 individual image feature pixel locations were determined and recorded. Manual identification and image location determination provided more accurate results over the use of the available automatic feature finding computer-based tools. Nevertheless, there were concerns about the consistency with which a human would find the pupil centers across the large number of images.

Humans are believed to interpret images in a fundamentally different way from computers, and their interpretations are not only more influenced by color, light, and shadow, but also by emotion: based on what the image represents and the context in which it is being interpreted. Because of the repetitive and tedious nature of manually locating all of the image features for this dissertation, there was a concern that only having a single individual (in this case, the author of this dissertation) locate all of the image features might somehow bias or skew the ultimate results.

In an attempt to determine if a bias was created by having a single individual manually locate all the features, two other individuals were recruited to re-determine the feature locations in a subset of the subject experiment images (experiment images for

Subjects 9 through 29). Only features from experiment images were re-determined. The experiment image locations of the corresponding features identified by the other two individuals were then compared to the pixel locations originally determined. In addition, the image pixel locations identified by the three individuals were then used to determine the gaze angle error for each image for each individual's pixel locations. The results of these efforts are summarized in Table 12.

Table 12 Multiple feature locator results.

		Original vs. #1	Original vs. #2	#1 vs. #2
Pixel Δ	Mean	0.417	0.460	0.416
	StDev	0.603	0.640	0.642
	Var	0.363	0.410	0.412
	Max	7[1]	6[2]	5[3]
[1]: Left Pupil X				
[2]: Right Pupil Y				
[3]: Left Pupil X, Right Pupil X, and Right Pupil Y				
		Original	#1	#2
Gaze Angle Error (degrees)	Mean	3.111	3.596	3.552
	StDev	2.589	4.460	4.097
	Var	6.701	19.893	16.783

Based on the average gaze angle errors, the original locations produced better results which might indicate that the original locator (the author) had higher motivation to be more careful and/or more consistent in locating the features.

7.2 Calibration Categories

The primary measure of gaze determination performance, the gaze angle error, was originally planned to only be determined for the actual experiment images. However, as already discussed, a gaze angle error was calculated during the preliminary calibration, and used to identify suspect images for removal from the final subject calibration process. Because of this, it was decided to re-calculate the gaze angle errors associated with the final calibration as well. In addition, because the subject calibration images were collected and processed as pseudo-stereo images and the actual experiment images were not, the subject calibration also provided an opportunity to compare the results of the single camera 3D optimization technique with those from the pseudo-stereo triangulation method for determining 3D locations. Therefore, the gaze angle errors for the final subject calibration images were determined using both triangulation and single camera 3D optimization. Finally, because of the occasional inconsistencies that were observed between the individual daily hardware calibration results, the gaze angle errors were calculated using each of the hardware calibration categories. This effort provided an opportunity to not only evaluate the differences in the overall categorizes, but also provided an opportunity for insight into the impact the lower level hardware calibration inconsistencies might be expected to have. The results of this subject calibration study are summarized in Table 13. The 'triangulation' results represent those found from pseudo-stereo processing. The 'optimization' results represent those found using images from a single camera. The 'perspective' category represents which single camera was used, or which camera was used as the primary coordinate system for transformations during pseudo-stereo processing.

From the results summarized in Table 13, it appeared that the calibration category made very little difference in the overall gaze angle error results, and that an average, or possibly a one-time calibration, could be satisfactorily used. Minimizing the frequency of the required hardware calibration would prove extremely beneficial in any real-world application. It also appeared that the camera used can have a noticeable impact on the overall results obtained. However, since only a single camera was actually used for the experiments, the importance of potential camera differences was not investigated. Fortunately, the left camera, which seemingly provided the better calibration results, was the camera used along with the middle camera to derive the transformation between the middle camera and the rig/monitor.

Table 13 Gaze angle error final subject calibration results.

	Left Camera Perspective											
	Day_Avg			Day_Day			Avg_Avg			Avg_Day		
	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'
Mean: Angle Error	2.850	2.998	2.557	2.849	2.998	2.559	2.849	2.997	2.556	2.849	2.997	2.557
StDev: Angle Error	2.198	2.574	2.102	2.195	2.573	2.100	2.197	2.574	2.103	2.194	2.574	2.101
Var: Angle Error	4.831	6.624	4.419	4.819	6.619	4.408	4.826	6.623	4.421	4.815	6.625	4.413
(triangulation)												
	Right Camera Perspective											
	Day_Avg			Day_Day			Avg_Avg			Avg_Day		
	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'
Mean: Angle Error	2.850	2.998	2.557	2.849	2.998	2.559	2.849	2.997	2.556	2.849	2.997	2.557
StDev: Angle Error	2.198	2.574	2.102	2.195	2.573	2.100	2.197	2.574	2.103	2.194	2.574	2.101
Var: Angle Error	4.831	6.624	4.419	4.819	6.619	4.408	4.826	6.623	4.421	4.815	6.625	4.413
(triangulation)												
Image Count	684											
	Left Camera Perspective											
	Day_Avg			Day_Day			Avg_Avg			Avg_Day		
	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'
Mean: Angle Error	2.961	2.892	2.390	2.959	2.894	2.391	2.961	2.891	2.388	2.956	2.892	2.388
StDev: Angle Error	2.462	2.645	2.317	2.459	2.647	2.314	2.463	2.648	2.318	2.462	2.649	2.316
Var: Angle Error	6.059	6.997	5.367	6.048	7.004	5.356	6.067	7.009	5.375	6.060	7.018	5.364
(optimization)												
	Right Camera Perspective											
	Day_Avg			Day_Day			Avg_Avg			Avg_Day		
	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'	Right Eye	Left Eye	'Average'
Mean: Angle Error	2.950	3.357	2.648	2.952	3.360	2.650	2.827	3.259	2.534	2.827	3.262	2.536
StDev: Angle Error	4.293	3.551	3.836	4.281	3.554	3.827	3.031	2.573	2.617	3.016	2.576	2.606
Var: Angle Error	18.429	12.612	14.714	18.323	12.629	14.648	9.187	6.621	6.850	9.096	6.633	6.794
(optimization)												
Image Count	576											

7.3 Single Camera 3D Optimization Anomalies

During the review of the individual results from determining gaze from the final subject calibration images, it was noticed that several (less than 2%) of the combined left and right camera images produced significant gaze errors when using the single camera 3D optimization method to determine 3D locations. Because the errors seemed to occur randomly (see Table 14), it was hypothesized that during the single camera 3D optimization, local minima were being found instead of the actual minima. To test this hypothesis, the 3D optimization was executed in the opposite direction: starting at the optimization upper bound and decrementing, instead of starting at the lower bound and incrementing as described in Section 4. It was believed that if the optimization was indeed converging around a local minima, starting the optimization from the opposite direction might allow avoidance of the local minima and convergence to the true function minimum. Because the same results occurred, it was concluded that the issue was not a result of finding a local minima.

While looking for potential causes, it was observed that the 3D feature locations found by optimization and used to determine the head coordinate system ('hairline' marker and markers directly under the left and right eyes) were consistent with the 3D locations found using triangulation. However, the feature markers not used to determine the head coordinate system (nose and eye bridge markers) had *Z* component values in the head coordinate system with an incorrect sign as compared to the triangulation values. The incorrect *Z* component direction occurred for all the images (bad '*Z*' images) presented in Table 14.

Based on the fact that the problem manifested itself for some images from one camera perspective and not from another and when using certain calibration categories and not others, and the fact that the optimization function was derived using only the feature pair distance values derived during calibration, it was hypothesized that the problem might be a result of the optimization's sensitivity to the feature pair distance values used to derive the optimization function. It was decided to examine this sensitivity hypothesis using the images identified in Table 14. However, because there were only two images that exhibited the bad '*Z*' phenomenon from a single camera's perspective and for only some of the camera calibration categories (Subject 45 and 51), it was decided to just examine the results from these two subjects.

Table 14 Final calibration bad '*Z*' images.

Calibration	Subject	Camera Perspective	Image	Monitor Location	Angle Error
Avg:Avg Cal	4	Left Camera	3	3	36.08
Avg:Day Cal	4	Left Camera	3	3	36.05
Day:Avg Cal	4	Left Camera	3	3	35.85
Day:Day Cal	4	Left Camera	3	3	35.85
Avg:Avg Cal	8	Left Camera	7	7	16.29
Avg:Avg Cal	8	Left Camera	8	8	17.85
Avg:Day Cal	8	Left Camera	7	7	16.27
Avg:Day Cal	8	Left Camera	8	8	17.84
Day:Avg Cal	8	Left Camera	7	7	16.75
Day:Avg Cal	8	Left Camera	8	8	18.27
Day:Day Cal	8	Left Camera	7	7	16.71
Day:Day Cal	8	Left Camera	8	8	18.24
Avg:Avg Cal	37	Left Camera	5	5	35.55
Avg:Day Cal	37	Left Camera	5	5	35.59
Day:Avg Cal	37	Left Camera	5	5	35.68
Day:Day Cal	37	Left Camera	5	5	35.66
Day:Avg Cal	45	Right Camera	4	4	48.69
Day:Day Cal	45	Right Camera	4	4	48.80
Avg:Avg Cal	49	Right Camera	2	2	56.44
Avg:Avg Cal	49	Right Camera	8	8	61.70
Avg:Day Cal	49	Right Camera	2	2	56.37
Avg:Day Cal	49	Right Camera	8	8	61.61
Day:Avg Cal	49	Right Camera	2	2	56.77
Day:Avg Cal	49	Right Camera	8	8	61.95
Day:Day Cal	49	Right Camera	2	2	56.74
Day:Day Cal	49	Right Camera	8	8	61.89
Avg:Avg Cal	51	Left Camera	3	3	26.35
Avg:Day Cal	51	Left Camera	3	3	26.31
Day:Avg Cal	51	Left Camera	3	3	26.23
Day:Avg Cal	51	Left Camera	7	7	27.37
Day:Day Cal	51	Left Camera	3	3	26.19
Day:Day Cal	51	Left Camera	7	7	27.38
'Bad' Images	9		Total Images	1368	

The distance values between the various feature pairs for image 4 from the right camera perspective for Subject 45 and for image 7 from the left camera perspective for Subject 51 (see Table 15) were determined using the Avg:Avg calibration parameters (provided correct results in all cases) and were compared to those determined using the Day:Day calibration parameters (did not provide correct results).

Table 15 Bad '*Z*' feature pair distance comparison.

Subject 45 (Image 4: Right Camera)

	Calibration	Feature Pair Distances (mm)									
	Category	T/R	T/L	T/M	T/B	R/L	R/M	R/B	L/M	L/B	M/B
Triangulation	Avg-Avg	76.7343	77.5094	37.2992	89.0312	62.5682	49.1961	47.7771	49.1564	51.0041	51.927
	Day-Day	76.6097	77.4086	37.1839	88.7667	62.5289	49.146	47.6352	49.1338	50.913	51.7803
Optimization	Avg-Avg	76.2395	77.5362	36.4595	87.9303	62.7893	49.0089	47.6861	49.2937	51.7149	51.8655
	*Day-Day	76.0635	76.1896	35.9185	87.8004	63.6416	50.3652	48.3247	50.1818	50.8035	51.9174
Calibration	Avg-Avg	76.2539	76.8805	36.3926	88.0204	63.3579	49.7379	48.1155	49.4876	51.347	51.8134
	Day-Day	76.1287	76.7745	36.2746	87.7498	63.3217	49.6897	47.9776	49.4666	51.2694	51.6634

* Didn't Work

Subject 51 (Image 7: Left Camera)

	Calibration	Feature Pair Distances (mm)									
	Category	T/R	T/L	T/M	T/B	R/L	R/M	R/B	L/M	L/B	M/B
Triangulation	Avg-Avg	68.954	64.9987	36.5208	75.7182	64.6913	40.0845	45.5377	41.4137	49.7914	41.0744
	Day-Day	68.9449	64.9949	36.511	75.6709	64.6971	40.0827	45.4947	41.4172	49.7737	41.0342
Optimization	Avg-Avg	69.5338	65.4014	36.4294	76.0606	64.9297	40.7594	46.9931	41.7355	50.376	41.7231
	*Day-Day	69.0276	65.3696	35.8271	76.1246	65.286	41.1968	46.7189	42.3592	50.3239	41.8058
Calibration	Avg-Avg	69.1647	65.2627	36.4912	76.1192	65.4827	40.4551	46.9028	41.8826	50.4185	41.6608
	Day-Day	69.1585	65.2596	36.4821	76.0727	65.4901	40.4541	46.8562	41.8862	50.4015	41.6207

* Didn't Work

Table 15 also presents the 'calibration' distance values for that were used to determine the optimization function for the 3D single camera optimization methodology discussed in Section 4. These values are independent of the actual images being discussed.

Examination of the distances revealed that the distance values determined using triangulation were often larger than those determined using optimization for Subject 45, but just the opposite for Subject 51. Looking at the differences between the distance values resulting from just the optimization, it indeed appeared that very small distance differences in the distance between feature points may have drastically influenced the success or failure of the optimization. With only very slight differences in the feature pair distances that resulted, the Day:Day calibration parameters resulted in the flipped sign (bad '*Z*') issue and a gaze angle error of 48.8 degrees (Subject 45) and 27.4 degrees (Subject 51), while the Avg:Avg calibration parameters did not produce the bad '*Z*' issue and resulted in a gaze angle error of 1.8 degrees (Subject 45) and 5.0 degrees (Subject

51). There is insufficient evidence to conclude that the slight differences in distance values are an indicator of the bad '*Z*' problem.

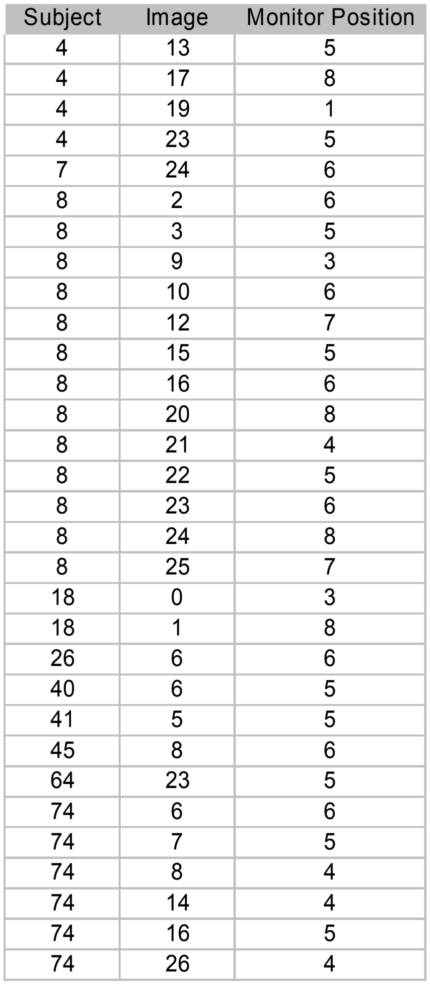
While no definitive conclusions could be drawn regarding the bad '*Z*' issue, the difference in sign between the *Z* component (in head coordinates) of either of the feature markers not used to determine the head coordinate system and the *Z* component (in head coordinates) of the same features determined during calibration appeared to be a good indicator of when the bad '*Z*' problem occurred. Because of the availability of this indicator, no additional effort to study or solve the problem was expended and a method to correct the bad '*Z*' problem has not been addressed. However, to address the bad '*Z*' issue for purposes of the calibration, the final calibration results were re-calculated after simply excluding the bad '*Z*' images.

Because the suspected sensitivity of the optimization to the feature pair distances makes it necessary that at least one of the facial features be available and used as part of the optimization but not be used in determination of the head coordinate system so that it is available to detect when the sensitivity issue manifests itself, at least four facial features (not including pupil centers) must be used in any real-world application that uses the single camera 3D optimization method. As was the case with this dissertation, if the problem is detected in a real-world application, the associated gaze determination result should be excluded from consideration.

7.4 Experiment Results

After the discovery of the bad '*Z*' issue, it was decided to calculate the gaze angle errors for the experiment images, but exclude those images that showed the bad '*Z*' phenomena because a detection method exists for excluding them (see Table 16). In addition, as with the final subject calibration, it was also decided to eliminate those images that the subject's themselves had identified during the collection of the images as potentially having stability (eyes blinking, head moving, looking away, etc.) issues. Also, a manual review of all of the images was conducted in order to identify instances where the subject's eyes were closed and/or a feature (pupil or green marker) was not visible, or significant facial deformation (smiling, laughing, etc.) had occurred. It was decided to also exclude these images from further evaluation. Finally, as an alternate to the facial feature movement/deformation check performed during calibration and discussed in Subsection 6.4, an 'optimization' check was performed.

Table 16 Experiment images identified with bad 'Z' issue (for 2009 images).



Subject	Image	Monitor Position
4	13	5
4	17	8
4	19	1
4	23	5
7	24	6
8	2	6
8	3	5
8	9	3
8	10	6
8	12	7
8	15	5
8	16	6
8	20	8
8	21	4
8	22	5
8	23	6
8	24	8
8	25	7
18	0	3
18	1	8
26	6	6
40	6	5
41	5	5
45	8	6
64	23	5
74	6	6
74	7	5
74	8	4
74	14	4
74	16	5
74	26	4

As with the movement/deformation check performed during calibration, it was believed that facial deformation during experiment image collection would manifest itself through facial feature pair distance anomalies. Unlike during calibration which used triangulation to determine locations and thus feature pair distances, feature locations during experiments were determined using an optimization (see Subsection 4.3) involving the feature pair distances. Therefore, the calibration check would be of no value during the experiment image processing.

As an alternative, a twofold check was employed. This check involved a comparison of the minimum value of *J* found during the optimization described in Subsection 4.3 for a particular image for a particular subject ($\overline{J_{\min,i,s}}$) to both an average *J* from calibration using all subjects ($\overline{calJ_{total}}$) and average *J* from calibration for that particular subject ($\overline{calJ_{s}}$). If the minimum value of *J* found during the optimization $\overline{J_{\min,i,s}}$ was greater than three[2] standard deviations above both $\overline{calJ_{total}}$ and $\overline{calJ_{s}}$, facial deformation was assumed to have occurred and the image was excluded. Table 17 provides a summary of those images that were excluded based on the 'optimization' check (bad '*D*') using 5105 for the value of $\overline{calJ_{total}}$ and 5774 as the standard deviation.

Table 18 presents a summary of all the images that were candidates for exclusion from further processing/analysis along with an indication of the prevalence of each type of issue. The table indicates how many images were involved for each category, how many subjects were involved for each category, and the maximum number of candidates for exclusion for any given subject for each category.

Unfortunately, neither stability (SI) nor facial deformation (FD) issues would be detectable, making the exclusion of images from these categories questionable. However, because the purpose of the experiments was to evaluate the underlying methods for determining gaze, not implementation issues, it was felt the exclusion of images containing issues resulting from either of these two categories was acceptable, at least for analysis purposes.

[2] The choice of three standard deviations was made on an empirical basis after studying the effect of several different values.

Table 17 Experiment images failing 'optimization' (bad 'D') check.

Subject	# Excluded
1	1
2	2
4	1
5	2
6	1
7	2
8	11
9	1
11	7
12	8
13	1
16	8
18	2
19	3
22	2
24	1
26	1
27	1
31	2
33	2
34	6
37	4
38	1
39	1
41	3
43	2
46	4
48	1
51	2
52	3
56	13
60	8
61	2
63	2
64	1
66	2
68	1
70	26
71	2
72	3
74	27
Total	173
Image Pool	2009

Table 18 Experiment rejection candidate images.

	Rejection Cause					
	Bad '*Z*'	Stability Issue (SI)	Feature Issue (FI)	Facial Deformation (FD)	Bad '*D*'	
Total Images Identified	31	295	9	121	173	Sum = 629
# of Different Subjects	10	61	7	35	41	
Max for Any Subject	13	19	2	10	27	
Total Images Rejected	520					
Total Image Pool	2009	(Subject 15 has 1 image excluded due to missing features) (Subject 17 excluded due to calibration image count criteria) (Subject 59 has 15 images excluded due to missing features)				

Table 19 provides a summary of the experiment average gaze angle error results broken out by various image removal categories. The rightmost column in Table 19 represents a quantity identical to that used in the preliminary subject calibration (see Table 11 or Subsection 6.4) to try and capture the likelihood that the subject was not actually looking at the location they reported. It represents the best gaze angle error resulting from assuming that the subject was looking at the monitor point with the smallest error.

As expected, the average overall performance as evaluated based on the magnitude of the gaze angle error decreased as suspect images were excluded. It was interesting that with the additional 260 stability issue images excluded (from FI/FD to FI/Stab/FD), there was only an approximate 0.004 degree increase in performance (gaze angle error was smaller). However, with only an additional 121 facial deformation issue images excluded (from Nothing to FD), there was an approximate 0.232 degree increase in performance. This would seem to indicate that the concerns expressed by the subjects as to their stability were unfounded or at least not as significant as the facial deformation issues. Also, the rightmost column, while only a potential indication of subject error, indicated that the difference in the average overall performance between using the reported location and using the 'minimum' location was between 0.25 degrees and 1.4 degrees. This would seem to indicate that the potential for subject error is significant.

Table 19 Experiment gaze angle errors.

Category Removed		Experiment Gaze Angle Errors			
		Right Eye	Left Eye	'Avg'	Min 'Avg'
Nothing	Mean: Angle Error	3.79	3.68	3.14	2.74
	Stdev: Angle Error	3.70	3.67	3.53	2.12
	Var: Angle Error	13.68	13.48	12.43	4.51
	Remaining Images	2009	2009	2009	2009
Bad 'Z'	Mean: Angle Error	3.54	3.45	2.90	2.64
	Stdev: Angle Error	2.74	2.81	2.57	1.78
	Var: Angle Error	7.49	7.89	6.59	3.16
	Remaining Images	1978	1978	1978	1978
Feature Issues (FI)	Mean: Angle Error	3.75	3.65	3.11	2.71
	Stdev: Angle Error	3.49	3.52	3.34	1.95
	Var: Angle Error	12.18	12.41	11.16	3.80
	Remaining Images	2000	2000	2000	2000
Stability (Stab)	Mean: Angle Error	3.74	3.64	3.08	2.71
	Stdev: Angle Error	3.66	3.56	3.46	2.10
	Var: Angle Error	13.39	12.71	11.96	4.42
	Remaining Images	1714	1714	1714	1714
Facial Deformation (FD)	Mean: Angle Error	3.57	3.45	2.91	2.65
	Stdev: Angle Error	3.40	3.25	3.16	2.00
	Var: Angle Error	11.56	10.60	9.98	4.01
	Remaining Images	1888	1888	1888	1888
Optimization' Check (Bad 'D')	Mean: Angle Error	3.43	3.31	2.78	2.59
	Stdev: Angle Error	2.80	2.76	2.59	1.70
	Var: Angle Error	7.85	7.63	6.73	2.88
	Remaining Images	1836	1836	1836	1836
FI and FD	Mean: Angle Error	3.53	3.41	2.88	2.62
	Stdev: Angle Error	3.15	3.07	2.93	1.80
	Var: Angle Error	9.93	9.42	8.60	3.26
	Remaining Images	1879	1879	1879	1879
FI, Stab, and FD	Mean: Angle Error	3.52	3.43	2.87	2.61
	Stdev: Angle Error	3.24	3.15	3.01	1.83
	Var: Angle Error	10.50	9.92	9.09	3.34
	Remaining Images	1619	1619	1619	1619
FI, Stab, FD, and Bad 'Z'	Mean: Angle Error	3.29	3.22	2.64	2.50
	Stdev: Angle Error	2.01	2.11	1.79	1.41
	Var: Angle Error	4.05	4.47	3.21	1.98
	Remaining Images	1598	1598	1598	1598
FI, Stab, Bad 'Z', and Bad 'D'	Mean: Angle Error	3.23	3.13	2.59	2.48
	Stdev: Angle Error	1.90	1.87	1.59	1.33
	Var: Angle Error	3.59	3.50	2.54	1.78
	Remaining Images	1544	1544	1544	1544
FI, Bad 'Z', and Bad 'D'	Mean: Angle Error	3.28	3.15	2.63	2.52
	Stdev: Angle Error	2.05	2.03	1.77	1.44
	Var: Angle Error	4.19	4.14	3.14	2.08
	Remaining Images	1810	1810	1810	1810
All Removed	Mean: Angle Error	3.19	3.10	2.55	2.45
	Stdev: Angle Error	1.88	1.86	1.57	1.32
	Var: Angle Error	3.52	3.46	2.46	1.74
	Remaining Images	1489	1489	1489	1489
Using Avg:Avg calibration parameters and 2009 images (Subject 15- 1, Subject 17-27, and Subject 59-15 images excluded)					

Unfortunately, there is no method by which to assess whether or not the subjects were really looking at the point they reported. Finally, there was only an approximate 0.081 degree improvement when additionally removing the stability and feature deformation issue images from the FI/Bad '*Z*'/Bad '*D*' (*FZD*) images. Because the *FZD* suspect images could be automatically determined and, therefore, do not require manual processing, only excluding these images would more accurately reflect what could be accomplished with the visual axes center method in an actual application. However, for purposes of comparison, several of the removal categories will continue to be analyzed along with the *FZD* category, although the *FZD* category will be considered the primary category for the remainder o f this dissertation.

7.5 Initial Graphical Analysis

In an attempt to take a more in-depth look at the experiment results, it was decided to graphically present the individual gaze errors rather than just simply average them. However, instead of using the gaze angle error, a notion more along the lines of a gaze distance error was presented.

In the following figures (Fig. 23, Fig. 24, and Fig. 25), the gaze vector monitor intercept locations in monitor coordinates for the experiment images are plotted on top of a series of small (white with red asterisk) squares representing the locations of the monitor gaze target points. For purposes of the figures, the monitor intercept locations can be assumed to have the same *Z* location (zero) in monitor coordinates (the figures are 2D). However, because the screen was not flat, the actual monitor locations of the monitor intercept point of the gaze vectors would have had *Z* values less than or equal to zero (the screen was curved in the negative *Z* direction in monitor coordinates). This would be an issue if gaze distance errors were being used to evaluate performance. However, because gaze angle errors are being used as the primary performance metric, the screen curvature is not an issue.

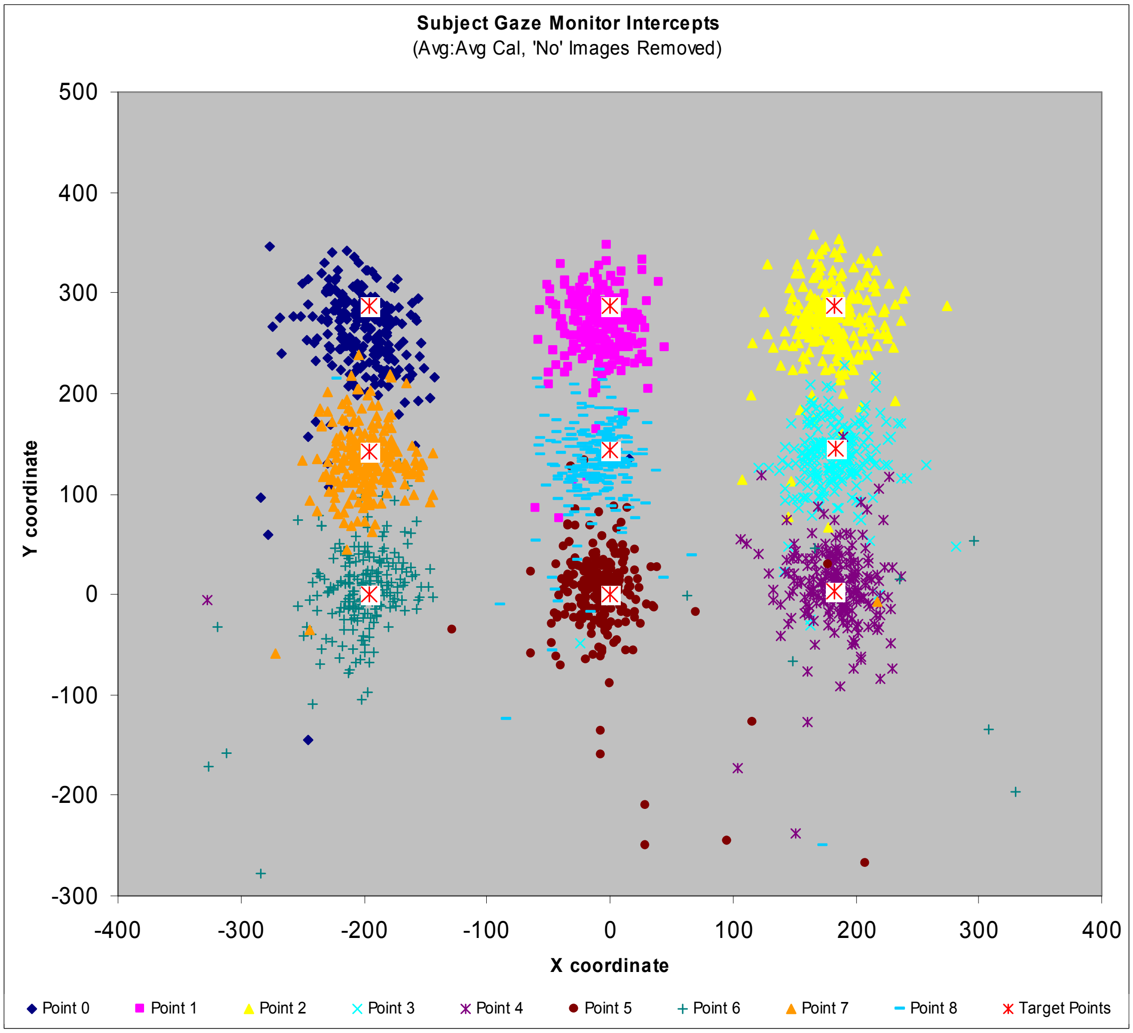


Fig. 23 Gaze monitor intercept (nothing excluded).

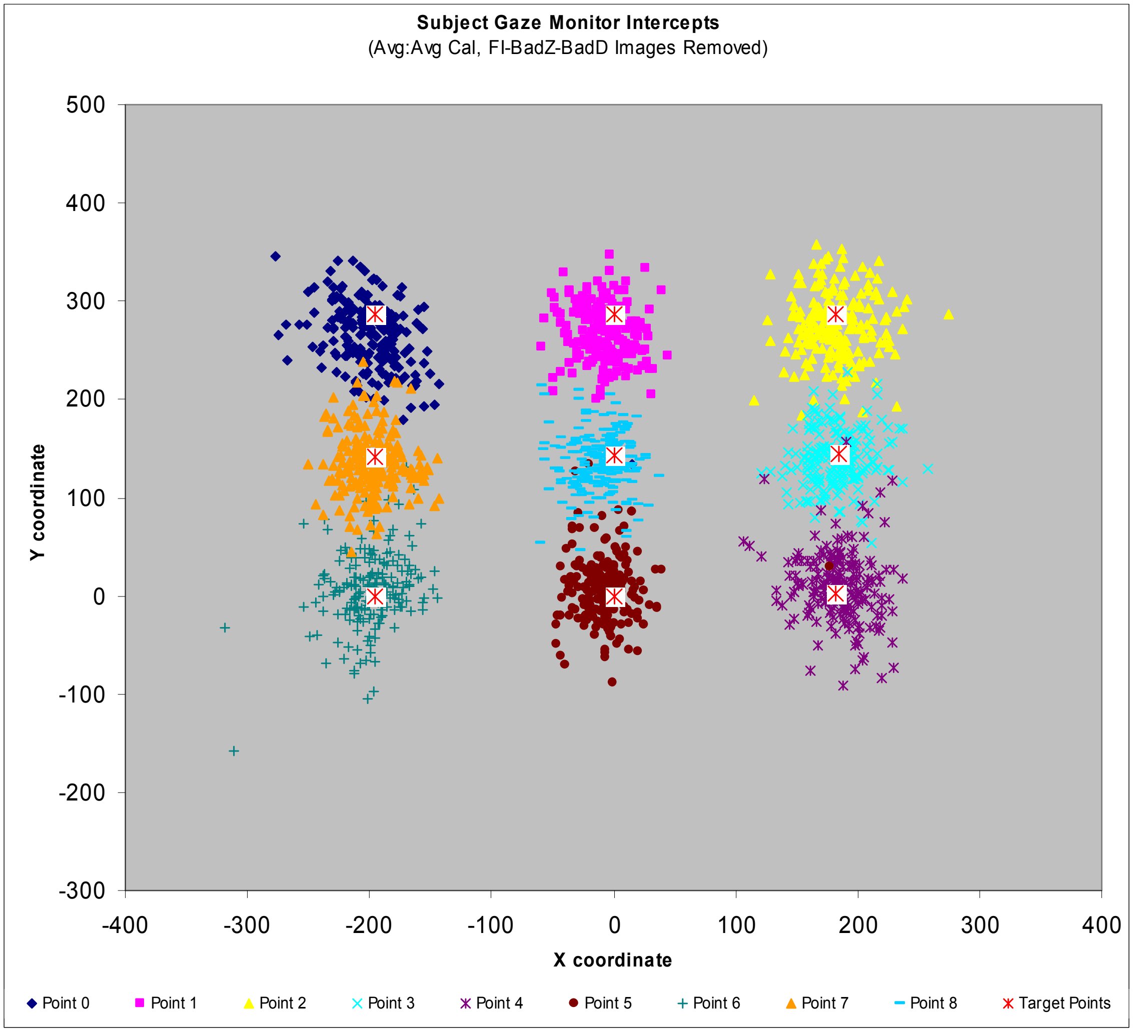


Fig. 24 Gaze monitor intercept (FI-bad*Z*-bad*D* images excluded).

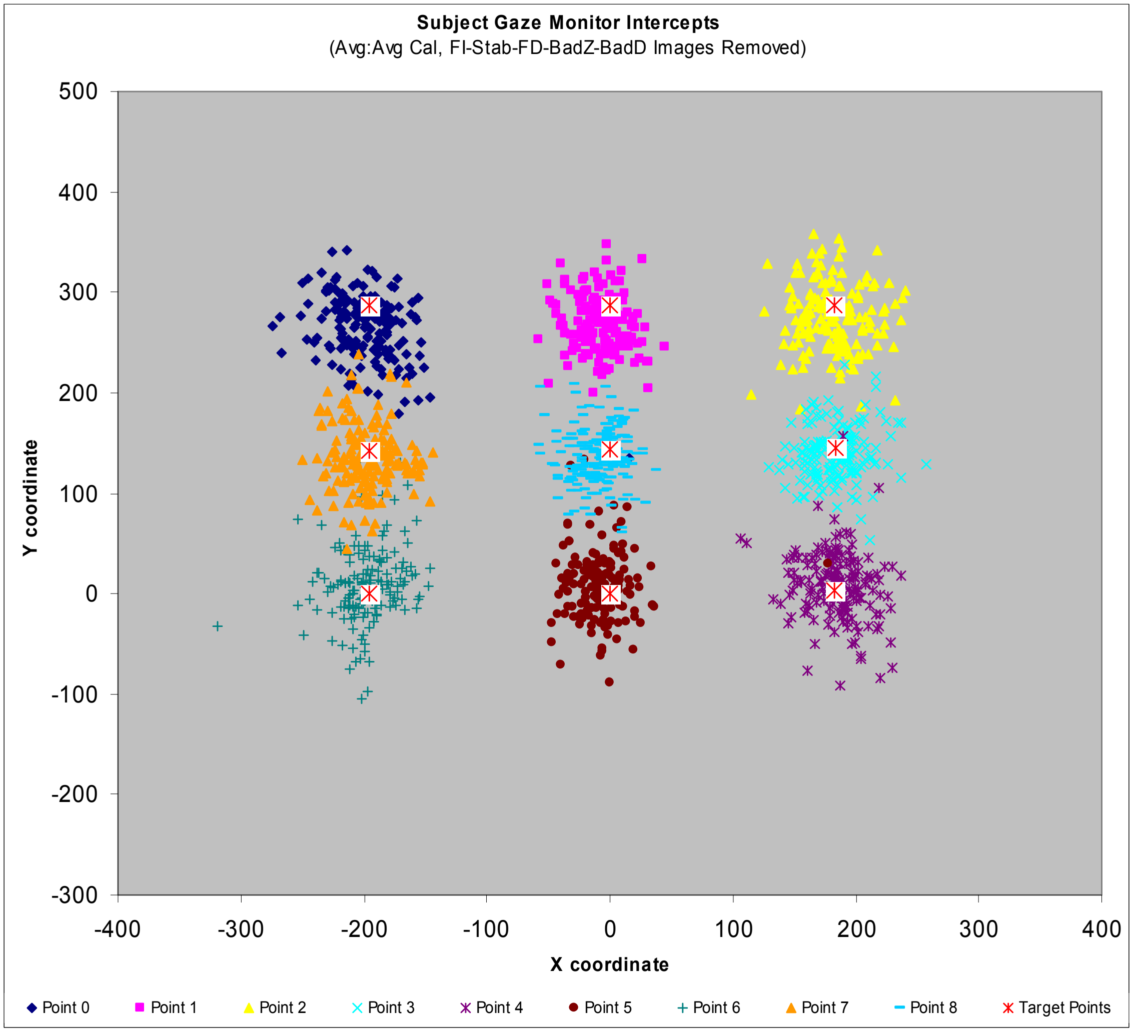


Fig. 25 Gaze monitor intercept (FI-Stab-FD-bad*Z*-bad*D* images excluded).

Although the clustering was reasonable, there were several images that resulted in gaze locations that were significantly in error. Some of these may have been as a result of subject error, but it is unreasonable to assume that they were all due to the subject looking in the wrong place. Another interesting observation was the seeming increased 'spread' associated with looking at the upper left corner of the monitor (monitor location 1). However, based on preliminary examination of the experiment results, it was suspected that there were other parameters such as the subjects' use of corrective lenses and the level of instruction received by the subjects that affected the magnitude of the gaze angle errors more significantly than the monitor location being gazed upon. Because of this, a more detailed examination of the gaze angle errors related to monitor gaze location will be presented later in Subsection 7.8 after the 'more significant' sources identified are discussed.

7.6 Protocol Modification Analysis

As was presented in Subsection 5.5, it was initially observed that the subjects remained fairly rigid (the head stayed fixed with relation to the body and the body stayed fixed with respect to the monitor) during the collection of both the calibration and experiment images. Some rigidity had been anticipated during the calibration phase, but not during the experiment phase. To address the rigidity during the experiments, additional instructions were given to the majority of the subjects participating after the fifth day of experiments (see Subsection 5.5). A random number of subsequent subjects were not given these instructions in an attempt to provide a control group for comparison with those having the additional instructions. To determine the impact, if any, of these additional instructions, it was decided to segregate the experiment results by whether additional instructions had been received or not (42 subjects received additional instructions and 34 did not) and observe the gaze angle error. The graphical representation of this segregation for the situation where the feature issue (FI), bad '*Z*', and bad '*D*' issue images were excluded is presented in Fig. 26 and Fig. 27. The average gaze angle error for those subjects who received instructions was 2.44 and was 2.90 for those who did not.

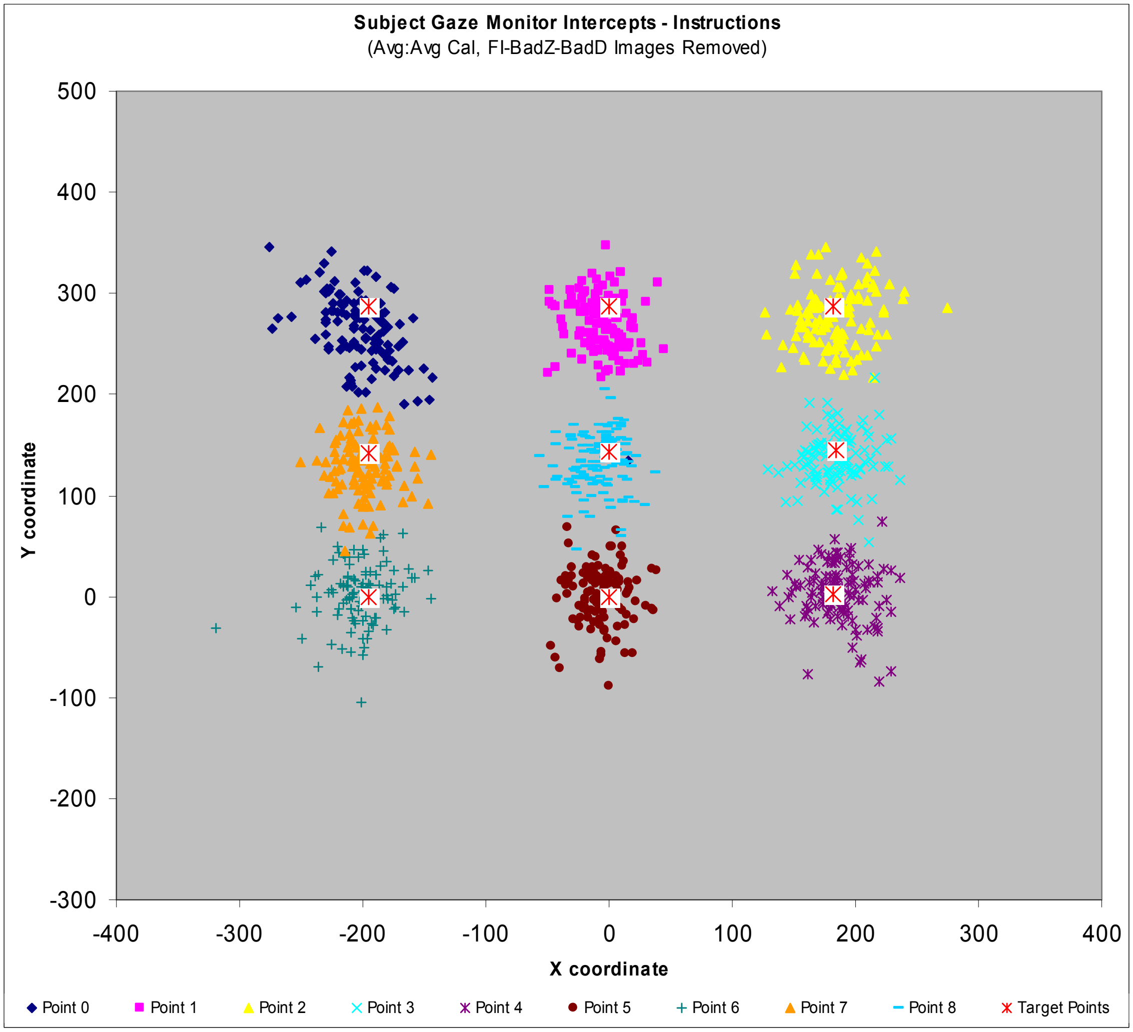


Fig. 26 Monitor intercepts (with instructions).

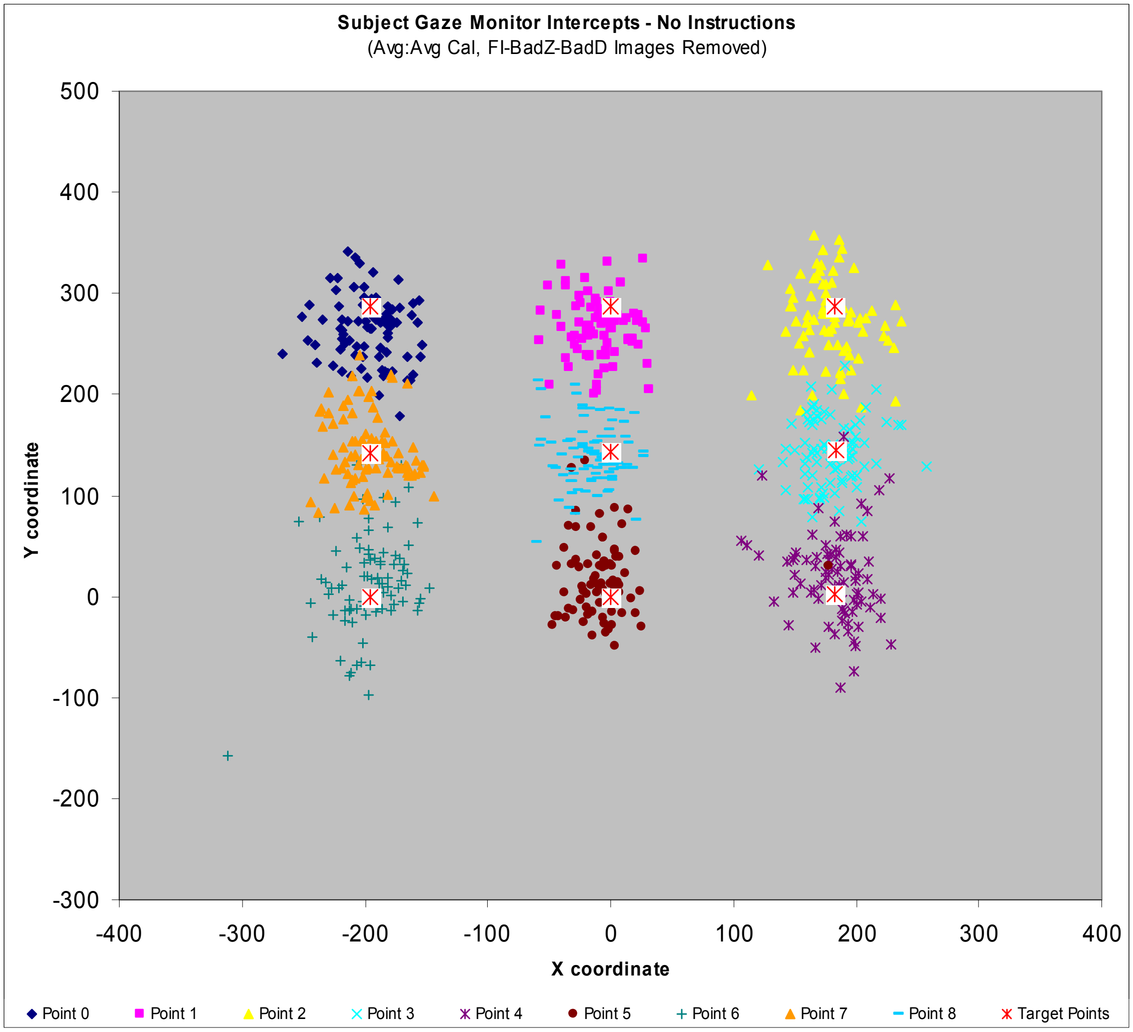


Fig. 27 Monitor intercepts (no instructions).

A t-test [92] was performed to determine if there was a statistical significance based on the gaze angle error between the experiments involving additional instructions and those that did not. The t-tests resulted in the determination of a 'p-value' representing the likelihood that the sample groups corresponding to the data sets (instructions/no instructions) came from the same population. If the sample groups came from the same population, there would be no statistical significance between the groupings, whereas, if the groups did not come from the same population a statistical significance would exist. According to Ostle and Mensing [92], a p-value of approximately 0.05 or less indicates that the data sets being used to determine the p-value did not come from the same population, and therefore, a statistical significance exists. However, there are at least two ways to determine the p-value for each t-test. One is to base the data on the number of images. The other is to base the data on the number of subjects. Unfortunately, the two methods do not result in the same, or even the same order of magnitude of p-values, though the difference is often just the degree of significance of the results. Therefore, both bases will be used whenever possible.

The resulting p-value from the t-test comparing the gaze angle errors for the instruction/no instruction experiments was 1.21 x 10^{-25} based on the number of images and 0.013 based on the number of subjects. Both of these p-values indicate that there was something statistically significant between providing and not providing the additional instructions.

In providing the additional instructions, there was also a concern that the experiments had been biased such that eye movement was discouraged (virtually eliminated) and head movement was encouraged (significantly increased). Because determining gaze resulting from primarily eye movement was believed to be more challenging, it was feared that eliminating or significantly reducing eye movement would cause the performance of the examined methods to be overstated: the average gaze angle error reported being less than what would actually exist in a real-world implementation.

To determine the impact of the instructions on eye movement, the amount of eye movement was compared between the 'instruction/no instruction' cases. An average pupil center position (*PC*) in head coordinates was determined for each eye (*e*), for each subject (*s*), for all experiment images that were not excluded. A vector between the visual axes center (${}^{Head}VAC_{e,s}$) and this average pupil center (${}^{Head}\overline{PC_{e,s}}$) was then determined:

$$\text{'}average\text{' vector} = {}^{Head}AV_{e,s} = {}^{Head}\overline{PC_{e,s}} - {}^{Head}VAC_{e,s} \tag{65}$$

The amount of angular movement away from this 'average' vector for each image (*i*) was then determined by determining the angle between the gaze vector (${}^{Head}GV_{e,s,i}$) and the 'average' vector (${}^{Head}AV_{e,s}$) for each eye:

$$angular\ movement = AM_{e,s,i} = \cos^{-1}\left(\frac{{}^{Head}GV_{e,s,i} \bullet {}^{Head}AV_{e,s}}{\left\|{}^{Head}GV_{e,s,i}\right\| * \left\|{}^{Head}AV_{e,s}\right\|}\right) \qquad (66)$$

It was this angular movement ($AM_{e,s,i}$) that was used as a measure of the amount of eye movement.

The average of the angular movement calculated for the 'instruction' subjects using Eq. 66 was approximately 5.9 degrees for the left eye and 6.1 degrees for the right eye. For the 'no instruction' subjects, there was an average movement of 6.7 degrees for the left eye and an average of 6.8 degrees for the right eye. While it appears that, as expected, there was some decrease in eye movement as a result of giving the additional instructions, it appeared that the instructions had in no way eliminated eye movement.

Head movement was also compared for the 'instruction/no instruction' cases. The amount of head movement was determined in a manner similar to eye movement, except that instead of an average pupil position being determined, the average head coordinate origin in monitor coordinates (the three point face plane centroid: $\overline{{}^{MC}CD_s}$) was determined in the middle camera (*MC*) coordinate system for each subject (*s*). The 'average' centroid was compared with the centroid (${}^{MC}CD_{s,i}$) for each of the subjects in each of the non-excluded experiment images (*i*) to obtain a distance or displacement. The displacement provided an indicator of head movement. For the 'instruction' case, the average difference between the centroids was 22.0 millimeters. For the 'no instruction' case the difference was 17.0 millimeters. As was the case with the eye movement, the instructions caused the anticipated effect: head movement increased for those subjects who were given the additional instructions. However, head movement was not increased so as to significantly reduce eye movement. Therefore, having provided the additional instructions had the intended effect of increasing head movement, but without biasing the reported gaze angle error results.

7.7 Prescription Lens Analysis

In addition to the 'instruction/no instruction' t-test, several t-tests were performed for other data groupings. Among these additional groupings were 'glasses/no glasses' and 'contacts/no corrective lenses.' During the design of the experiments, it was anticipated that the use of corrective lenses might have an impact on the gaze angle error, thus the derivation of these groups. It was anticipated that the wearing of glasses would be a significant factor affecting the gaze angle error. Not just because of the distortion and reflections caused by the lenses, but also because of the potential for the lenses, depending on the subject's orientation with the camera, to occlude the pupils and other facial features. The p-value for the glasses/no glasses comparison based on the number of images was 3.15×10^{-22} and 0.044 based on the number of subjects. Both values again indicate a statistical significance between the groups, although the subject-based p-value is close to the boundary indicating statistical significance. However, because there were only eight subjects who wore glasses during the experiments, the significance of the p-value results is questionable. In addition, of the eight subjects, five were not given any additional instructions (average gaze angle error of these five subjects was 3.66 degrees). The remaining three subjects who were given instructions had an average gaze angle error of 2.81 degrees. Due to all these factors, a more complete investigation of the use of glasses is needed before any conclusions should be made on the effectiveness of the methods being studied when glasses are involved. Such an investigation was not performed as part of this dissertation. However, for the remainder of this dissertation it will be assumed that there is a significance between the glasses/non-glasses subjects with respect to average gaze angle error.

The p-value for a contacts/no prescription lenses t-test based on the number of images was 0.411 and 0.575 based on the number of subjects. Based on the fact that both p-values strongly indicate no statistical significance, it is assumed, for purposes of further analysis with respect to gaze angle errors, that there is no statistical difference between subjects wearing contact lenses and subjects wearing no corrective lenses. Therefore, combining the results of the protocol modification t-test discussed in Subsection 7.6 and the prescription lens t-tests of this subsection, further analyses will concentrate on those subjects who received the additional instructions and who did not where glasses. Any deviations will be noted on a case by case basis.

7.8 Monitor Point 1 Analysis

As a mentioned in Subsection 7.5, the graphical presentations of the gaze monitor intercept points determined from the experiment images using the visual axes center and single camera 3D optimization methods, indicated there was a larger 'spread' associated with looking at monitor point 1 as compared to looking at other monitor points. A determination of the average gaze angle error associated with looking at the different monitor locations confirmed this observation (see Table 20).

While trying to theorize a plausible explanation for this additional 'spread,' it was observed by looking at the subjects monitor point viewing patterns, that the subjects seemed to choose a relatively non-random order to look at the monitor points (numerical order - one through nine) and that this order was repeated three times so that each monitor point was looked at a total of three times (as instructed). As a result of this non-random order, monitor point 1 was often the first point looked at by a subject during the conduct of the experiments (see Table 21).

Table 20 Average gaze angle error based on monitor location.

	Angle Error (Total)			Angle Error (Instructions)			Angle Error (No Instructions)		
Monitor Point	Mean	StDev	Var	Mean	StDev	Var	Mean	StDev	Var
0	3.122	2.000	4.002	3.207	2.202	4.847	3.003	1.684	2.836
1	2.633	1.446	2.090	2.514	1.367	1.867	2.805	1.545	2.387
2	2.744	1.525	2.325	2.580	1.270	1.614	2.981	1.813	3.289
3	2.301	1.268	1.608	2.112	1.204	1.451	2.576	1.315	1.728
4	2.509	1.660	2.756	2.160	1.243	1.546	3.007	2.024	4.097
5	2.480	1.702	2.896	2.262	1.229	1.511	2.781	2.166	4.690
6	2.718	1.849	3.418	2.342	1.372	1.884	3.216	2.248	5.052
7	2.677	2.623	6.882	2.374	1.348	1.816	3.079	3.662	13.409
8	2.479	1.383	1.912	2.383	1.305	1.704	2.618	1.485	2.204

Table 21 Subject experiment image viewing patterns.

Image	Monitor Location (Total)									Monitor Location (Instructions)									Monitor Location (No Instructions)								
Order	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	18	2	2	2	0	3	9	4	17	10	2	1	0	0	2	6	3	10	8	0	1	2	0	1	3	1	7
2	6	24	2	4	2	1	4	11	5	3	14	2	3	1	1	3	6	3	3	10	0	1	1	0	1	5	2
3	8	8	26	5	3	5	5	3	2	5	4	14	3	3	5	3	1	1	3	4	12	2	0	0	2	2	1
4	5	9	5	25	2	5	5	7	2	3	6	3	14	2	2	3	4	2	2	3	2	11	0	3	2	3	0
5	3	6	9	5	25	4	3	7	7	2	2	7	4	15	2	2	3	4	1	4	2	1	10	2	1	4	3
6	4	5	6	14	10	21	2	2	3	2	3	4	9	7	12	1	0	2	2	2	2	5	3	9	1	2	1
7	5	4	9	1	10	7	29	3	1	3	3	5	1	7	3	15	1	1	2	1	4	0	3	4	14	2	0
8	4	3	3	3	6	12	5	27	3	2	2	3	2	3	7	2	17	2	2	1	0	1	3	5	3	10	1
9	11	1	4	4	9	7	4	2	28	6	1	2	1	3	4	4	1	17	5	0	2	3	6	3	0	1	11
10	33	11	4	2	0	1	5	4	7	20	3	3	2	0	1	4	3	2	13	8	1	0	0	0	1	1	5
11	8	33	3	6	1	1	2	8	6	5	19	2	2	1	1	1	5	5	3	14	1	4	0	0	1	3	1
12	8	6	27	3	3	7	3	4	6	6	4	15	1	2	6	2	1	3	2	2	12	2	1	1	1	3	3
13	4	7	6	25	4	3	3	10	5	2	5	4	15	1	2	2	5	3	2	2	2	10	3	1	1	5	2
14	4	6	6	9	25	3	6	1	7	1	4	5	7	16	0	4	1	2	3	2	1	2	9	3	2	0	5
15	2	4	6	10	10	24	3	4	3	1	3	3	6	6	14	1	2	2	1	1	3	4	4	10	2	2	1
16	4	3	6	2	8	9	30	6	1	3	1	4	2	3	5	17	3	1	1	2	2	0	5	4	13	3	0
17	1	2	3	9	4	11	8	29	4	1	1	3	6	2	5	3	18	2	0	1	0	3	2	6	5	11	2
18	3	3	2	3	10	3	5	7	33	2	3	0	1	5	2	3	3	22	1	0	2	2	5	1	2	4	11
19	25	4	8	2	4	4	8	6	9	15	2	4	2	2	2	4	4	5	10	2	4	0	2	2	4	2	4
20	6	24	5	9	3	8	1	7	4	2	15	1	8	2	3	1	5	2	4	9	4	1	1	5	0	2	2
21	4	9	25	4	7	4	9	2	2	4	6	15	1	3	3	4	0	2	0	3	10	3	4	1	5	2	0
22	6	9	5	29	3	4	1	7	5	2	8	3	16	2	2	0	5	1	4	1	2	13	1	2	1	2	4
23	4	3	6	9	29	5	2	0	9	2	3	5	5	16	3	1	0	4	2	0	1	4	13	2	1	0	5
24	3	5	7	10	7	27	2	4	1	3	1	3	5	5	18	0	2	1	0	4	4	5	2	9	2	2	0
25	7	4	9	0	9	8	26	4	1	4	1	5	0	6	3	17	3	0	3	3	4	0	3	5	9	1	1
26	4	8	1	5	3	5	8	35	2	2	4	1	3	2	4	3	21	1	2	4	0	2	1	1	5	14	1
27	7	2	3	1	7	3	7	3	35	4	1	0	0	5	1	5	1	23	3	1	3	1	2	2	2	2	12

Theorizing that the order had more to do with the visual 'spread' than the location, the average gaze angle errors for experiment images were determined based on the order in which the images were collected (see Table 22). For a graphical presentation of the data in Table 22, see Fig. 28.

A t-test comparing the gaze angle errors for those images collected as the first image versus those images collected as the 2nd through 27th image, produced a p-value of $2.86\text{x}10^{-4}$ based on the number of images, indicating the order of image collection was indeed statistically significant (it should be noted that the image set included instruction/no instruction images as well as images of subjects wearing glasses). In fact, the general trend for the average gaze angle errors over the course of the experiment decreases during the initial portions of the experiment and then levels off toward the completion of the experiment (see Fig. 29). Fig. 29 represents a rolling window average based on image order (the window is five image order positions wide) for the average gaze angle error of each of the given image order positions. One possible explanation for this trend might be that the subjects either became more familiar and/or more comfortable with the experiment as the experiment progressed.

In an attempt to gain more insight into the impact of looking at monitor point 1 versus other monitor points, several t-tests were conducted using the same set of images that were used for the order t-test. These monitor location t-tests involved comparing the gaze angle error results for looking at monitor point 1 with the gaze angle error results from looking at all other monitor points. The first t-test simply involved a comparison of gaze angle errors when looking at monitor point 1 versus when looking at all other monitor point locations. The p-value for this t-test was $2.43\text{x}10^{-4}$, indicating there was a statistical significance. For the next t-test, any image that was the first experiment image collected for a particular subject was excluded (eliminating the 'first' image or order effect discussed previously). The first t-test resulted in a p-value of $2.63\text{x}10^{-3}$.

Table 22 Average gaze angle error by image order.

Image	Angle Error (Total)			Angle Error (Instructions)			Angle Error (No Instructions)		
Order	Mean	StDev	Var	Mean	StDev	Var	Mean	StDev	Var
1	3.570	1.878	3.527	3.497	1.859	3.457	3.677	1.942	3.771
2	2.818	1.587	2.519	2.334	1.179	1.391	3.575	1.856	3.446
3	3.258	1.654	2.737	2.930	1.222	1.493	3.750	2.078	4.317
4	2.572	1.691	2.858	2.265	1.559	2.430	3.033	1.804	3.253
5	2.768	1.693	2.866	2.238	1.146	1.314	3.544	2.055	4.221
6	2.340	1.367	1.869	2.199	1.162	1.350	2.549	1.626	2.643
7	2.672	1.558	2.428	2.632	1.569	2.460	2.724	1.570	2.464
8	2.405	2.303	5.306	2.648	2.766	7.651	2.031	1.273	1.620
9	2.800	1.469	2.158	2.781	1.675	2.806	2.824	1.186	1.408
10	2.889	1.585	2.512	2.796	1.456	2.120	3.010	1.759	3.093
11	2.469	1.440	2.073	2.510	1.295	1.677	2.406	1.660	2.755
12	2.403	1.260	1.588	2.345	1.269	1.610	2.488	1.266	1.604
13	3.071	4.156	17.269	2.474	1.463	2.140	3.903	6.162	37.969
14	2.596	1.772	3.140	2.372	1.633	2.666	2.928	1.944	3.780
15	2.592	1.336	1.784	2.538	1.263	1.595	2.665	1.449	2.099
16	2.684	1.905	3.629	2.379	1.365	1.864	3.079	2.404	5.780
17	2.608	1.439	2.072	2.441	1.272	1.619	2.837	1.636	2.675
18	2.322	1.379	1.902	2.279	1.195	1.428	2.385	1.634	2.669
19	2.475	1.631	2.661	2.119	1.225	1.500	2.949	1.977	3.907
20	2.407	1.248	1.558	2.366	1.082	1.170	2.465	1.468	2.155
21	2.735	1.697	2.881	2.347	1.304	1.700	3.263	2.026	4.105
22	2.562	1.418	2.010	2.336	1.411	1.990	2.856	1.395	1.947
23	2.485	1.539	2.369	2.280	1.257	1.581	2.770	1.850	3.422
24	2.284	1.119	1.253	2.314	1.259	1.585	2.243	0.917	0.841
25	2.377	1.361	1.852	2.302	1.303	1.698	2.477	1.453	2.110
26	2.233	1.273	1.619	2.100	1.205	1.452	2.414	1.359	1.848
27	2.760	2.330	5.430	2.123	1.490	2.219	3.670	2.968	8.809

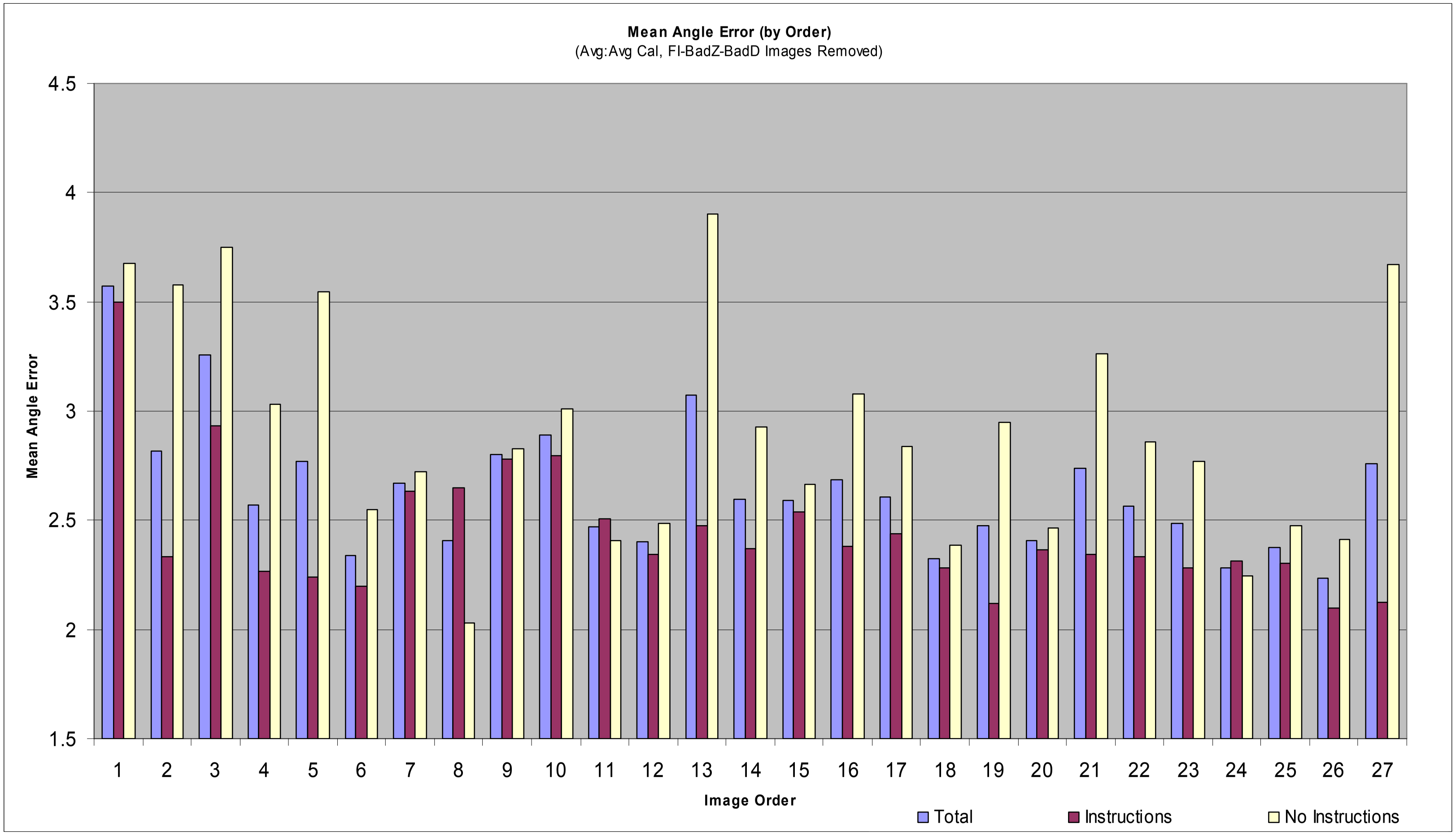


Fig. 28 Average gaze angle error by image order.

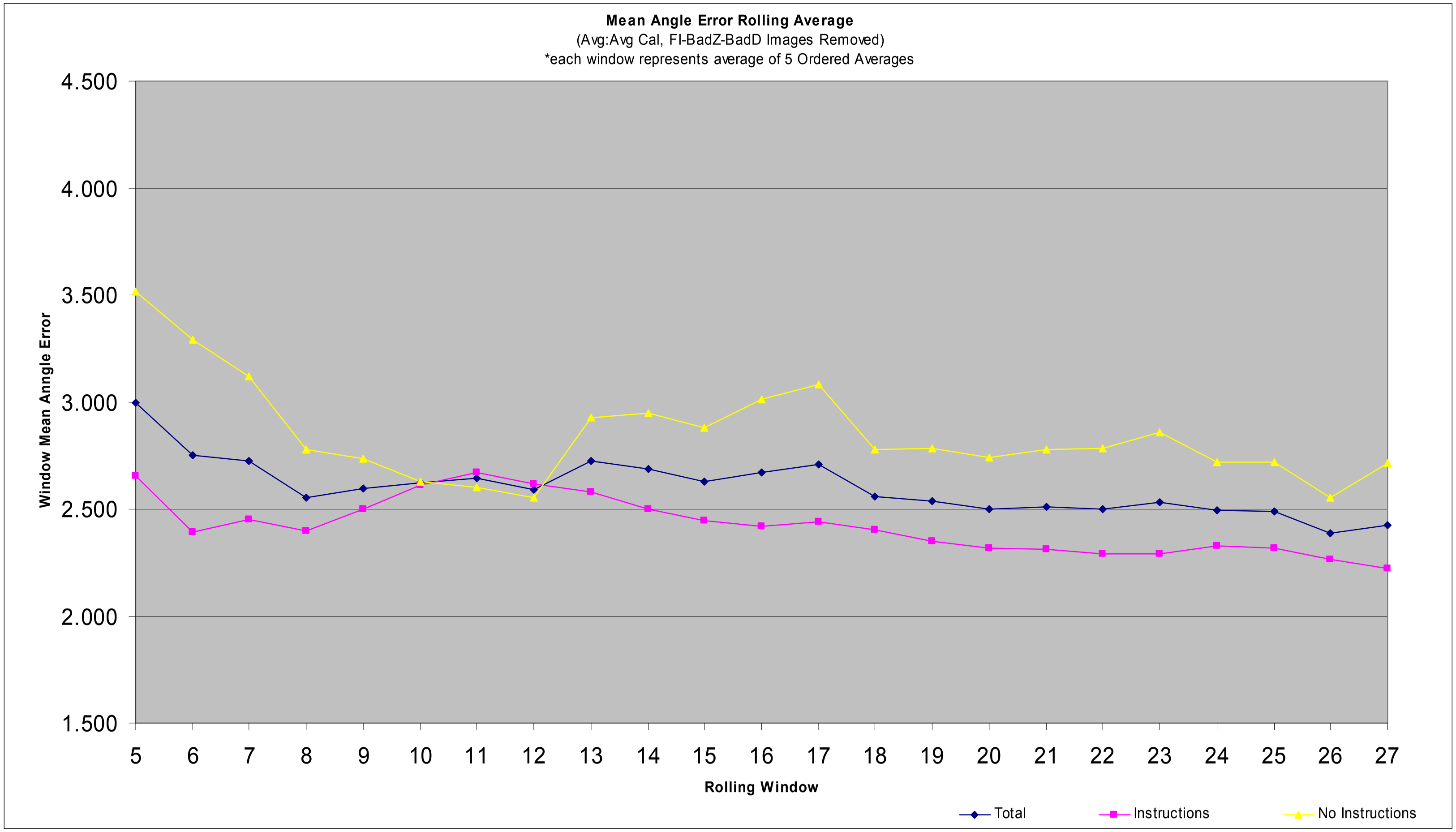


Fig. 29 Mean gaze angle error rolling window average.

The third t-test also compared gaze angle errors for subjects looking at monitor point 1 versus other points. Only in this test, images collected as the 1st through 9th images were excluded. The p-value for this t-test was $2.56\text{x}10^{-2}$.. The final t-test was for the same grouping, except only the final nine images of the experiment sequence were included for each subject. This p-value was 0.375. The sequence of these four t-test results appear to indicate that the 'spread' phenomenon had more to do with where a particular image/viewing location occurs in the image collection sequence rather than with the specific monitor point the subject was looking at. Unfortunately, a broader experiment where monitor location viewing was more tightly controlled would be required to further analyze the relationship between sequence and viewing location.

7.9 Final Gaze Angle Errors

As a result of the analysis, it was decided to re-calculate the average gaze angle error excluding those subjects with glasses and those who were in the 'no instruction' category. The individual gaze angle errors ($\alpha_{s,i}$) for each image (*i*) for each subject (*s*) were determined by finding the angles between the vectors from the midpoints of the visual axes centers (*midVAC*s) to the reported target gaze points (*TP*s) and the vectors from the midpoints of the visual axes centers (*midVAC*s) to the midpoints of the gaze intercept points (*GIP*s) (see Fig. 22). Each of these quantities was determined as described in Subsection 6.5. The individual $\alpha_{s,i}$ were then averaged for all images and subjects using the following data/results from the initial pool of 2052 images for 76 subjects:

a. only subjects who were successfully calibrated were included (now 2025 images from 75 subjects),
b. the subject calibration results found using Avg:Avg hardware calibration parameters were used,
c. only subjects who were given additional instructions to promote more natural movement between gaze positions were included (now 1134 images from 42 subjects),

d. only subjects not wearing glasses were included (20 subjects wore contacts: Subject 59 had 15 images excluded due to out of field-of-view features, now 1053 -15 = 1038 images from 39 subjects),
e. the experiment images for each subject believed to have bad 'Z' issues (four images), feature issues (four images), and bad 'D' issues (56 images, two of which also had bad 'Z' issues) were excluded (now 976 images from 39 subjects), and
f. the gaze angle error was determined using the 'average' gaze vector using a total of 976 images from 39 subjects.

The re-calculated average gaze angle error was 2.40 degrees, with a standard deviation of 1.45 degrees.

7.10 Pupil Center Location Sensitivity Analysis

In addition to the analysis of the experiment data and the determination of an average gaze angle error, it was originally planned to conduct a detailed error analysis. In general, this analysis was to have identified all of the possible contributing sources to the gaze angle error, and then evaluate the magnitude of each of the sources' contribution to the overall gaze angle error. Such efforts as deriving the distance error resulting from moving an image location by a single pixel, the physical dimension represented by an image pixel for images collected at various distances from the camera, and the magnitude of errors associated with manually locating image features were all initiated. However, as these error analysis efforts progressed, it was evident that the inter-dependency with respect to the gaze angle error of the various parameters for which errors ranges were being determined was so great, that a meaningful error component analysis was not reasonable.

Therefore, it was decided to simply determine the sensitivity of the gaze angle error produced using the visual axes center and single camera 3D optimization methods to the pixel locations of image features. While this scheme did not isolate the errors associated with measurement and with the calibration procedures, it did incorporate them, and it did attempt to emphasize those errors associated directly with the methods being studied. In addition, the measured values and calibration parameters utilized were already averages, decreasing the likelihood of a single error changing the results.

The sensitivity analysis was initiated by selecting the image feature that would impact the gaze angle error the most if its pixel location were incorrectly identified in the image. While errors in locating other facial features would ultimately impact the calculated gaze angle error, it is the pupil center (see Subsection 3.1) that was anticipated to have the most direct and significant impact on the gaze angle error because it factors directly into the determination of the gaze vector from each eye for a given image.

The general idea behind the analysis was to allow the location of the pupil center in an image to vary, and determine the gaze angle error associated with that varied pupil center pixel location. The differences between the original gaze angle error and the 'varied pixel location' (adjusted) gaze angle errors represented the sensitivity metric.

For the first phase of the analysis, the pixel location of the pupil centers was adjusted by plus or minus one pixel in either the *X* direction, the *Y* direction, or in both the *X* and *Y* directions (see Fig. 30). Only one pupil center was adjusted at a time. But during the course of the first phase, each pupil center was eventually adjusted: just not both at the same time. After each pupil center location adjustment, the adjusted gaze angle error ($\beta_{e,i,d}$) for that adjusted image *i* was determined (the subscript *e* represents the eye for which the pupil movement occurred and *d* represents the range of possible pixel movements for a given pupil center).

After all possible one pixel adjustments were made to both eyes for each image, the average adjusted gaze angle error ($\overline{\beta_{total}}$) for all images specified in the previous subsection (976 images * 8 pixel movements/pupil * 2 pupils/image + 976 original location gaze angle errors) was determined to be 2.63 degrees:

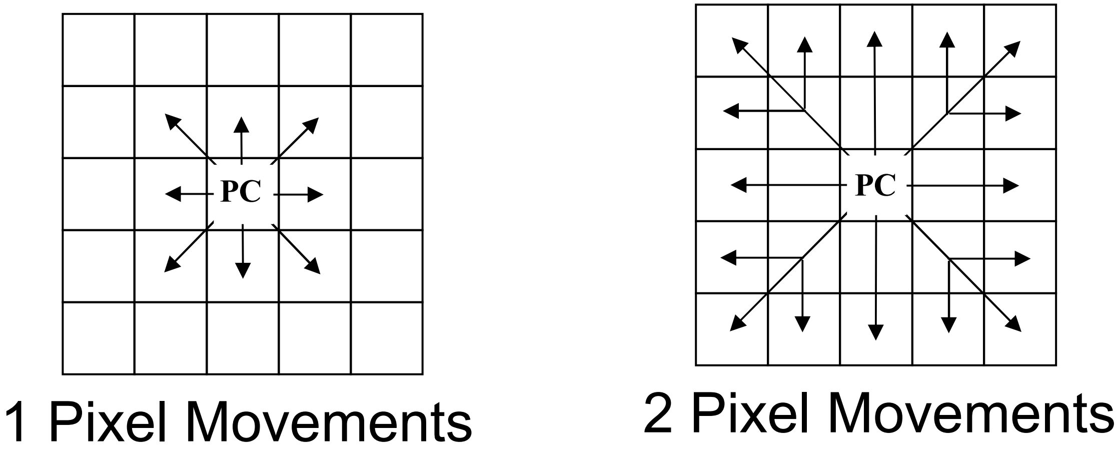


Fig. 30 Pixel movement patterns.

$$\overline{\beta_{total}} = \frac{\left(\sum_{i=1}^{976}\sum_{e=1}^{2}\sum_{d=1}^{8}\beta_{e,i,d}\right)+\sum_{i=1}^{976}\alpha_i}{(976*2*8)+976} = 2.633° \tag{67}$$

where α_i is the original gaze angle for image *i* with no pupil movement applied. As expected, $\overline{\beta_{total}}$ (2.63 degrees) is greater than the value of the original average gaze angle error ($\overline{\alpha_{total}}$) from the previous subsection, 2.40 degrees.

Re-computing the average adjusted gaze angle errors including only one pixel of movement in the *X* direction ($\overline{\beta_{1inX}}$), one pixel of movement in the *Y* direction ($\overline{\beta_{1inY}}$), or one pixel of movement in both the *X* and *Y* directions ($\overline{\beta_{1inX\&Y}}$) for both eyes results in errors of 2.58, 2.56, and 2.73 degrees respectively.

The minimum gaze angle error for each image/subject specified in the previous subsection using the original location pupil center locations, as well as all possible one pixel adjustments of both eyes was then determined ($\overline{\beta_{\min}}$). These minimum adjusted gaze angle errors for each image were then averaged across all subjects and resulted in an average value of 1.39 degrees.

In addition to the average adjusted gaze angle error ($\overline{\beta_{total}}$) and the average minimum adjusted gaze angle error ($\overline{\beta_{\min}}$), the average of the absolute value of the difference between the original gaze angle errors (α_i) for each image *i* before any pixel movement and the gaze angle error for an image *i* after pixel movement ($\beta_{e,i,d}$) was determined:

$$average\ angle\ change = \frac{\sum_{i=1}^{976}\sum_{e=1}^{2}\sum_{d=1}^{8}\left|(\alpha_i - \beta_{e,i,d})\right|}{976*2*8} = 0.744° \tag{68}$$

The '*average angle change*' parameter was interpreted to be an estimate on how much, on average, one would expect the gaze angle error to change if a one pixel change in the pupil center location was applied. The minimum and maximum '*angle change*' values were also calculated and determined to be 0.478 degrees (minimum) and 1.31 degrees (maximum), indicating the maximum and minimum gaze angle error deviation that would be expected with a one pixel error in locating a pupil center.

The entire first phase of the analysis was then repeated using a two pixel movement instead the one pixel movement (see Fig. 30). This meant that for the images/subjects specified in the previous subsection, there were 16 different pupil center locations to which each pupil center was moved. The results of both the first and second phases of the sensitivity analysis are presented in Table 23.

Table 23 Pixel change sensitivity analysis results.

	Mean	Standard Deviation	Variance	Minimum	Maximum
Overall Gaze Angle Error (degrees: 1 Pixel Δ)	2.63	1.33	1.76	_	_
Best' Gaze Angle Error (degrees: 1 Pixel Δ)	1.39	1.27	1.62	_	_
ABS(No Movement - 'Best' Gaze Angle Error) (degrees: 1 Pixel Δ)	0.744	0.11	0.012	0.478	1.31
Overall Gaze Angle Error (degrees: 2 Pixel Δ)	3.09	1.17	1.37	_	_
Best' Gaze Angle Error (degrees: 2 Pixel Δ)	1.02	0.938	0.879	_	_
ABS(No Movement - 'Best' Gaze Angle Error) (degrees: 2 Pixel Δ)	1.43	0.246	0.06	0.902	2.62

The results of the final gaze angle error determination presented in Subsection 7.9 and the pixel change sensitivity analysis presented in this subsection clearly indicate the viability of using the visual axes center and single camera 3D optimization methods to determine gaze. However, in that moving the pupil center by one pixel changes the gaze angle error, on average, by 0.744 degrees, the results also highlight the sensitivity of the methods to the accuracy of locating features in the images used; particularly the pupil centers.

8. CONCLUSIONS, CONTRIBUTIONS, AND FUTURE WORK

In this dissertation, a method to determine a human's gaze using a single camera (after calibration) was presented. The method is image-based, analytic, non-intrusive, and relies on the estimation of the visual axes center for each eye of a given subject. It involves the determination of an approximation of the visual axes center for a subject through image processing of pseudo-stereo images collected during a calibration phase. A 3D visual axes center is approximated by the mean of the most likely intersection points of a collection of vectors from known target gaze points through pupil centers located from captured images. The visual axes center location is then transformed into a head coordinate system derived using several facial features that are non-deformable with respect to each other. The spatial relationships of the facial features, also determined during calibration, are then utilized to facilitate the post-calibration determination of the 3D locations of facial features from images collected from a single camera. With the facial features located, the visual axes center is also defined. Utilizing the assumption of a constant distance between the visual axes center and the pupil center allows the determination of the pupil center in 3D. Projecting from the visual axes center through the pupil center defines a gaze vector.

The fundamental goal of this research was to develop and experimentally validate a method of gaze vector determination that could potentially be low cost and utilize only a single camera, after an initial calibration, the tradeoff being the acceptance of slightly less accuracy. This tradeoff was deemed acceptable as there are many applications that do not require high accuracy. The experiments that were conducted show that the fundamental goal has been achieved. The experiments also indicated a number of interesting aspects to the use of the visual axes center method. In particular, the issues of human variability and reliability, the effects of glasses/contacts/no glasses on the method, the approximations required in determining the *VAC*, the impact of eye dominance on the method, the effects of providing instruction/no instruction regarding head and body movement, accuracy of hardware and subject calibration, and the sensitivity of the gaze angle to pupil center location.

8.1 Discussion

Dealing with humans in experiments such as conducted in this research is fraught with possible anomalies that might distort the results but are not related to the fundamental method being tested. Things such as a subject looking at the wrong target point, or changing their facial expression in a way that moves the facial features used to identify the head location and orientation, or glasses sometimes occluding part of a feature that must be accurately located can all cause difficulties. For testing purposes, considerable care was taken to identify instances of such anomalies. Those cases identified were removed from the experimental data to help determine what could be achieved under reasonably good circumstances. In many practical applications, such as the examination monitoring application that motivated this research, it is not necessary that every image processed yield usable results, only that useful data be obtained with reasonable frequency. What is important, however, is the ability to detect the anomalies so that their occurrence can be neglected. Alternatively, further research might identify additional features or feature location methods that improve upon the ability to accurately locate a subject's head.

The most significant aspect of the human subject testing discovered was completely unexpected. The 'instruction versus no instructions' differences that were found were statistically very significant. Yet, it was originally perceived as only a minor change to get the subjects to relax a bit more during the tests. The test data showed no significant difference in the amount of head or eye movement. What causes the difference in accuracy between these two cases is still not understood. The best guess is that it has something to do with how relaxed and natural the subject felt during the test. Understanding this better may be coupled with a better understanding of the effects of prolonged used of the method.

The tests also showed that there was a definite trend toward improved accuracy as the number of images viewed increased, though this trend was more pronounced for the 'instructions' case. Again, the cause for this is not well understood, and the best guess is again that the effect is related to how comfortable the subjects are with the conduct of the experiment. Nevertheless, when anomalous cases and the 'no instruction' cases are removed, an average gaze error of 2.44 degrees was achieved. While not in the realm of being considered 'high accuracy' (less than one degree of angular error), the results were well within those needed to support the exam proctoring application that motivated this research.

Since anomalous behavior does occur with human beings, it is important to be able to recognize situations in which the data should be considered suspect and new data obtained. Fortunately, the analysis of the experimental data also led to ways to perform such detection. One of the most obvious is whether or not the subject was wearing glasses. Utilizing the visual axes center method to determine subject gaze in experiments conducted with 39 subjects who did not wear glasses ('instructions' and 'no instructions') and were not excluded for other anomalous reasons, resulted in average gaze direction angular errors of less than 2.6 (~ 2.56) degrees. On the other hand, the average gaze direction error for those subjects wearing glasses was 3.25 degrees. Interestingly, the impact of wearing contacts was less clear and those wearing contacts were included in the final analysis. This is probably because most of the difficulties associated with the wearing of glasses seemed to be due to partial occlusion of some of the facial features from some part of the glasses. That also suggests that a different choice of facial features might make the method equally useful for people wearing glasses.

From the experiments, it was also concluded that the notion of eye dominance played virtually no role in determining gaze. The gaze angular errors calculated using individual eyes showed virtually no difference in the results using the dominant eye (3.22 degrees) versus using the non-dominant eye (3.17 degrees), indicating that eye dominance seems to play little, if any, role during gaze fixation.

The use of the pupil center as a substitute for the nodal point in determining gaze vectors in the visual axes center method had been predicted to add an additional 0.49 degrees of error. While the magnitude of error actually contributed by the pupil center substitution is unknown, the observed magnitude is deemed acceptable given the acceptability of the resulting average gaze angle errors. Therefore, it is concluded that the pupil center provides a reasonable approximation for the nodal point in the visual axes center gaze determination method.

Finally, an estimate of the sensitivity of the method to pupil center location accuracy was determined: a one pixel movement of the pupil center location resulted in an average of 0.744 degrees of difference in the determined gaze direction. More importantly, if one uses the pupil center location in the 3x3 pixel window about the original pupil center location that yields the lowest angle error, an average minimum gaze angle error of 1.39 degrees is obtained. The results emphasize the importance of accurately determining the pupil center in the image and suggest that if sub-pixel accuracies can be achieved with better image processing, there is considerable room for improvement in the method.

8.2 Research Contributions

The primary achievement of this research has been the development of a methodology to determine 3D gaze locations without the continuous use of stereo cameras or the need for specialized illumination such as that needed to obtain consistent corneal reflections. Moreover, this is accomplished with the use of a single low cost COTS camera (after calibration). The method uses the notion of a visual axes center which has been stated in the literature to be fixed with respect to the head and to be relatable to the fixation object through the anterior nodal point. An error analysis shows that use of the pupil center to approximate the nodal point has an acceptable error, and this approximation was validated through experimentation. In order to successfully use a single camera, a novel approach of utilizing known (from measurement or calibration) distances of a set of facial features to determine the 3D locations of these features from a single image was introduced.

Experimental data has validated the approach and approximations used. Moreover, it has led to identification of on-line techniques for identification of anomalous images, allowing them to be discarded and not influence the application. Since the intended application domain does not include the instantaneous tracking of gaze, this allows subsequent images to be used. Moreover, the experimental analyses point to a number of future studies that are likely to yield important additional results.

8.3 Future Work

The gaze determination method presented in this dissertation clearly provides adequate capabilities for the exam monitoring application and the similar class of applications for which it was proposed. However, several questions surfaced during the conduct of this research whose answer could provide valuable insight and possibly enhance any implementation efforts. The remainder of this subsection highlights some of the primary activities that have a high potential for yielding information important for the practical implementation of the method or for better understanding of the underlying principles.

8.3.1 Automatic Feature Identification

In order to eliminate the need for manual processing of images, methods that accurately locate human faces and subsequently human features such as eye corners, nostrils, lip corners, pupils, etc. should be developed/incorporated with the visual axes center gaze determination method. Subsection 2.5 discusses several possible techniques for developing accurate face localization, as well as feature finding capabilities. However, these must be successfully implemented such that they are able to robustly identify, given a reasonable quality webcam image, a single point representation for the feature using sub-pixel accuracy.

In addition, efforts should be made to investigate the spatial deformation potential between the locatable features. If spatially non-deformable features cannot be reliably located, alternate techniques to robustly and accurately determine the 3D location of the head must be found, e.g., higher redundancy in the number of features.

8.3.2 Single Camera 3D Estimation

The method used to determine 3D locations using a single camera relies heavily on the availability of physical distances between object features that are always visible in the images to be processed. In this dissertation, the relationship between five feature points (10 distances) determined during calibration using pseudo-stereo images was used to estimate the 3D locations of the corresponding feature points in subsequent images from a single camera. The ability to utilize feature distances that are physically measured instead of relying on distances determined using stereo triangulation should be investigated, as this could eliminate the need for multiple cameras entirely for subject calibration (if done in combination with the revised pupil center determination discussed below).

In addition, the minimum number of feature points required to obtain a good estimate for the 3D facial feature locations should be investigated. Would having higher redundancy (in the sense of more inter-feature distances) provide a better estimate? This should be investigated from both a theoretical and practical standpoint.

Finally, the ability to estimate the visual axes center for a given eye of a given subject without the need for multiple cameras should be investigated. During the test phase of the experiment the 3D locations of the facial features were determined using an optimization based on knowing a set of inter-feature distances. If one uses calibration images corresponding to gazing at multiple fixation points, it may then become possible to modify the J function so that minimizing it allows one to determine the VAC and dVP directly. This technique would derive a function J' based on the premise that gaze vectors should meet at the visual axes center and the distances between the VAC and the pupil centers remain constant. J' would be similar to the current function used for single camera 3D feature location, but would produce the *X*, *Y*, and *Z* location of the *VAC* and the distance to the pupil. It is complicated, however, by the fact that the head may move from one calibration image to the next. Hence, it would be necessary to first determine the head coordinate system as in the present method and use this to remove the effects of head movement.

8.3.3 Facial Deformation Estimation

In Subsection 6.4, a simple experimentally determined technique for estimating the likelihood of facial deformation during calibration was presented. A similar technique involving the deviation from a mean *J* value determined during calibration was then used for the actual experiment phase. The adequacy and effectiveness of both of these techniques needs to be investigated further. If adequate, efforts to justify the acceptance criteria (five or fewer features deviating from the mean by 1.25 standard deviations or more during calibration and function values less than three standard deviations from both the mean for the subject and the mean from all subjects) needs to be put forth. Alternatively, one might look at different approaches to detect facial distortion, such as identifying additional facial features that are more likely to indicate facial deformation, e.g., lip corners and eyebrows.

8.3.4 Bad '*Z*' Phenomenon

During the determination of the 3D facial feature locations using the single camera 3D location method, it was observed for ~ 2% of the images, that the 3D locations determined were in error. When represented in head coordinates, the *Z* component of several of the feature locations exhibited an incorrect sign when compared with calibration values. Given the closeness of the magnitude of the *Z* component values to the calibration values, one suspects that there might be some sort of reflective set of solutions, at least one of which has the opposite *Z* components. However, the phenomenon should be investigated further in an attempt fully understand what is happening.

8.3.5 Gaze Angle Error vs. Gaze Location

As discussed in Subsections 7.5 and 7.8, the gaze angle errors were strongly influenced by their position in the sequence of points viewed by the subjects, and there was some indication that one particular location might be having an effect. These issues need to be better understood.

For example, how would longer sequences impact the very significant 'instruction versus no instruction' condition results that were obtained? Does the difference between the two diminish or disappear after long enough use? Are there other experimental conditions that help determine the reasons for the difference between the 'instruction versus no instruction' conditions?

The apparent impact of location one on the results also bears further investigation. There is some indication that this effect will disappear as the sequence becomes longer. However, conducting experiments with different ordering of the test locations might help determine if the effect is real or not. However, designing experiments that do not introduce other, unintended, effects, such as the difficulty in dealing with randomly ordered target points or the cultural habit of reading from left to right, top to bottom, may be difficult and require the involvement of cognitive science specialists.

8.3.6 *VAC*/Pupil Center/Nodal Point Relationships

During the course of preparing this dissertation and performing the research associated with it, the lack of consistent/definitive information with respect to the optical features of the human eye and the relationships between these optical features became evident. Many authors referenced eye models with which to perform studies, but even the parameters used for these models varied from author to author, and bounds for many of the parameters were not specified. As a result, the error estimations regarding the substitution of the pupil center for the nodal point and the impact of the variation of the distance between the visual axes center and the pupil center are not as good as might be desired. Therefore, a more definitive representation of the eye is needed in order that a thorough error estimation can be obtained. A more thorough error analysis may facilitate a theoretical justification for the use of the visual axes center gaze determination method rather than the experimentally-based one presented in this dissertation. These investigations, however, lie in the domain of the eye physiologists.

8.3.7 Camera Calibration

Prior to the conduct of the experiments for this dissertation, it was anticipated that a single camera calibration sequence for each camera would be sufficient to characterize the cameras throughout the conduct of the experiments. However, as documented in Subsection 6.3, there were inconsistencies between the daily calibration results that could not be explained. The fact that there seemed to be accuracy/precision discrepancies between the cameras (particular between the right and the left), leads one to believe the camera calibration issues are not fully understood. In addition, it was speculated that which communication channel (which USB channel in this case) that the camera was connected to may have played a role. Also the camera power on/off duration may also have had an impact. Therefore, efforts are needed to try and further study the calibration process incorporating variables such as camera power on/off duration and camera communication channel connection. In addition, incorporating metrics based on the quality and consistency of the images should also be considered.

8.3.8 Subjects Wearing Glasses

As stated previously, the average gaze angle error of 2.40 degrees for the experiments conducted for this dissertation involved subjects who were not wearing glasses. However, for the entire subject pool of the experiment, there were eight subjects who wore glasses. In visually analyzing the images from these subjects and comparing the visual analysis to the gaze angle error result calculations, it was clear that the primary issue with glasses and the use of the visual axes center gaze determination method was one of feature occlusion (or partial occlusion). For those subjects who were wearing glasses, poor gaze angle error results occurred almost exclusively (neglecting facial deformation issues) when one or more of the features were occluded by the frames or occluded and partially distorted by a lens. While believed to be an issue of any feature-based technique, some effort should be extended to more fully characterize the issue and, if truly an occlusion/distortion problem, investigate methods to use alternate or additional features to overcome the problem.

8.4 Conclusion

Assuming the additional efforts suggested in the previous subsections were expended and satisfactory results were implemented, it is believed that the visual axes center method in combination with the single camera 3D determination techniques presented in this dissertation would provide an adequate, low cost foundation with which to actually implement many applications requiring gaze determination. Particularly, those applications like the exam proctoring application that do not require high levels of accuracy.

REFERENCES

[1] A. T. Duchowski, "A Breadth-First Survey of Eye Tracking Applications," Behavior Research Methods, Instruments, & Computers, vol. 34, pp. 455-470, 2002.

[2] Technology and Systems for Eye Tracking 6000 Series, accessed on 4/19/06, http://www.a-s-l.com/5000_series.htm, Applied Science Laboratories.

[3] V. Bakic and G. Stockman, "Real-time Tracking of Face Features and Gaze Direction Determination," in: 4th IEEE Workshop on Applications of Computer Vision, 1998.

[4] V. Bakic, "An Interface for Human-Computer Interaction Based on Face Feature Tracking in 2D," in Computer Science and Engineering, Ph.D. Michigan State, 2000, pp. 277.

[5] National Institutes of Health National Eye Institute, accessed on 3/19/2005, "Eye Diagram," 300dpi, eye12-300.tif, ftp.nei.nih.gov/eyean/eye12-300.tif.

[6] Rods and Cones, accessed on 5/6/2006, http://www.cis.rit.edu/people/faculty/ montag/vandplite/pages/chap_9/ch9p1.html.

[7] "Facts and Figures Concerning the Human Retina," accessed on 3/19/2005, http://www. webvision.med.utah.edu/facts.html.

[8] B. Noureddin, P.D. Lawrence, and C.F. Man, "A Non-contact Device for Tracking Gaze in a Human Computer Interface," Computer Vision and Image Understanding, vol. 98, pp. 52-82, 2005.

[9] A. J. Glenstrup and T. Engell-Nielsen, "Eye Controlled Media: Present and Future State," in Computer Science. Copenhagen, University of Copenhagen, 1995, http://www.diku.dk/~panic/eyegaze/article.html.

[10] Digital Image Definitions, accessed on 5/17/2006, http://www.ph.tn.tudelft.nl/ Courses/ FIP/frames/fip.html.

[11] Introduction, accessed on 5/17/2006, http://www.ph.tn.tudelft.nl/Courses/FIP/ frames/fip.html.

[12] C. H. Morimoto and M. R. M. Mimica, "Eye Gaze Tracking Techniques for Interactive Applications," in: Computer Vision and Image Understanding, Academic Press, Orlando, 2005.

[13] Gaze Direction Determination, accessed on 5/7/2006, http://homepages.inf.ed. ac.uk/rbf/CVonline/LOCAL_COPIES/WANG2/CVonline.htm, J. Wang and E. Sung.

[14] The Free Dictionary, accessed on 5/17/2006, http://medical-dictionary. thefreedictionary. com/gaze, Farlax.

[15] R. H. S. Carpenter, Movements of the Eyes, 2nd ed. Pion Ltd., London, 1988.

[16] H. Ono, A. P. Mapp, and R. Barbeito, "What Does the Dominant Eye Dominate? A Brief and Somewhat Contentious Review," Perception & Psychophysics, vol. 65, pp. 310-317, 2003.

[17] R. Barbeito, "Sighting Dominance: An Explanation Based on the Processing of Visual Direction in Tests of Sighting Dominance," Vision Research, vol. 21, pp. 855-860, 1981.

[18] Modeling Off-axis Vision - I: The OPTICAL Effects of Decentering Visual Targets or the Eye's Entrance Pupil, accessed on 4/28/2006, http://research.opt.indiana.edu/Library/ModelOffAxisI/ModelOffAxisI.html, A. Bradley and L. N Thibos.

[19] Visual Optics, Lecture 28, accessed on 5/18/2006, http://www.opt.indiana.edu/optlib/Class%20of%202007%20General%20Notes/V663%20Visual%20Optics%202004-%2028(Blaine).doc, Blaine.

[20] L. N. Thibos, A. Bradley, D. L. Still, X. Zhang, and P. A. Howarth, "Theory and Measurement of Ocular Chromatic Aberration," Vision Research, vol. 30, pp. 33-49, 1990.

[21] R. S. Park and G. E. Park, "The Center of Ocular Rotation in the Horizontal Plane," American Journal of Physiology, vol. 104, pp. 545-552, 1933.

[22] Cardinal Point (optics), accessed on 5/17/2006, http://en.wikipedia.org/wiki/Cardinal_point_(optics)#Nodal_points, The Free Encyclopedia Wikipedia.

[23] F. Martin, "The Importance and Measurement of Angle Alpha," British Journal of Physiological Optics, vol. 3, pp. 27-45, 1942.

[24] G. E. Park and R. S. Park, "Further Evidence of a Change of Position of the Eyeball During Fixation," Archives of Opthamology, vol. 23, pp. 1216-1230, 1940.

[25] D. Scott and J. M. Findley, "Visual Search, Eye Movements and Display Units," University of Durham, Human Factor Report 1993.

[26] P. Hallett, "Chapter 10," in: Eye Movements: Wiley, New York, 1986, pp. 25-28.

[27] M. L. Córdoba, A. Pérez, A. García, R. Méndez, M. L. Muñoz, J. L. Pedraza, and F. Sánchez, "A Precise Eye-Gaze Detection and Tracking System," in: 11th International Conference in Central Europe on Computer Graphis, Visualization and Computer Vision (WSCG), Plzen-Bory, Czech Republic, 2003.

[28] S. Baluja and D. Pomerleau, "Non-intrusive Gaze Tracking Using Artificial Neural Networks," CMU CS Technical Report, vol. CMU-CS-94-102, 1994.

[29] Smart Eye Pro, accessed on 4/19/06, http://www.smarteye.se/smarteyepro.html, Smart Eye.

[30] iView X HED, accessed on 4/13/2006, http://www.smi.de/iv/index.html, Senso Motoric Instruments GmbH.

[31] Eye Gaze Tracking, accessed on 4/19/06, http://www.is.cs.cmu.edu/mie/eyegaze.html, ISL.

[32] EYEGAZE - A Straight Line to the Future of Computer Technology, accessed on 4/19/06, http://www.eyegaze.com/, LC Technologies, Inc.

[33] faceLAB 4, accessed on 4/19/06, http://www.seeingmachines.com/, Seeing Machines.

[34] K. Tan, D. J. Kriegman, and N. Ahuja, "Appearance-based Eye Gaze Estimation," in: IEEE Workshop on Applications of Computer Vision, 2002, http://citeseer.ist.psu.edu/tan02appearancebased.html.

[35] Product Description Model 50, accessed on 4/16/2006, http://www.tobii.se/downloads/Tobii_50series_PD_Aug04.pdf, Tobii.

[36] Eye Movement Equipment Database, accessed on 3/19/2005, http://ibs.derby.ac.uk/cgi-bin/emed/emedsrch.cgi?opr1=OR&fld1=name&key1a =*.

[37] Fundamentals and Applications of Scleral Search Coils, accessed on 4/13/2006, http://www. skalar.nl/epmintro.html.

[38] D.A. Robinson, "A Method of Measuring Eye Movements Using a Scleral Search Coil in a Magnetic Field," IEEE Transactions on Biomedical Engineering, vol. 10, pp. 137-145, 1963.

[39] A. Kaufman, A. Bandopadhay, and B. Shaviv, "An Eye Tracking Computer User Interface," in: Research Frontier in Virtual Reality Workshop, 1993.

[40] J. G. Daugman, "High Confidence Visual Recognition of Persons by a Ttest of Statistical Iindependence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp. 1148--1161, 1993.

[41] A. L. Yuille, David S. Cohen, and Peter W. Hallinan, "Feature Extraction from Faces Using Deformable Templates," in: IEEE Computer Vision and Pattern Recognition, San Diego, CA, 1989.

[42] H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance," International Journal of Computer Vision, vol. 14, pp. 5-24, 1995.

[43] J. Zhu and J. Yang, "Subpixel Eye Gaze Tracking," in: 5th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, D.C., 2002.

[44] Stereo Vision: Triangulation, accessed on 4/20/06, http://www.dis.uniroma1.it/ ~iocchi/ stereo/triang.html, L. Iocchi.

[45] History Of EyeTracking Technology, accessed on 3/18/2005, http://www. eyemouse.com/3Solutions/HistoryofET.htm.

[46] H. Koesling, "Visual Perception of Location, Orientation and Length: An Eye-Movement Approach," in: Neuroinformatics Group and the Collaborative Research Center. Bielefeld, University of Bielefeld, 2003, pp. 304, http://bieson.ub.uni-bielefeld.de/volltexte/2003/244/pdf/diss.pdf.

[47] R. Canosa, "Seeing, Sensing, and Selection: Modeling Visual Perception in Complex Environments," in Center for Imaging Science. Rochester, Rochester Institute of Technology, 2003, pp. 1-249, http://www.cs.rit.edu/~rlc/ Dissertation/.

[48] Eye Tracking in Advanced Interface Design, accessed on 3/18/2005, http://www. cs.tufts.edu/~jacob/papers/barfield.html, Robert J.K. Jacob.

[49] H. B. Crane and C. M. Steele, "Accurate Three-dimensional Eyetracker," Applied Optics, vol. 17, pp. 691-705, 1978.

[50] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, "Real-Time Stereo Tracking for Head Pose and Gaze Determination," in: International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000.

[51] K. R. Park, J. J. Lee, and J. Kim, "Gaze Position Detection by Computing the Three Dimensional Face Positions and Motions," Pattern Recognition, vol. 35, pp. 2559-2569, 2002.

[52] Y. Matsumoto and A. Zelinsky, "An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement," in: 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000.

[53] Human Face Processing: A Survey, accessed on 5/7/2006, http://ugweb.cs.ualberta.ca/~ayman/face/faceSurvey.htm#4.1%20Deformable%20Template%20 Matching, A. Ammoura.

[54] P. Bilek, "Face Localization From Disciminative Regions (Thesis Proposal)," in: Electrical Engineering, Master's, Czech Technical University, Prague, 2001, pp. 31, http://citeseer.ist.psu.edu/cache/papers/cs/29453/ftp:zSzzSzcmp.felk.cvut.czzSzpubzSzcmpzSzarticleszSzbilekzSzBilek-TR-2001-24.pdf/face-localization-from-discriminative.pdf.

[55] K. Sung and T. Poggio, "Example-based Learning for View-based Human Detection," MIT A.I. Lab, Technical Report Technical Report 1521, 1994.

[56] T. Sakai, M. Nagao, and S. Fujibayashi, "Line Extraction and Pattern Detection in a Photograph," Pattern Recognition, pp. 233-248, 1969.

[57] G. Yang and T. S. Huang, "Human Face Detection in a Complex Background," Pattern Recognition, vol. 27, pp. 53-63, 1994.

[58] G. Burel and D. Carel, "Detection and Localization of Faces on Digital Images," Pattern Recognition Letters, vol. 15, pp. 963-967, 1994.

[59] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation Invariant Neural Network-based Face Detection," in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Sanata Barbara, CA, 1998.

[60] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Application to Face Detection," in: The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR 1997.

[61] J. Yang and A. Waibel, "A Real-time Face Tracker," in: 3rd IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 1996.

[62] Binary Image Formation and Analysis, accessed on 5/17/2006, http://www.cs.rpi.edu/~stewart/comp_vision/classes/class2/.

[63] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature Extraction from Faces Using Deformable Templates," International Journal of Computer Vision, vol. 8, pp. 99-111, 1992.

[64] A. Pentland, B. Moghaddam, and T. Starner, "View-based and Modular Eigenspaces for Face Recognition," in: Computer Vision and Pattern Recognition, IEEE Computer Society Press, Loas Alamitos, CA 1994.

[65] Y. H. Kwon and N. da Victoria Lobo, "Age Classification from Facial Images," in: Computer Vision and Pattern Recognition, IEEE Computer Society Press, Loas Alamitos, CA 1994.

[66] R. Stiefelhagen, J. Yang, and A. Waibel, "A Model-based Gaze Tracking System," in: IEEE International Joint Symposia on Intelligence and Systems:

Image, Speech and Natural Language Systems, IEEE Computer Society Press, Loas Alamitos, CA 1996.

[67] T. S. Jebara and A. Pentland, "Parameterized Structure from Motion for 3D Adaptive Feedback Tracking of Faces," in: The IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR 1997.

[68] J. Wang, E. Sung, and R. Venkateswarlu, "Estimating the Eye Gaze from One Eye," in: Computer Vision and Image Understanding, Academic Press, Orlando, 2005.

[69] P. V. C. Hough, "Method and Means for Recognizing Complex Patterns," U.S. Patent Office, Ed. USA, 1962.

[70] Listing's Law, accessed on 3/18/2005, http://www.bme.jhu.edu/labs/chb/glossary/listing.html, Domonik Straumann.

[71] H. von Helmholtz, "Handbuch der Physiologischen Optik," Leipzig: Leopold Voss, vol. 1, 1867.

[72] D. J. Fischer, "Gradient-index Ophthalmic Lens Design and Polymer Material Studies," in: The Institute of Optics. Rochester, University of Rochester, 2002, pp. 1-272, http://www.shoutingman.com/DigitalResume/Chapter-0.pdf.

[73] J. L. Davis, J. Ayers, and A. Rudolph, "Neurotechnology for Biomimetic Robots," The MIT Press, Cambridge, MA, 2002.

[74] Visual Sensors Using Eye Movements, accessed on 3/18/2005, http://www.klab.caltech.edu/Papers/421.pdf, O. Landolt.

[75] O. Bolina and L. H. A. Monteiro, "A Note on Eye Movement," arXiv:physics/9811031, vol. 1, pp. 1-8, 1998, http://arxiv.org/PS_cache/physics/pdf/9811/9811031.pdf.

[76] W. P. Medendorp, B. J. M Melis, C. C. A. M Gielen, and J. A. M. Van Gisbergen, "Off-centric Rotation Axes in Natural Head Movements: Implications for Vestibular Reafference and Kinematic Redundancy," Journal of Neurophysiology, vol. 79, pp. 2025-2039, 1998.

[77] G. A. Fry and W. W. Hill, "The Mechanics of Elevating the Eye," American Journal of Optometry, vol. 39, pp. 707-716, 1963.

[78] R. Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," in: IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach , FL, 1986.

[79] R. Y. Tsai, "A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," IEEE Journal of Robotics and Automation, vol. 3, pp. 323-344, 1987.

[80] P. E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in: SIGGRAPH 97, 1997.

[81] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," International Journal of Computer Vision, vol. 9, pp. 137-154, 1992.

[82] M. Pollefeys, R. Koch, and L. van Gool, "Self-Calibration and Metric Reconstruction in Spite of Varying Unknown Internal Camera Parameters," in: 6th International Conference on Computer Vision, Bombay, 1998.

[83] Tsai Camera Calibration, accessed on 5/11/2006, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/DIAS1/, P. Dias.

[84] Comments: Description of the Calibration Parameters, accessed on 5/11/2006, init_intrinsic_param.m, Matlab.

[85] J. Heikkila and O. Silven, "A Four-step Camera Calibration Procedure with Implicit Image Correction," in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, 1997.

[86] Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations," in: Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999.

[87] J. Bouguet, "Camera Calibration Toolbox for Matlab," 2005, http://www.vision.caltech.edu/bouguetj/calib_doc/download/TOOLBOX_calib.zip.

[88] Camera Calibration Toolbox for Matlab: Description of the Calibration Parameters, accessed on 5/11/2006, http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/parameters.html, J. Bouguet.

[89] Rodrigues' Rotation Formula, accessed on 8/29/06, http://mathworld.wolfram.com/RodriguesRotationFormula.html, S. Belongie.

[90] J. J. Craig, Introduction to Robotics Mechanics and Control, 2nd ed. Addison-Wesley, Reading, MA, 1989.

[91] K. Fung, A. Chau, K. Pak, and M. Yap, "Is Eye Size Related to Orbit Size in Human Subjects?," Opthalmic and Physiological Optics, vol. 24, pp. 35, 2004.

[92] Stereo Vision: Epipolar Geometry, accessed on 5/24/2006, http://engnet.anu.edu.au/DEcourses/engn4528/Lectures/18_epipolar_geom.pdf, G. Loy.

[93] B. Ostle and R. W. Mensing, Statistics in Research, 3rd ed. The Iowa State University Press, Ames, Iowa, 1975.

[94] A Step-by-step Guide to the Use of the Intel OpenCV Library and the Microsoft DirectShow Technology, accessed on 5/25/2006, http://www.site.uottawa.ca/~laganier/tutorial/opencv+directshow/cvision.htm, R. Laganiere.

APPENDIX 1

CONSENT FORM

for participation in the study titled
"Gaze Determination Capability/Accuracy"
taking place from July, 2005 through December, 2005
Computer Science Department, Texas A&M University, College Station, TX

I have been asked to participate in a research study whose purpose is to determine the capability/accuracy of determining where a person is looking using video images of a person's face. This study is being undertaken in partial fulfillment of the requirements for a PhD in Computer Engineering at Texas A&M University. The following points comprise my understanding of the terms of my participation:

1. The purpose of this study is to determine where I am looking using video images of my face. I will look at locations on/near a computer monitor, and video images will be captured while I do so by 1 to 3 webcams mounted under the monitor. I will be asked to look at no more than 36 locations (plus repeats, if necessary, to capture good images). Each location viewing and image collection will last no more than 10 seconds, for a total viewing time of less than 10 minutes. I may take a break or discontinue my participation at any time should I become fatigued or uncomfortable.

2. I will be asked to look at a set of monitor locations by either glancing or staring at the intended location. I will signify to the person in charge of the study that I am ready, at which time my image will be saved. I will be expected to continue looking at the particular location until I am told the image collection is complete (<= 10 seconds for that location) and that I may relax and look elsewhere.

3. I am one of no more than 100 subjects whose facial image will be collected during this study. I will be asked to record the order in which I looked at the locations on a form provided by the person in charge of the study. I will also provide an estimate of my confidence that I was actually looking at the recorded location during the collection of the images.

4. After the collection of images, I may leave the image collection area and my active participation is complete. The images of my face will be processed at some later point. I understand that only the researchers involved in this study will be allowed to view these images, and then, only to process the images and determine the image pixel locations of various facial features.

5. The facial features identified and located will include the my pupil centers, as well as the location of 5 adhesive dots which I will place on my face in the specified locations (on my forehead below my hairline, on the bridge of my nose between my eyes, on the tip of my nose, and just below my lower eyelid above each cheek) prior to the experiment starting. If I desire, I may ask the person in charge of the study to assist me in affixing the dots correctly. I understand that these dots will remain on my face for up to 30 minutes. I will be asked to remove the dots before leaving the study area, and provide a written assessment of whether I experienced any irritation or discomfort during the study, from either looking at the monitor or affixing/wearing the adhesive dots.

6. My participation in the study will be monitored by the individual in charge of the study. This person will not be my EPSY 435 instructor. However, my EPSY 435 instructor will be notified of my participation in the study so that I receive 3 points of credit for participating. I understand that all participants will receive the same amount of credit for participating.

7. In order to participate in the study, I must register in advance and must at least complete a review of this Consent Form during the time for which I registered in order to receive credit.

8. Any information collected for this study that identifies me as an individual, other than this form, will be destroyed upon completion of the study. In the interim, all information will be kept secure and confidential, and will only be used for the study and to provide my EPSY 435 instructor with evidence of my participation. This form will be kept in a private location for at least 3 years after the study has ended. At the completion of this time, this form will be shredded. I have a right to obtain a copy of this form at any time during the 3 year period.

9. Participation in the study is voluntary. If I do not wish to participate in the study, I will not be penalized in any overt or covert way. However, if I do not participate, I will not be eligible for the course credit associated with this study. The offer of any alternate credit option is at the sole discretion of my EPSY 435 instructor.

This research study has been reviewed by the Institutional Review Board - Human Subjects in Research, Texas A&M University. For research-related problems or questions regarding subjects' rights, I can contact the Institutional Review Board through Ms. Angelia M. Raines, Director of Research Compliance at (979) 458-4067 (araines@vprmail.tamu.edu).

I have read and understood the explanation provided to me. I have had all my questions answered to my satisfaction. I voluntarily agree to participate in this study and to answer all questions truthfully and to the best of my ability. I will be given a copy of this consent form upon request.

____

Subject **Date**

____

Principal Investigator **Date**

Points of contact:
Jeffery L. Beckmann, Texas A&M University, College Station, TX 77843-3112, (979) 862-6910
Richard A. Volz, Texas A&M University, College Station, TX 77843-3112, (979) 845-8873

APPENDIX 2

Subject #: __________ Date:

Gaze Determination Capability/Accuracy Subject Data

Stare Seq. #	**Location Viewed**	**Stability During Capture**
1		___ Good ___ Unsure ___ Bad
2		___ Good ___ Unsure ___ Bad
3		___ Good ___ Unsure ___ Bad
4		___ Good ___ Unsure ___ Bad
5		___ Good ___ Unsure ___ Bad
6		___ Good ___ Unsure ___ Bad
7		___ Good ___ Unsure ___ Bad
8		___ Good ___ Unsure ___ Bad
9		___ Good ___ Unsure ___ Bad
10		___ Good ___ Unsure ___ Bad
11		___ Good ___ Unsure ___ Bad
12		___ Good ___ Unsure ___ Bad
13		___ Good ___ Unsure ___ Bad
14		___ Good ___ Unsure ___ Bad
15		___ Good ___ Unsure ___ Bad
16		___ Good ___ Unsure ___ Bad
17		___ Good ___ Unsure ___ Bad
18		___ Good ___ Unsure ___ Bad
19		___ Good ___ Unsure ___ Bad
20		___ Good ___ Unsure ___ Bad
21		___ Good ___ Unsure ___ Bad
22		___ Good ___ Unsure ___ Bad
23		___ Good ___ Unsure ___ Bad
24		___ Good ___ Unsure ___ Bad
25		___ Good ___ Unsure ___ Bad
26		___ Good ___ Unsure ___ Bad
27		___ Good ___ Unsure ___ Bad

Eye Dominance: ____ Left ___ Right ___ None

Overall Discomfort Level: ___ None ___ Minor ___ Significant

(over)

Comments:

VITA

Jeffery Linn Beckmann
9301 Amberwood Court - College Station, TX 77845
jbeckmann@tamu.edu

EDUCATION

Texas A&M University, College Station, TX

Ph.D. in Computer Engineering 2007

Dissertation: "Single Camera 3D Gaze Determination"

M.S. in Safety Engineering 1985

Thesis: "Fault-tree Construction and Calculations on a Microcomputer"

B.S. in Chemical Engineering 1983

TEACHING EXPERIENCE

Graduate Assistant Lecturer - Texas A&M University, College Station, TX 1997 - 2003

Instructor of record for ENGR111, ENGR112, CPSC203, and BANA207

PROFESSIONAL EXPERIENCE

Mary Kay O'Connor Process Safety Center, College Station, TX	2001 - Present
Department of Computer Science, College Station, TX	1997 - 2003
Raytheon, Inc., Kwajalein, Republic of the Marshall Islands	1995 - 1997
Lockheed Martin Missiles & Space Company, Inc.	1985 - 1995