STARS

Electronic Theses and Dissertations, 2004-2019

2008

Meshfree Approximation Methods For Free-form Optical Surfaces With Applications To Head-worn Displays

Ozan Cakmakci
University of Central Florida

STARS Citation

Showcase of Text, Archives, Research & Scholarship

MESHFREE APPROXIMATION METHODS FOR FREE-FORM OPTICAL SURFACES WITH APPLICATIONS TO HEAD-WORN DISPLAYS

by

OZAN ÇAKMAKÇI
B.S. Michigan State University, 1999
M.S. University of Central Florida, 2004

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in CREOL, The College of Optics and Photonics
at the University of Central Florida
Orlando, Florida

Fall Term
2008

Major Professor: Jannick P. Rolland

ABSTRACT

Compact and lightweight optical designs achieving acceptable image quality, field of view, eye clearance, eyebox size, operating across the visible spectrum, are the key to the success of next generation head-worn displays. The first part of this thesis reports on the design, fabrication, and analysis of off-axis magnifier designs. The first design is catadioptric and consists of two elements. The lens utilizes a diffractive optical element and the mirror has a free-form surface described with an *x-y* polynomial. A comparison of color correction between doublets and single layer diffractive optical elements in an eyepiece as a function of eye clearance is provided to justify the use of a diffractive optical element. The dual-element design has an 8 mm diameter eyebox, 15 mm eye clearance, 20 degree diagonal full field, and is designed to operate across the visible spectrum between 450-650 nm. 20% MTF at the Nyquist frequency with less than 3% distortion has been achieved in the dual-element head-worn display.

An ideal solution for a head-worn display would be a single free-form surface mirror design. A single surface mirror does not have dispersion; therefore, color correction is not required. A single surface mirror can be made see-through by machining the appropriate surface shape on the opposite side to form a zero power shell. The second design consists of a single off-axis free-form mirror described with an x-y polynomial, which achieves a 3 mm diameter exit pupil, 15 mm eye relief, and a 24 degree diagonal full field of view. The second design achieves 10% MTF at the Nyquist frequency set by the pixel spacing of the VGA microdisplay with less than 3% distortion. Both designs have been fabricated using diamond turning techniques.

Finally, this thesis addresses the question of "what is the optimal surface shape for a single mirror constrained in an off-axis magnifier configuration with multiple fields?" Typical optical

surfaces implemented in raytrace codes today are functions mapping two dimensional vectors to real numbers. The majority of optical designs to-date have relied on conic sections and polynomials as the functions of choice. The choice of conic sections is justified since conic sections are stigmatic surfaces under certain imaging geometries. The choice of polynomials from the point of view of surface description can be challenged. A polynomial surface description may link a designer's understanding of the wavefront aberrations and the surface description. The limitations of using multivariate polynomials are described by a theorem due to Mairhuber and Curtis from approximation theory. This thesis proposes and applies radial basis functions to represent free-form optical surfaces as an alternative to multivariate polynomials. We compare the polynomial descriptions to radial basis functions using the MTF criteria. The benefits of using radial basis functions for surface description are summarized in the context of specific head-worn displays. The benefits include, for example, the performance increase measured by the MTF, or the ability to increase the field of view or pupil size. Even though Zernike polynomials are a complete and orthogonal set of basis over the unit circle and they can be orthogonalized for rectangular or hexagonal pupils using Gram-Schmidt, taking practical considerations into account, such as optimization time and the maximum number of variables available in current raytrace codes, for the specific case of the single off-axis magnifier with a 3 mm pupil, 15 mm eye relief, 24 degree diagonal full field of view, we found the Gaussian radial basis functions to yield a 20% gain in the average MTF at 17 field points compared to a Zernike (using 66 terms) and an *x-y* polynomial up to and including 10th order. The linear combination of radial basis function representation is not limited to circular apertures. Visualization tools such as field map plots provided by nodal aberration theory have been applied during the analysis of the off-axis systems discussed in this thesis. Full-field displays are

iii

used to establish node locations within the field of view for the dual-element head-worn display. The judicious separation of the nodes along the x-direction in the field of view results in well-behaved MTF plots. This is in contrast to an expectation of achieving better performance through restoring symmetry via collapsing the nodes to yield field-quadratic astigmatism.

To family: Sheri, Yaman, Inci and Cevdet.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

The amount of information we deal with in our personal living is increasing. We remember passwords for our email accounts, bank accounts, several various digit numbers to lock/unlock keys around the home or at work, phone numbers, shopping lists, to-do lists, and the list goes on. However, there is a finite limit to our immediate memory capacity. George Miller, in a seminal paper [123], estimated the working capacity of immediate memory to be seven "chunks" with a variance of plus or minus two by applying the concept of channel capacity from information theory. Computing devices provide one way of extending our limited capacity. The emergence of several trends such as the increased availability of wireless networks, miniaturization of electronics and sensing technologies, and novel input/output devices is giving rise to user interfaces suitable for use "on-the-go" during our daily living.

In parallel to these trends, cell phones are becoming the center of communication and personal information management needs. For example, in Japan there are ~90 million cellular telephone subscribers as of July 2005; 86% of these mobile phones are configured for internet access [118]. One potential implication of the convergence of these trends is future user interfaces that are constantly available to the users. At the time of this writing, cell phone displays are not fully equipped to support such visual interfaces that are constantly available to the users. Furthermore, it is awkward to hold a cell phone as the primary choice of a mobile augmented reality display and walk around while looking at the computer generated imagery superimposed on the environment.

1.1 Motivation

The idea of a head-worn display that would overlay computer generated imagery on the real-world has existed since the 1960s [187]. We believe that it is the physical obtrusiveness of the conventional designs that made them socially unacceptable for both horizontal and vertical markets. Eyeglass based displays are potentially a socially acceptable candidate for providing the visual hardware platform for future visual user interfaces. Microdisplay technology has matured to a level where high resolution (800x600 pixels) displays with a .25" diagonal are commercially available. Optical system designs that magnify these small microdisplays within the most compact volumes remain a challenge in delivering head-worn display solutions that will be widely adopted across several applications. The research in this thesis is motivated by the optical design of compact and lightweight eyeglass displays to support future mobile visual displays.

1.2 Research Summary

Free-form optical surfaces are being proposed for many optical systems as a means of creating systems with larger fields of view (> 20 degrees) having fewer optical surfaces. Personal displays packaged as eyewear is a technology that could be enabled by free-form surfaces. In this thesis, a new class of free-form optical surfaces defined using radial basis functions is introduced. It is demonstrated that this class of optical surfaces can be optimized to create a personal display with excellent MTF across up to a 25 degree full diagonal field of view using two optical elements. This advance in displays developed using conventional optical surfaces provides a viable path to a functional personal display packaged as conventional looking eyeglasses.

1.3 Dissertation Outline

Chapter 2 reviews the state-of-the-art in head-worn display design and development [33, 158]. Head-worn display design is inherently an interdisciplinary subject fusing optical engineering, optical materials, optical coatings, electronics, manufacturing techniques, user interface design, computer science, human perception, and physiology for assessing these displays. The review in chapter 2 is focused on optical engineering aspects, divided into different sections to explore principles and applications. Building on the guiding fundamentals of optical design and engineering, the principles section includes a summary of microdisplay or laser sources, the Lagrange invariant for understanding the trade-offs in optical design of HWDs, modes of image presentation (i.e., monocular, binocular, and stereo) and operational modes such as optical and video see-through. A brief summary of the human visual system pertinent to the design of HWDs is provided. Two optical design forms, namely, pupil forming and non-pupil forming are discussed. Chapter 2 summarizes the results from previous design work using aspheric, diffractive, or holographic elements to achieve compact and lightweight systems. The applications section of Chapter 2 is organized in terms of field of view requirements and presents a reasonable collection of past designs.

Chapter 3 summarizes the fundamentals of optical design techniques pertinent to the design of head-worn displays, which include understanding designing with diffractive optical elements, aspheric surfaces, nodal aberration theory, and non-linear optimization techniques.

Chapter 4 will apply the techniques discussed in Chapter 3 to design, analyze, and fabricate head-worn displays tending towards the eyeglass form factor. The eyeglass form factor restricts the geometry and the number of optical elements within a volume on the order of 20 to

$30\ cm^3$. This volume is estimated from a laser scanned CAD model of a human head and close fitting sunglasses. The whole display assembly can be broken down into sub-assemblies and certain sub-assemblies can be placed elsewhere on the human body. As an example, fiber bundles have been used to decouple the drive electronics and the image sources from the optical sub-assembly coupling to the human visual system [176]. However, the optical sub-assembly coupling to the human eye must reside approximately within this volume. The demanding performance and volume requirements are addressed in Chapter 4 by exploiting the degrees of freedom in free-form surfaces, such as the local anamorphisms. Chapter 4 presents the design, fabrication, and analysis of single and dual element magnifiers. All designs in Chapter 4 use one free-form surface described with x-y polynomials.

Chapter 5 questions the appropriateness of multivariate polynomials as being the choice of basis in representing free-form optical surfaces. The limitations of multivariate polynomials are described by the Mairhuber-Curtis theorem. In chapter 5, we propose the radial basis functions as a potential candidate for describing optical free-form surfaces. Radial basis functions are used to describe the surfaces of mirrors in single and dual-element magnifier configurations. The weights of a linear combination of radial basis functions are optimized under the specific constraints in the single and dual element magnifier systems. Radial basis functions will be compared to multivariate polynomials such as an anamorphic asphere, Zernike polynomial, and an *x-y* polynomial. Benefits of using radial basis functions to describe free-form optical surfaces will be discussed.

Emerging technologies in the design of head-worn displays include displays capable of occlusions [98], multi-focal head-worn displays [161, 167, 208], and eye tracking in head-worn

displays [202]. Chapter 6 focuses on preliminary work related to the design of an occlusion display.

CHAPTER 2 REVIEW OF HEAD-WORN DISPLAYS: OPTICS AND THE HUMAN VISUAL SYSTEM

The emergence of several trends, such as the increased availability of wireless networks, miniaturization of electronics and sensing technologies, and novel input and output devices is giving rise to user interfaces suitable for use in our daily living [9]. A range of displays, both on-body and in the environment, are being developed to provide visual output for the users. See [16] for a recent review of displays in the environment and [143] for an interesting instance of a projector based display. This chapter will review approaches to the design of visual output devices, in particular, displays worn on the body, to support mobile users. Section 2.2 reviews the principles of optical design for HWDs. HWDs are coupled with the human eye, thus, in Section 2.3 we provide a brief summary on the human visual system parameters of interest. Section 2.4 will review the basics of pupil-forming versus non-pupil-forming designs. Section 2.5 will review specific design techniques such as the use of aspheric surfaces, diffractive optical elements, holographic optical elements, and free-form surfaces in HWDs. Section 2.6, 2.7 and 2.8 are organized by field-of-view (FOV) requirements. Low-FOV designs (<40 degrees), suitable for integration into an eyeglasses form factor are reviewed in Section 2.6. The mid-FOV (between 40 and 60 degrees) and the wide-FOV (>60 degrees) designs are reviewed in Section 2.7 and Section 2.8, respectively. Eyeglasses based displays are particularly interesting because they are well suited for mobile applications and they are more likely to enjoy higher social acceptance due to aesthetics compared to bulkier displays.

2.1 Principles of Optical Design for HWDs

Ideally, head-worn displays are designed for each specific application. Requirements of each application will guide the design process. Example specifications such as the usage of the display indoors, outdoors, or both, will determine the luminance requirement on the microdisplay. Luminance is defined as the flux per unit area per unit solid angle. Microdisplay spectrum combined with the spectral response of the human eye at the ambient illumination level of interest (e.g., scotopic or photopic) determines the spectral band of operation, which can be made narrower based on the application (i.e., a monochrome display). For example, applications such as displaying thermal camera imagery may not require color displays. In order to aid in the selection of appropriate parameters for an application, there are tables available, organized by FOV, resolution, color depth, head mobility, tracking and stereo/accommodation requirements [158].

Field of view is an important design specification, for example, a compact text-based display used in reading electronic mails, could be monocular and may require less FOV whereas an immersive training application could require a stereo display with a wider field of view. It is important to set the FOV based on the task and informed by the functioning of the visual pathways. However, independent of the target application all head-worn displays must meet some minimum requirements in terms of eye clearance, eye box size, image quality, and spectral characteristics. These minimum requirements are discussed in Section 2.3.

Depending on the luminance requirements of the application, each design category can employ a laser, a laser-diode, or a light-emitting diode (LED)-based source in combination with a microdisplay, an organic LED (OLED) or other self-emissive display, and historically cathode

ray tubes (CRTs). Average outdoors scene luminance is about 2000fL. The outdoors scene luminance can be as high as 12000fL [204]. Currently, active-matrix liquid-crystal displays (AM-LCDs), OLEDs, and liquid-crystal-on-silicon (LCOS) technologies constitute popular choices for HWD microdisplays. Liquid crystals (for example, AM-LCD or LCOS) can be transmissive or reflective [221]. OLEDs are self-emissive.

OLEDs suffer not only from a shorter life span but also from non-uniform degradation of luminance for the various colors over their lifespan. The choice of OLED as the microdisplay can lead to compact HWDs compared to LCOS panels. LCOS panels typically require illumination optics. The trade-off in using illumination optics is brightness versus compactness. OLED can be considered relatively dim compared to optimized illumination for LCOS. Compact LCOS illumination systems are an active research and optical engineering area. LED-based illumination schemes constitute one of the directions in compact LCOS illumination system development. The promise of LEDs in LCOS illumination is of a long lifespan, small physical size, and low operating voltages. LEDs can operate in sequential mode emitting red, green, and blue light at a specific time instance synchronized with the display. The LED luminance is improving, and LEDs are becoming suitable for low light output systems (i.e., on the order of 50 lm/W per LED). System level considerations, such as optical filters, integrators, polarization components, and projection lenses are discussed in [186], and some of these considerations apply to HWD design as well. Ferrin has compiled a more complete list of microdisplays suitable for HWD applications [61]. An in-depth discussion of microdisplay technology that covers the structure and performance of microdisplays can be found in [5].

Laser sources have been employed in scanning displays. Typically, laser-based sources have been brighter compared to their microdisplay counterparts. The perceived brightness is a function of both the laser dwell time (i.e., time allocated for the laser beam at each pixel) and the source luminance. Laser-based sources can also be divided into high and low power sources. Laser-based displays based on high power laser sources have been demonstrated with >1500fL luminance for military applications, however such systems are not portable and the demonstrators have been monochromatic. Laser-based displays based on laser-diodes with luminance up to 900fL have been demonstrated for medical, aerospace, and industrial applications. An alternative to higher luminance sources is to dim the light from the scene through electrochromic or photochromic mechanisms, see [165] for an example and [127] for a review of organic and polymeric electrochromic materials.

Design trade-offs in HWDs can be understood by applying the Lagrange invariant, in relation to the FOV and the exit pupil, or eye box size that defines the limits of user eye position from which the entire display is visible. The derivation of the Lagrange invariant can be found in many classical geometrical optics textbooks, see, for example, [97]. The Lagrange invariant, axiomatically stated and applied to the pupils, can be written as follows

$$\text{LI} = n \cdot \theta \cdot y_{pupil} = n' \cdot \theta' \cdot y'_{pupil} \tag{2.1}$$

where θ represents the semi field of view at the entrance pupil, y_{pupil} is the radius of the pupil in object space, n is the refractive index in the object space, θ' is the chief ray angle at the exit pupil, y'_{pupil} is the radius of the exit pupil, and n' is the refractive index in the image space.

Using equation 2.1, for a fixed value of the Lagrange invariant, the FOV in image space is inversely proportional to the exit pupil height. A drawback of a small exit pupil is that the eye naturally moves within the eyebox of the display, vignetting and in a worse case scenario 100% vignetting may occur, yielding a blackout of the image. In Section 2.4, we will relate this property to the design of retinal scanning displays.

The light is emitted from a microdisplay, and then collected by the optics within its numerical aperture, before being redistributed onto an image plane through the imaging process. The numerical aperture, or equivalently the aperture size of the optics, sets the amount of light flux contributing to the imaging process. The conservation of light flux together with the Lagrange invariant now applied to the microdisplay and the image plane yield conservation of luminance. Therefore, according to the Lagrange invariant, the luminance of the virtual image formed through a HWD will be constant. The statement of constant luminance may be counter-intuitive; the common misconception is to reason that because the virtual image is magnified, the luminance would decrease. However, we need to consider both the cone angle and the magnified image plane area and remember that their product is constant [157].

Another fundamental trade-off in HWDs exists between the FOV and the resolution. This trade-off exists because a functionality of the HWD optics is to spread the pixels on the microdisplay across the FOV. Thus as the FOV is increased, the number of pixels per degree decreases. Approaches to overcoming this trade-off has been to use a high resolution area of interest [96], [163], resort to partial binocular overlap [120] or implement optical tiling [86].

HWDs can be monocular where a single image is presented to a single eye, biocular (see [155] for a review of biocular optics design) where the same image is presented to both eyes, or

binocular where two distinct images are presented to each eye [151]. There are optical design and perceptual issues associated with each mode.

HWDs can be designed in optical see-through, opaque, or video see-through modes. Optical see-through systems typically make use of beamsplitters to combine the computer generated objects with the real-world. In video see-through mode, video cameras capture the scene that is combined with computer-generated imagery. In this mode the user views the world through the displays, therefore does not get a direct view of the world. Rolland and Fuchs compare the optical see-through and video see-through modes in [9] and conclude that the most important issues are latency, occlusion, the fidelity of the real-world view, and user acceptance. Historically, occlusion has been technically easier to achieve with video see-through display even though optical approaches to occlusion, both transmissive [98], [190] and reflective [198], [38], are actively pursued. A compact solution to the occlusion capable HWDs remains a research challenge.

View point offset is also a critical issue in systems that capture the real-world with optical systems before presented to the eyes. Magnitude of the viewpoint offset has been found to impact the sensorimotor adaptation [17]. Orthoscopic displays that do not introduce a view-point offset have been built for a modest FOV [182]. In [182] the FOV was 26 by 19.6 degrees. Biocca and Rolland comment that it is challenging for video-based see-through HWDs to perfectly match the natural viewpoint of the user for wide FOV displays [17].

2.2 A Brief Summary of the Human Visual System Pertinent to the Design of Head-Worn Displays

In optical design of HWDs we characterize the human visual system in terms of its object space parameters and aim to deliver a high quality image in the object space of the human eye. Designing in the object space means that we do not compensate for the aberrations of the human eye or rely on the results from the encoding and processing that occurs starting at the retina and going through the optical nerve, the lateral geniculate body, the primary visual cortex (V1), and the extrastriate areas [94]. However, a basic understanding of aberration tolerances of the eye and the influence of accommodation is useful in designing HWDs.

In this section, we shall briefly summarize some of the object space parameters of primary interest for designing HWDs. The functional primary parameters of interest are the variation in pupil size under various levels of illumination, the depth of field, the smallest feature size that the human eye can resolve, e.g., the resolution of the eye, the spectral characteristics of the light absorbed by the cones in the retina, and the aberration tolerances of the human eye. Binocular properties of the visual system that are of interest include the inter-pupillary distance and the stereo acuity, which is the threshold discrimination of depth from very small disparities [137]. Finally, a summary of perceptual issues in HWDs will be provided in this Section.

The field of view of the human eye is 200 degrees by 130 degrees with a 120 degrees overlap [204]. The lens of the eye is a gradient index element with a higher index at the center. The front focal length of the eye is about 17 mm and the rear focal length is about 23 mm, the difference is due to the fact that the refractive index of the vitreous material is 1.336 [55]. Most

of the optical power is provided by the cornea. Depending on age, the lens can contribute a 10 diopter optical power change in order to focus on closer objects. There exist several optical models of the human eye, some of the more schematic but still useful models of the eye are: the reduced eye (single surface), the simplified schematic eye (three refractive surfaces), and the unaccommodated schematic eye (four refractive surfaces) [41]. The first-order parameters of interest such as the location and size of the pupil and the center of rotation are the same across these models; therefore, they are equivalent for the design of HWDs.

The pupil is the optical aperture of the human eye that can change its size through dilation and contraction of the muscles that control the iris. The diameter of the pupil changes from 2 mm under sunlight to about 8 mm in the dark. A recent study, under ambient hospital lighting conditions (the authors do not quantify these ambient conditions), was done in 300 healthy participants and the mean resting pupil size was found to be 4.11 mm [18]. The normal eye is considered to be near-diffraction limited for a 2 mm pupil. The entrance pupil location is about 3 mm from the vertex of the cornea and resides behind the cornea [6]. Ogle and Schwartz [131] quantify that the depth of field decreases in steps of 0.12diopters as the pupil size increases in steps of one mm.

The classical approach to determining the resolution of the eye has been the application of the Nyquist frequency to the anatomical data on the spacing of the cone mosaic. Curcio *et al.* studied the spatial density of cones and rods in eight human retinas obtained from 7 individuals between 27 and 44 years of age. An average of the center to center spacing of the cones reported in [50] yields 2.5 μm and can be converted to a visual angle of 0.5 arcminutes. However, we should note that there exists studies on the cone spacing and the resolution limit reporting that,

under certain conditions, observers can see fine patterns at the correct orientation when viewing interference fringes with spatial frequencies as much as about 1.5 times higher than the nominal Nyquist frequency of the underlying cone mosaic [215]. Currently, the trend in HWD design is to aim for 1 arcminute of visual acuity. We should also note that the higher visual acuity occurs within the fovea region, spanning roughly ±5 degrees around the optical axis of the eye. Finally, dynamical visual acuity [110], while challenging to measure, is of critical importance to HWD design. It is well known that the human visual system summates signals over a time period of about 120 ms. A compounding factor in quantifying the visual performance for moving factors is motion smear. Burr reports that if the target is exposed long enough the amount of smear is far less than may be expected [25].

Stray light, caused by ocular scattering, affects the contrast specification of an HWD [217]. De Wit shows that for a 24x18 degrees display, the maximum achievable contrast for a 4x4 checkerboard image is approximately 80:1. This is calculated by convolving the ideal image with the point spread function of the eye. Hasenauer and Kunick analyzed the impact of veiling glare in an HWD system and offered techniques to isolate the sources of stray light and minimize their impact in the design [77].

Spectrally, light is transmitted starting at about 400nm and up to 1400nm through the ocular media [41]. However, only a selected portion of these wavelengths are absorbed by the cones, depending on the level of illumination. Based on the Commission International de L'Eclairage (CIE) curve, we can consider the photopic (luminance levels above about 3 cd/m^2) visible spectrum to span from 400 to 700nm. The photopic response curve peaks at 555nm.

Based on the photopic response curve, optical designers choose a spectral weighting scheme and incorporate it as a meaningful factor throughout the optimization process.

Burton and Haig [26] simulated the effects of Seidel aberrations (defocus, spherical aberration, astigmatism and coma) on four images and determined the tolerances of the human visual system to different levels and combinations of the aberration types by a forced-choice discrimination task [73]. The basic idea in the experiment is to convolve targets with aberrated point spread functions having varying levels of aberration and to compare the aberrated image against the original in a controlled setting to quantify the level of just-noticeable differences between the images. Measurements of aberration tolerances were performed with 2 mm pupils. A summary of the averaged peak-to-valley aberration thresholds for 60%, 75%, and 90% discrimination probabilities, across four targets and two observers is given in Table 2.1.

In terms of chromatic aberrations of the eye, Rodgers calculated the chromatic aberrations of the eye by combining the photopic response curve with the chromatic aberration curve of the eye and found an error of $\lambda/4$ for a pupil of 2.4 mm [156]. Hence, the chromatic aberration of the eye under photopic conditions, for an average outdoor scene, is rather benign. Mouroulis *et al.* [128] reported that the 2.5 arcminutes recommendation in the military handbook 141 [55] represents a measurable drop in performance and constitutes a realistic tolerance for the lateral chromatic aberration.

Table 2.1 Aberration thresholds in the form of just-noticeable differences listed in terms of probability of discrimination.

Peak to Valley Wavefront Aberration (in units of Wavelengths)			
	60%	75%	90%
Defocus (W20)	0.14±0.013	0.22±0.016	0.3±0.019
Spherical (W40)	0.13±0.013	0.21±0.013	0.29±0.014
Astigmatism (W22)	0.20±0.017	0.30±0.019	0.41±0.023
Coma (W31)	0.30±0.02	0.46±0.031	0.62±0.045
Strehl Intensity Ratio			
Defocus	0.93±0.013	0.85±0.022	0.73±0.031
Spherical	0.94±0.012	0.86±0.017	0.77±0.02
Astigmatism	0.92±0.017	0.80±0.023	0.68±0.027
Coma	0.95±0.006	0.89±0.015	0.81±0.026

Binocular properties of the visual system are important for all modes of operation monocular, biocular, and binocular displays. Briefly, with monocular displays the visual system in many cases may respond with binocular rivalry, a perceptual conflict when the two eyes receive very different stimulation. With biocular displays, the two eyes receive the same stimulation, which would signal to the visual system that the scene contains zero disparity throughout the visual field (which would dilute the visual effect of any perspective depth cues present). See [206] for a discussion of binocular issues associated with the design of stereo displays. A review of human stereopsis is provided by Patterson and Martin [137], readers interested in depth perception in stereoscopic displays may also refer to [138]. Patterson organizes his review around five functional topics that may be important for the design of stereoscopic display systems: geometry of stereoscopic depth perception, visual persistence,

perceptual interaction among stereoscopic stimuli, neurophysiology of stereopsis, and theoretical considerations.

Important concepts in stereopsis stem from the retinal positions of images referred to as disparity. The horopter can be thought of as a baseline of zero disparity based on the retinal images, and can be defined geometrically (Vieth-Muller circle) or psychophysically [137]. In terms of the geometry of stereoscopic depth perception, Patterson notes that the processing for points inside (crossed disparity) or outside (uncrossed disparity) of the horopter is done by separate mechanisms [129]. The consequence is that depth perception quality can be different for points residing in the crossed and uncrossed disparity regions. The small area slightly in front or behind the horopter is called the *Panum's fusional area*. In terms of binocular fusion, fusion limits depend on the disparity magnitude. Fusion limits are given to be 10 arcminutes (e.g., small disparity), 2 degrees (e.g., medium disparity) and 8 degrees (e.g., high disparity). Beyond these limits, objects may be seen double (diplopic). The disparity limits of fusion are reported to covary directly with stimulus size or scale, and inversely with spatial frequency. The limit does not vary with contrast. The vergence eye movements are known to improve the range of disparities that can be processed by the human visual system. Clinical tests carried at a distance of 40 cm with a fixed 65 mm IPD assumption yield a stereoacuity between 20–40 arcseconds [41] and stereoacuity can be as low as 8 arcseconds [137]. It is known that stereoacuity varies with luminance and spatial frequency [129]. Patterson found that high stereoacuity is obtained with high spatial frequencies and low temporal frequencies [135]. High stereoacuity is achieved in the fovea region. Therefore, high quality depth discrimination over the visual field requires eye movements [137].

The optical design consequence of supporting eye movements is to have either a large exit pupil or eye tracking capability [202], [49]. The optical consequences of eye movements in visual instruments include full or partial vignetting and was analyzed in [164]. In terms of visual persistence, Patterson and Martin [137] recommend that for displays operating using the field sequential technique, longer durations of 10 ms or higher would be expected to produce better depth perception due to greater binocular integration for each channel. Other trade-offs such as color smear will need to be investigated.

According to MIL-STD-1472C and the literature on the statistics of IPD values, the IPD adjustment range for a binocular HMD is recommended to be at least 50–78 mm. U.S. Air Force anthropologists measured the IPD of 4000 flying personnel and the mean yielded 63.3 mm [81]. Self provides a review of optical tolerance limits for binocular displays [171]. Based on the rather sparse data available in this field, recommendations of Self are based on studies of Gold [70] and Gold and Hyman [71]. Both of these studies are limited to no more than four observers, hardly representative of a population as noted by Self. Self recommended vertical misalignment of the two optical axes, horizontal misalignment of the two optical axes, rotational difference between the optical axes to not exceed 3.4 arcminutes. Self recommended that the tolerance for the magnification difference between the two images at the edge of the overlap area to not exceed 10 arcminutes with a less than 3.4 arcminutes preferred value. Self recommended a luminance difference tolerance between the two images to be less than 25% and preferably less than 10%. Self recommended the collimation tolerance shall be such that the optical distance to the displayed image shall be between 100m to ∞.

Patterson, Winterbottom and Piece provided a comprehensive review of perceptual issues in HMDs [139] which we shall summarize in this paragraph. Patterson organizes his review in five sections: luminance and contrast and their effect on depth of field, dark focus and dark vergence, accommodation-vergence synergy and its effect on perceptual constancy (e.g., size and speed constancy), eye strain and discomfort, FOV and its relationship to the functioning of different visual pathways, binocular input and its relationship to visual suppression and head movements.

In terms of luminance and contrast, as the luminance is increased, the pupil size reduces and the depth of field increases. The competing trend is that as the target resolution increases, the depth of field decreases. Therefore, both the luminance level and the target resolution should be considered when designing for a specific depth of field value. The depth of field is a critical distance that determines objects that appear in sharp focus. The depth of field should be set appropriately so that the user can properly perceive both the computer generated imagery and the real objects lying at the same depth. Under low levels of illumination or degraded stimulus conditions, the accommodation will rest around a 1 m value with some variation [80].

The human eye has evolved in such a way as to converge and accommodate at the same point of interest as it saccades across the FOV. Monocular systems can present the magnified virtual image of the microdisplay at a fixed distance from the user and the information appears to reside on this single plane. The consequence is that the users can accommodate and converge on the plane of the virtual image, which is consistent with the accommodation and convergence mechanism of the human eye. Stereo displays demand that the users focus on the plane of the virtual images formed by the optics and converge potentially at different depths away from that

plane in order to perceive three-dimensions. As a guideline, the human eye requires that accommodation and convergence to match to within ±0.1 of a diopter [226]. Accommodation and convergence conflict is known to result in eye strain and discomfort. Moffitt reports a ±0.25diopter to be workable and a ±0.5diopter as the maximum level of mismatch [125]. If the accommodation and convergence planes are further apart than the workable mismatch, the eyes would converge to the correct location to avoid diplopia and suffer from excessive blur as a trade-off. Multi-focal displays have been proposed to resolve the accommodation-convergence conflict [108, 161]. When the accommodative response becomes a compromise between the stimulus and the dark focus value, size, depth, and speed may be misperceived. Vergence seems to be valid down to 0.02fL [109] and the accommodation seems to be valid down to between 2–100fL [93]. In terms of the accommodation-vergence conflict, Patterson concludes that more research is necessary to determine the exact tolerance limits.

Multi-focal planes could be based on non-moving parts. The idea in multifocal planes is to construct a miniature 3-D display, image the 3-D source into a volume perceived as continuous dictated by the limits of human visual acuity. The natural question to ask is the minimum number of planes one would need to cover the whole visual space. Rolland *et al.* has shown that 14 planes can cover a focusing distance from 0.5 m to infinity, under the assumptions of a 4 mm pupil and a 1 arcminute visual acuity [161]. When considering moving parts, Wann proposed tracking the eye and coupling eye movements to adjust the image plane depth via a servo-lens mechanism [206]. Alignment of the optics for each eye is a critical issue in stereo displays.

In terms of field of view, the human visual system seems to process two cortical streams [109]: Dorsal stream, which connects the central and peripheral retinal areas to the magnocellular pathways, and seems to be responsible for optical flow processing as well as visuomotor abilities such as reaching. The dorsal stream processes low spatial frequency and high temporal frequency information. The ventral stream connects mainly the central retina to the parvocellular pathways that are responsible for high spatial frequency and low temporal frequency information processing as well as color. For tasks such as targeting and object recognition, processed through the ventral stream, a full FOV of 50 degrees can be sufficient. For tasks requiring peripheral stimulation, Patterson recommends greater than 60-degrees. The perceptual trade-off in wider field of view is between the level of immersion and nausea.

References [140], [139] provide a discussion of visual issues in the use of a head-mounted monocular display. The monocular, partially occluded mode of operation interrupts binocular vision. Moreover, the presentation of disparate images to each eye results in binocular rivalry. Peli recommends using a peripheral positioning in order to maintain normal binocular vision of the environment. Peli also notes that under conditions of rivalry, the brighter field will dominate. Display resolution and contrast are reported to be of secondary importance for rivalry. In a recent study Sheedy and Bergstrom [174] performed a text-based task comparison of five displays: monocular virtual, binocular head-mounted virtual, hard copy, flat panel, and a small format portable display. Performance speed on a monocular HWD was found to be comparable with performances on flat panel and hard copy. Interestingly the authors found that the performance speeds on the binocular HWD were about 5% slower than normalized performances, 6.75% slower compared with the traditional flat panel and hard copy displays. The

authors also report that the symptoms of eyestrain and blurry vision were significantly higher on monocular HWDs than on other displays.

Finally, a key perception issue of binocular HWDs is the quantification of depth perception. While there is a large body of literature relating to depth perception, for example Cormack [45], Fox and Patterson [136], Becker *et al.* [13], some of the earlier works that studied depth perception in see-through HWDs were conducted by Ellis [57] and Rolland [159]. In his work, Ellis quantified the accuracy and precision of perceived depth for real and virtual objects, where the real objects included some surface in front or behind the virtual objects. Results showed that under the observation conditions some bias in depth perception was measured. Rolland in a first study investigated the accuracy and precision of perceived depth for two real, two virtual, and one real and one virtual object side by side. No bias was measured for the two first cases in spite of having the 3-D objects displayed at either 0.8 m or 1.2 m while the optical images were collimated. For the real/virtual condition a bias of about 30 mm and 50 mm for 0.8 and 1.2 m was reported. The study pointed to the difficulty in presenting virtual objects at targeted precise depths and the need for optimized methodology. In recent work, Rolland revisited aspects of the methodology to quantify depth perception in HWDs and concluded that the method of adjustments with two stimuli of different shapes was most appropriate [162].

.

2.3 Pupil Forming versus Non-Pupil Forming Designs

It is possible to categorize optical designs for HWDs into pupil forming and non-pupil forming. Pupil-forming designs include eyepieces that magnify an intermediary image while imaging the pupil of the optical system forming the exit pupil, projection optics, and retinal scanning displays which can optically be considered as scanning projection systems. Non-pupil forming designs include magnifiers. The optical systems designed for HWD applications fall into two broad design form categories: magnifier/eyepiece based designs and projection based designs. Once a microdisplay suitable for the task is chosen, the role of the optical system is to relay a magnified virtual image of the microdisplay to the human visual system. The gain in apparent size is typically quantified by comparing the image seen through the lens to the image seen with the unaided or naked eye. For the unaided human eye, the apparent size of a microdisplay can be increased by bringing the microdisplay closer to the eye up to the accommodation limit. Compact and lightweight optical systems are necessary in many applications, especially when the users are mobile. Compactness of the microdisplay necessitates a magnification requirement since the microdisplay typically is too small to view with the unaided eye.

Regardless of complexity (i.e., types or number of surfaces), an optical system can always be characterized by an effective focal length along with principal planes and nodal points. A *magnifier* forms a virtual image when the object lies inside of its focal length, as shown in Fig. 2.1. The distance x is the object distance and it is measured from the principal plane P. The distance x' is the image distance and it is measured from the principal plane P'. We assume the object and image spaces to have the same index of refraction (i.e., air in our case); therefore, the

nodal points (*N* and *N'*) are at the same points as the principal planes. The formation of the virtual image is illustrated by tracing an off-axis ray through the nodal point and a ray parallel to the optical axis of the system, the intersection point of these rays determine the location and size of the virtual image. The virtual image formed by a magnifier has the same orientation as the object. In HWDs, the virtual image can be placed within about 1 m for near-field tasks or beyond 6 m for far-field tasks. Optical distances can also be measured in diopters, which is a reciprocal of the optical distance expressed in meters.

The pupil of the observer becomes the limiting aperture and the exit pupil in a magnifier. Magnifiers due to their large eyebox, accommodate a range of eye movements such as translations and rotations while observing the virtual image formed by the magnifier. The eye motion requirement combined with a desire to achieve good image quality motivates designs more complicated than a single lens. However, we shall explore the limits of what can be achieved with a single mirror used as a magnifier in Chapter 4. In terms of image quality, we desire that the points in the object map to points in the image, as limited by the size of the pixel or the resolution of the eye, planes map to planes, and the desired magnification remains constant across the image. Deviations from these desires are called optical *aberrations* and it is the task of the optical designer to minimize them.

As an example of noncollimated single lens magnifier, Coulman and Petrie [46] discuss the design of a binocular magnifier with a conic surface. They compare the shapes and relative positions of images formed for each eye through a binocular magnifier with spherical surfaces and another magnifier combining a spherical and a conic surface. For each line of sight of each eye, there will be an astigmatic focus, different for the tangential and the sagittal directions. The
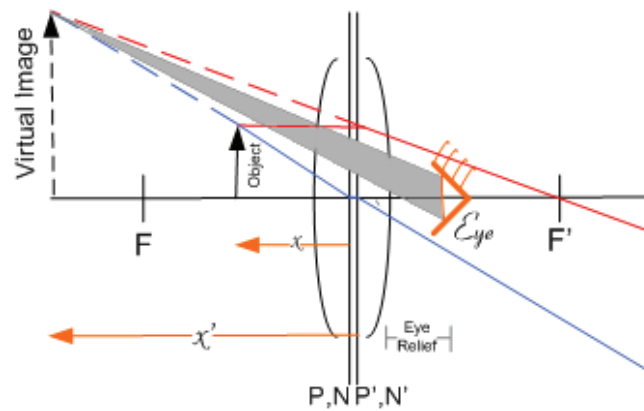
(a)
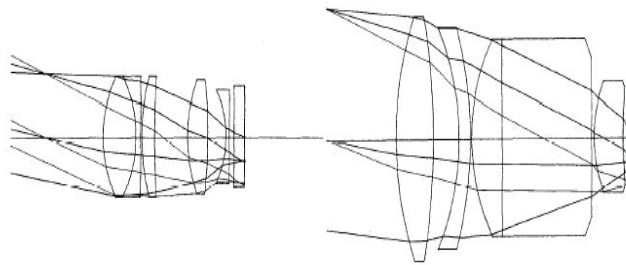
(b)

Figure 2.1 (a) Geometry of a magnifier. A magnifier presents a virtual image and the observer's eye-pupil becomes the exit pupil of the system. The distance from the vertex of the last physical surface of the system to the exit pupil is called the eye relief. (b) Example 45 mm focal length magnifiers. Monocular (left) and biocular (right). Both systems are working with a 30 mm screen diameter. The monocular system has a 40 mm 'exit pupil' and the biocular system has an 84 mm 'exit pupil', the latter being large enough to cover both eyes. (Adapted from Williamson [216])

middle point between the tangential and the sagittal surfaces can be taken as the focus surface, which is not necessarily the convergence surface. Therefore, the magnifier based on spherical surfaces creates images considerably different for each eye which requires the eyes to fuse two completely different images, especially toward the edge, leading to a disturbance of the accommodation and convergence relation. Rogers and Freeman reoptimized the design given by Couldman and Petrie to have a diameter of 79 mm and having a sixth-order aspheric surface and they reported a magnification of 2.3 [46].

An *eyepiece*, in addition to creating a virtual image for the human visual system, forms an exit pupil by imaging the pupil of the system prior to it to the image space. The distance between the edge of the last surface and the exit pupil is referred to as *eye clearance*. The distance from the vertex of the last surface to the exit pupil is called the *eye relief*. A minimum eye clearance distance is considered to be 20 mm according to Kocian [101] as well as Self [171]. Self [171] recommends the minimum exit pupil size of an HWD to be 10 mm. In the case of a telecentric design, the eye relief distance determines the focal length of an eyepiece. In a telecentric lens, as the field angle increases, the height of the lens will scale accordingly causing an increase in weight. HWDs with relaxed telecentricity conditions have been built successfully [88].

For compact and telecentric systems, the designers' goal is to optimize the design for a specific eye relief distance and have it as close as possible to the focal length. In a recent study, we compared two eyepieces: one closely resembled an Erfle eyepiece and a second one where the doublets in the Erfle-like eyepiece were replaced with diffractive optical elements. We quantified that the eyepieces having diffractive optical elements can support an eye relief range

of about 80% of their focal lengths, much higher than the doublet based system. In terms of FOV, 30 degrees is typical for simpler eyepieces while higher fields of view with considerable aberration correction have been achieved [54, 170].

The aberrations of interest in eyepiece designs are lateral color, spherical aberration of the pupil, distortion, field curvature, and astigmatism. In the case of HWDs, some amount of distortion in the optics can be accepted as it is possible to correct distortion by prewarping the images on the microdisplay in hardware or software or optically [22]. In wide angle eyepiece designs 8%–10% of distortion is considered typical [178]. If the distortion is to be corrected in electronics, it is crucial to achieve real-time correction (<10 ms delay per frame). Spherical aberration of the pupil refers to the change in the exit pupil position with field angles. Spherical aberration of the pupil is closely linked to distortion [178]. Distortion can be defined as nonlinear magnification across the field of view. The consequence of forming an exit pupil outside of the eyepiece is broken symmetry about the pupil. Broken symmetry makes aberration control challenging.

Head-worn optical projectors are an alternative optical design form to the eyepiece based designs. The basic idea in an optical head-worn projector is to place the microdisplay outside of the focal length of the projection lens in order to form a real image in front of the user. The real image formed by the projection lens can be displayed on a diffuse [102] or a retroreflecting screen [63, 134, 160]. An account of conventional projection screens can be found in [91]. A retroreflecting screen eliminates optical crosstalk between users and allows for multiple users to interact within a virtual environment. A retroreflecting screen also maximizes light throughput compared with a diffused screen. The retroreflecting screen can be placed physically or optically

27

in the environment depending on the application. Optical placement allows for using the display outdoors, eliminating the requirement to have a screen physically placed in the environment [116].

The optical path in a head-worn projector with an external retroreflective screen is folded using a beam splitter to direct it towards the screen placed in front of the user. The retroreflective screen will direct the light back, at least in principle, exactly in the same direction when the light meets the retroreflective screen. An example optical layout for the HMPD is shown in Fig. 2.2. The first miniature optics for HMPDs was developed by Rolland's group, see for the historical evolution of HMPDs [160].

Optical projectors based on a Tessar lens form have been designed. The optical layout of a Tessar lens for a head-mounted projector application is shown in Fig. 2.3. The design form shown in Fig. 2.3 evolved from a double-gauss design form as discussed in the original paper [88]. Double-Gauss designs descend from the double meniscus anastigmat lenses [178]. Double-Gauss lenses can achieve a flat field by balancing the positive and the negative surfaces. Symmetry around the pupil in the double-Gauss form helps with aberration correction. Double-Gauss lenses are limited by oblique spherical aberration caused by the anamorphism seen by the bundles at maximum field angles as they pass through the lens [113]. It is well known that systems exhibiting symmetry about the pupil will have all odd aberrations minimized, such as coma, distortion and lateral color. In an HMPD, the pupil resides within the optical system, symmetrically, and can be conjugated to the eye through the beamsplitter forming a virtual pupil [117]. Such designs allow large fields of views without any significant scalability in the optics, thus yielding wide angle compact designs.
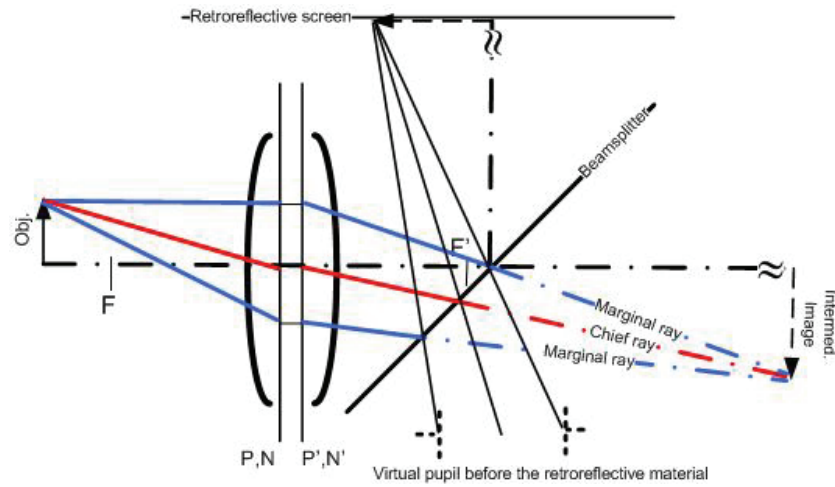
Figure 2.2 Optical layout of a head-mounted projector display.
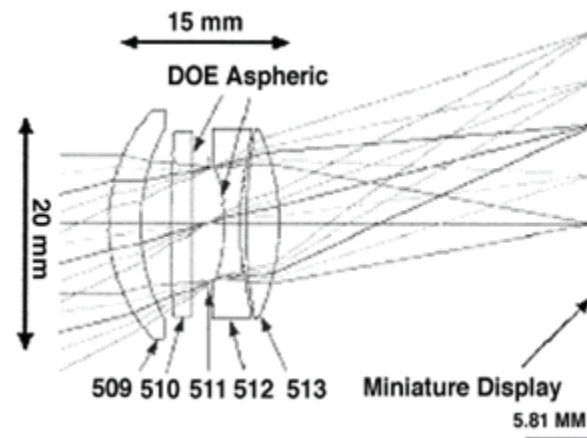
Figure 2.3 Tessar lens with a diffractive optical element (adapted from [88])

Retinal scanning displays (RSD) aim to scan a beam of light onto the viewers' retina. The early prototypes developed in the 1990's used lasers and acousto-optic modulators to implement the retinal scanning concept [66]. Different scanning techniques such as rotating polygons, galvanometers, piezoelectric deflectors, or acousto-optic deflectors, have been compared by Holmgren and Robinett [84] in terms of resolution, sweep frequency and cost. Based on the commercially available scanning techniques of the time, Holmgren and Robinett concluded that none of these scanning technologies promised a display system that would be clearly superior to current CRT and LCD displays in HMD applications. Recent work on the design of MEMS scanners shows that it is possible to design and fabricate MEMS scanners that can operate at frequencies high enough to support a resolution of 1280x1024. A comparison of thirteen state-of-the-art scanners published in the literature is provided in [222]. DeWit designed, implemented and assessed a retinal scanning display based on a scanning polygonal mirror [218]. De Wit points out the interesting similarity between the retinal scanning devices and the scanning laser ophthalmoscopes which can be regarded as the predecessor of the retinal scanning devices [209]. A comparison of state-of-the scanners is provided in [201].

At the system level, the building blocks of retinal scanning displays can be broken down into video electronics, a photonics module, a scanner module, an exit pupil expander, and a viewing optic. In the following paragraphs, we will summarize the functionality of each module and the interested reader should refer to [199], [219] for further details and references.

Based on the desired video signal input, the video electronics module controls the intensity and the mix of colors in the photonics module and generates the time-synchronized control signals for the scanner module. For each pixel, the photonics module is responsible for

delivering a light beam to the scanner module with the desired intensity and color as requested by the input video signal. In LED or laser-diode based systems, the pixel intensity can be controlled directly by varying the drive current. The alternative to direct modulation is to use an external modulator such as an acousto-optic or an electro-optic modulator. Luminance control can be implemented as a separate module if necessary, for example, in the form of polarizers, fiber-optic or electro-optic attenuators. In principle, given a fast modulation scheme, it should be possible to adjust the convergence for each pixel, therefore, imposing depth on each pixel [208].

The scanner module is responsible for directing the beam of light delivered from the photonics module to the desired image plane location. Relying on persistence of human vision, it is the fast and repeated scanning onto the retina that creates the perception of a static 2-D image. A number of scanning methods exists; for example, polygonal scanners, oscillatory scanners, acousto-optic and electro-optic scanners, and holographic scanners. There are several ways to implement scanners and biaxial MEMS scanners have been fabricated for use in commercial products by Microvision. In a biaxial MEMS scanner, the mirror is suspended with two flexures in a gimbal and is actuated via electrostatic plates under the mirror. Diffraction limited performance requires the scan mirror to be optically flat ($\lambda/10$). Pixel location and luminance variations caused by nonlinearities in the scanning can be overcome by synchronizing the output based on a look-up table that accounts for the nonlinearities.

In order to create a comfortable eye-motion box, a 10–15 mm exit pupil is preferred. However, without exit pupil expansion, a typical exit pupil size in RSD systems is reported to be 1–2 mm as imposed by the Lagrange invariant discussed in Section 2.2. Therefore, for improved

usability and to accommodate positioning error, it is essential to use an exit pupil expander unless the system makes use of an eye tracker and moves the exit pupil accordingly.

In laser-based RSDs, a Gaussian beam propagates through the scanner optics and forms a Gaussian spot, therefore, beam truncation impacts resolution and contrast [200]. For a detailed discussion on fundamental work in imaging with Gaussian pupils, see [111].

A light emitting diode (LED) based scanner has been designed by Wells and Becker [12] as shown in Table 2.2. In this system, LEDs, a magnifying lens which forms a virtual image, and a vibrating mirror are employed to scan a few rows at a time. The imaging optics is designed for adjustable focus. This display provides 720 (H) x 480 (V) pixels across a 21° x 14° FOV. The display is monochromatic, which makes the magnifying optics simpler since achromatization is not necessary.

2.4 Applications of Aspheric, Diffractive, Holographic and Freeform Surfaces in HWDs

In spite of cost, difficulties in fabrication and testing, and a deficiency of less than 100% light throughput in the case of diffractive and holographic surfaces, freeform [87, 90, 144, 145, 223], diffractive [47, 197, 211] and holographic surfaces have been used successfully in previous HWD designs [3, 7, 75, 78, 119], motivated mainly by weight constraints on HWDs.

The main use of diffractive optical elements is rooted in their ability for color correction, specifically, their approximately -3.5 Abbe number, which provides complimentary dispersion

characteristics when used with optical glasses and plastics; Stone and George [183] present a derivation of this fact. It is possible to replace the doublets used in color correction with diffractive optical elements yielding weight reduction. Missig and Morris point out in [124] that as the task of color correction is shifted to the diffractive optical element, the steep curvatures for negative color correction elements reduce leading to reduced aberrations. Details of aberrations inherent in the diffraction process can be found in [76].

The contours of diffractive optical elements can be fabricated in discrete steps using lithographic techniques, diamond turning, laser writing as well as electron-beam lithography. Therefore, the minimum feature size becomes an important parameter in the utilization of diffractive optical elements as constrained by the fabrication methods. Sub-micron (0.1 m) feature sizes have been demonstrated [175]. Many diffractive elements are based on plastic materials for their substrate. Popular plastic materials include acrylic (PMMA), Styrene, NAS, SAN, polycarbonate, tpx, abs, and nylon. Materials choices are limited in plastics compared to glasses. Injection molding and compression molding processes make volume manufacturing feasible for plastic optics. Aspheric surfaces can be manufactured on plastic by diamond turning and rms surface roughness of less than 50 Å can be achieved.

A concrete comparison between a pure refractive solution and a hybrid refractive-diffractive design has been provided in [47] by Cox *et al.* and the systems compared are shown in Fig. 2.4. Fig. 2.4 shows two systems having equivalent optical performance. Each system is reported to have a 15 mm exit pupil, 20 degrees FOV, and a 75% diffraction limited modulation transfer function (MTF) at the Nyquist frequency of the display for the visible spectrum. The

design using diffractive optical elements has only two elements and is reported to weigh ten times less than the conventional design.
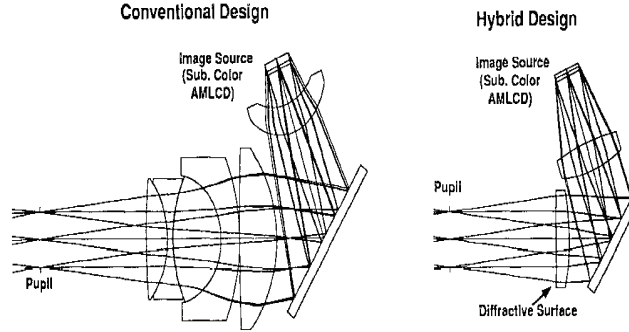
Figure 2.4 Comparison of a conventional and a hybrid design using diffractive optical elements.

A second example is given by Stone and George [183]. These authors study in detail hybrid lenses with diffractive and refractive elements and show that two and three wavelength achromats can be achieved with smaller refractive curvatures at larger attainable apertures. The authors also point out the capability of hybrid elements to achieve an Abbe number ($V = \dfrac{n_d - 1}{n_f - n_c}$, where n_d, n_f, and n_c are the refractive indices of the D (587.6nm), F (468.1nm), and C (656.3nm) lines, respectively).

A third example, utilizing diffractive–refractive optical elements, discussing the relationship between the MTF and the diffraction efficiency is provided in [58], based on a relay lens system. The authors compared a plano–convex refractive element against a plano-convex with a diffracted optical element. This study reports the integrated diffraction efficiency of the hybrid element to be 91% and concludes that the effect of the wavelength dependent diffraction

has little impact on the imaging performance. The authors use the modulation transfer function as their criteria for image quality comparison between the hybrid and the pure refractive element.

A fourth example is the design by Missig and Morris [124], where the authors designed hybrid elements with >60 degrees FOV and reported a 70% weight reduction compared with the Erfle eyepiece as well as a 50% decrease in the pupil spherical aberration and a 25% reduction in distortion.

A fifth example is provided by Knapp *et al.* [100] where the authors compared an eyepiece using a diffractive optical element to an eyepiece using aspheric elements. The system in [100] uses spherical glass elements and a diffractive surface compared with a system that uses spherical glass elements and a plastic aspheric element. The development criteria for the lenses were 60 degrees full FOV, 24 mm focal length, 10% distortion, 12 mm exit pupil and a 20 mm eye relief over the 540–558nm spectral band. Aberration tolerances were ±2 D on the focus range, ±0.25 D on the field curvature, <0.5 D for astigmatism. There were additional packaging constraints as well. They draw the conclusion in their study that the diffractive optical element yields a better optical design from a chromatic, field curvature and MTF standpoint for their set of specific application requirements.

A sixth example of hybrid refractive-diffractive solution is provided by Wood [220], where the author replaces conventional doublets with a hybrid refractive-diffractive element, operating within the 3-5μm band. Wood confirms the reduction in mass and material cost as well as comparable level of aberration correction (i.e., half the mass and one third of the cost in the specific geometry Wood compared). Wood notes that an all-spherical DOE can correct spherical aberration using the phase terms in the DOE; however, this introduces excessive

spherochromatism. An aspheric surface can be used as the substrate of the DOE to match the spherical aberration correction of the doublet.

Due to multiple diffracted orders, power in orders other than the one a designer is using serves to reduce the contrast in the final image. Buralli and Morris define a quantity called the integrated efficiency that serves as a useful figure of merit to describe diffractive lenses [24]. Missig and Morris [124] note that numerous factors can affect the diffraction efficiency, including the surface blaze profile, the zone spacing, the surface coatings, the illumination wavelength, the incidence angle, the polarization, and the substrate index of refraction.

The use of holographic optical elements [189] is motivated by aberration correction, suitability in off-axis geometries, applicability to semi-transparent optical combining, and compact and light packaging requirements. The disadvantages of holographic elements include their low efficiency across a broad spectrum. Ando *et al.* [3] experimented with a holographic combiner that can transmit the rays from the scene and reflect a narrow band of specific wavelengths. The authors used a single HOE of 4"x5" size and recorded two beams incident at 30 degrees, one for each eye, in order to generate the necessary parallax per eye from a single HOE. A He–Ne laser (632.8nm) was used in the recording process and a semiconductor laser (635nm) was used in the reconstruction process. Finally, the use of volume hologram written in the UV part of the spectrum but replayed in the visible part of the spectrum were explored towards use in HWDs [75]. The advantage of a volume hologram is higher efficiency as compared to surface holograms.

2.5 Applications of Head-Worn Displays: Low Field of View Designs (<40 Degrees)

 "Wearable computing allows graphics and text to be associated with physical objects and locations. Through overlay displays, the user can concentrate on the task instead of constantly looking down as is the case with notebook computers or pen-based systems. Techniques are available to sense what object the user is looking at, the computer may deduce the user's intention and update the display automatically." [181]

Table 2.2 Low field-of-view designs (<40 Degrees)
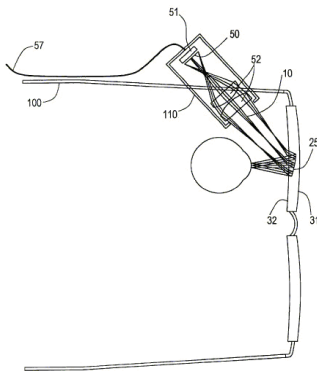
Spitzer. Eyeglass Display Lens System Employing Off-Axis Optical Design.
US 6,353,503
Mar. 5, 2002
(a)
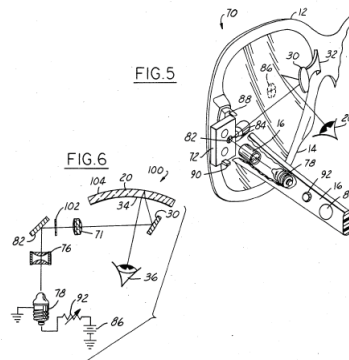
Bettinger. Spectable-mounted ocular display apparatus.
US 4,806,011
Feb. 21, 1989
(b)
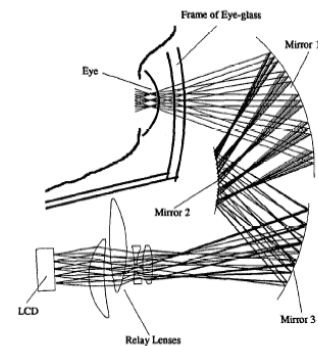
Hoshi et.al. Off-axial HMD optical system consisting of aspherical surfaces without rotational symmetry.
In Proc. of SPIE Vol. 2653
(c)
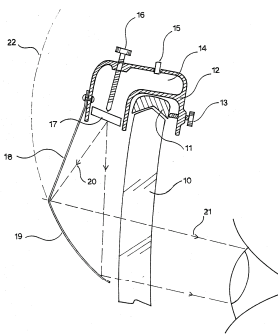
Perera. Display Projection Optical System for Spectacles or Sunglasses.
US 4,867,551
Sep. 19, 1989
(d)
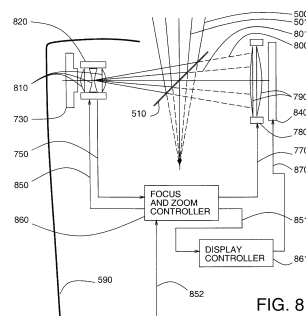
Mann. Wearable Camera System With Viewfinder Means.
US 6,307,526
Oct. 23, 2001
(e)
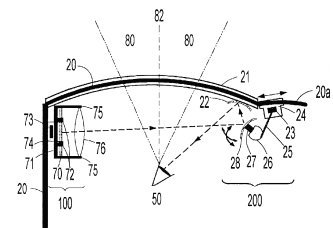
Geist. Head-mounted virtual display appa with
near-eye deflecting element in the periph field of view.
US 6,771,423
Aug. 3, 2004
(f)
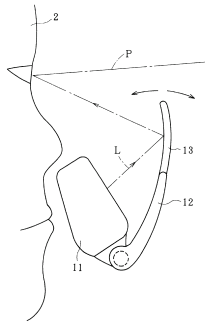
Amafuji. Head Mounted Display Device.
US 6,359,602
Mar. 19, 2002
(g)
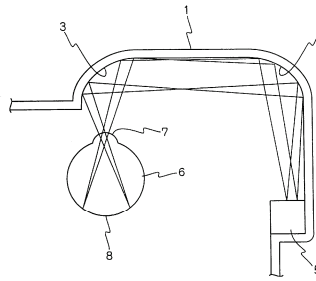
Kuriyama. Image Display Apparatus.
US 6,081,304
Jun. 27, 2000
(h)
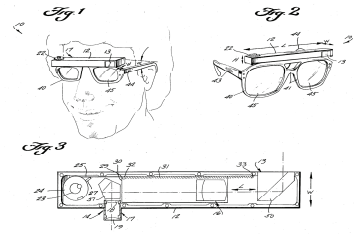
Pekar. Vision enhancing system.
US 4,704,000
Nov. 3, 1987
(i)
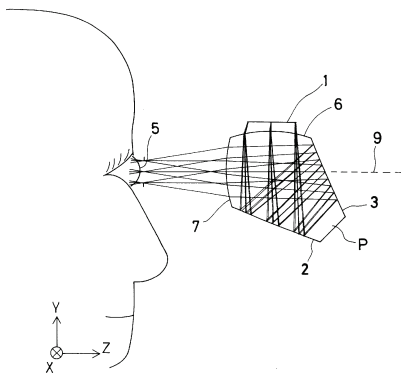
Togino. Prism Optical System.
US 5,991,103
Nov. 23, 1999
(j)
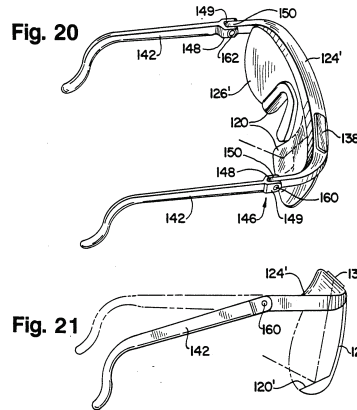
Furness. Display System for a Head Mounted
Viewing Transparency.
US 5,162,828
Nov. 10, 1992
(k)
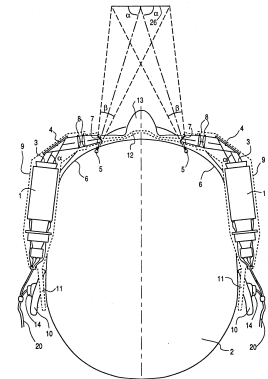
Holakovszky. Stereoscopic video
image display
appliance wearable on head like
spectacles.
US 5,129,716
Jul. 14, 1992
(l)
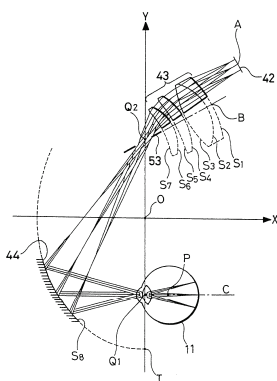
Iba. Image Observation Device.
US 5,384,654
Jan. 24, 1995
(m)
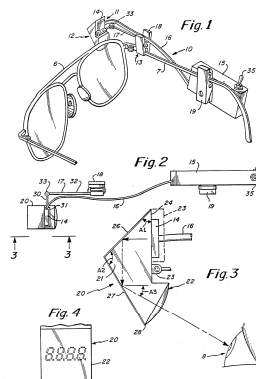
Kubik. Headwear-mounted Periscopic Display
Device.
US 4,753,514
Jun. 28, 1988
(n)
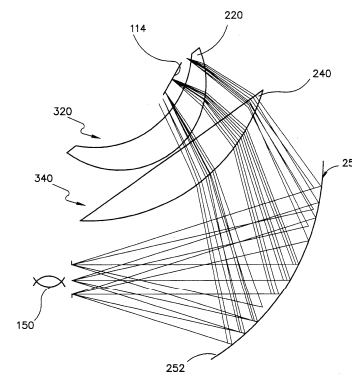
Ferrin. Headgear Display System
Using
Off-axis Image Sources.
US 5,576,887
Nov. 19, 1996
(o)
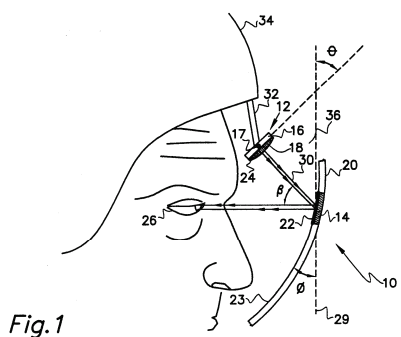
Fig.1

Lippert. Visor Display with Fiber Optic
Faceplate Correction.
US 5,309,169
May. 3, 1994
(p)
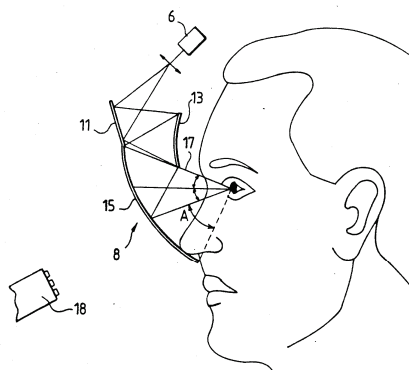
Lacroix. Device for the Display of Simulated
Images for
Helmets.
US 5,184,250
Feb. 2, 1993
(r)
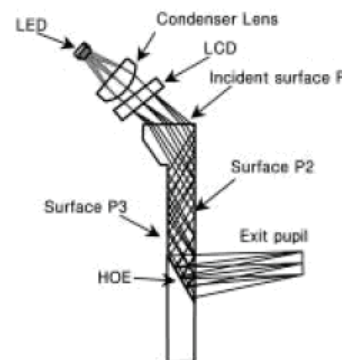
Kasai. A Forgettable Near-Eye
Display.
ISWC 2000.
(s)
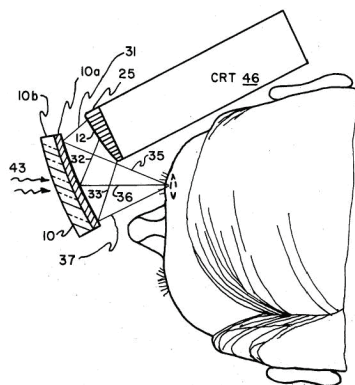
Bosserman. Toric reflector display.
US 4,026,641
May. 31, 1977
(t)
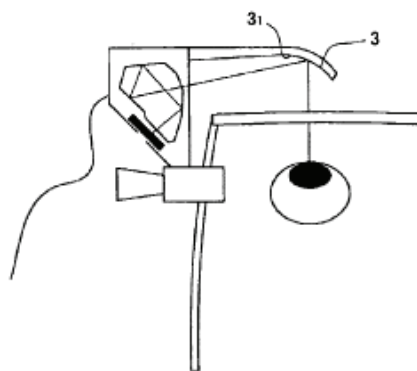
Nagaoka. Light weight head mounted image
display device.
US 6,697, 200
February 24, 2004
(u)
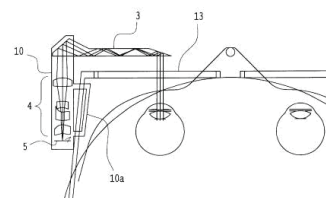
Takeyama. Observation optical system
US 6,710,902
Mar 23, 2004
(v)
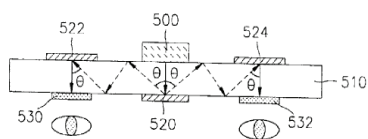
Song. Wearable display system.
US 6,882,479
Dec. 17, 2001
(w)
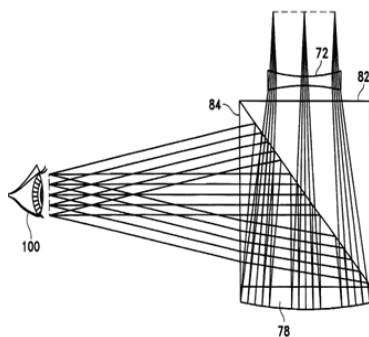
Robinson. Video headset.
US 5696521.
Dec 9, 1997.
(x)

FIG. 1.
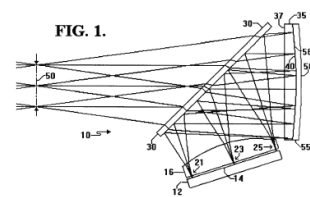
Fritz. Head mounted display using
mangin mirror combiner.
US 5838490.
(y)

39

Several researchers are exploring the applications of HWDs in various fields. To date, HWDs have been used in applications such as supporting human memory [148], [95], future machine tools [51], factory automation [192], computer-supported collaborative work [15], personal imaging [114], telemedicine [141], urban environment guides [60], remote collaboration [11], computer games [191], [132], [44], archeological sites [205], [48], astronaut extravehicular activity [39], and medical applications [133]. A survey of augmented reality technology between 1997 and 2001, containing information on displays, new tracking sensors and applications, calibration and autocalibration, interfaces and visualization components and application can be found in [8].

Even though the markets might be driven initially by vertical applications, such as training, maintenance, manufacturing, and medical, we believe that it is the more horizontal adoption that will embed HWDs into societies around the world. In this low FOV regime, most of the HWD applications are expected to target mobile users. As the Lagrange invariant is relatively low compared to regimes with higher FOVs, the designs are likely to have a lower element count [35] leading to potentially wearable, compact and lightweight displays with acceptable image quality.

Table 2.1 presents a collection of optical designs reasonably representative of HWDs in the low FOV regime that aim to fit to the eyeglass form factor. In terms of the number of elements, designs fall between anywhere from one and six elements. The majority of these designs is within the one to three element(s) range, and laid out in off-axis geometries. Off-axis geometries can be desirable for conforming to the shape of the human head. All of the designs shown in Table 2.2 are catadioptric, except one HWD that makes use of a holographic element.

An important metric is the value of the MTF at the Nyquist cut-off frequency given as 1/(2*pixel spacing). Typically a designer would aim at an as-built MTF value of 20% at the Nyquist cut-off frequency. The as-built MTF accounts for degradations imposed by the optical and the optomechanical tolerances. The spatial frequency corresponding to 50% MTF value is typically selected to analyze the robustness of the system to tolerances. Tolerances are determined by typically allowing an overall drop in MTF performance of less than 15%. Another metric that can differentiate these designs is the required number of compensators (i.e., adjustable groups or elements at system assembly) required to meet the as-built MTF specification. If a design is sensitive to positioning during the assembly phase, it will impose tight tolerances and additional optomechanical mechanisms will be necessary that add to the volume, weight and complexity of the overall design. It is desirable to have loose tolerances that allow for simpler optomechanical arrangements.

2.6 Applications of Head-Worn Displays: Mid Field of View Designs (40 to 60 Degrees)

In this regime, example applications target scientific visualization, simulation and training, and medical training [4]. Table 2.3 provides seven designs each having between 40 and 60 degrees FOV.

The Canon Mixed Reality Laboratory has developed a hybrid HWD with a camera that is able to record the FOV the user is looking at and can operate in see-through mode simultaneously. The freeform prism designed by Yamazaki *et al.* [224] successfully folds the optical path in a compact package. The optical layout of one of the prototypes is shown in Fig. 2.5. The "COASTAR" HWD has a 60 degrees diagonal FOV, 12 mm exit pupil diameter, 20 mm eye relief. The total thickness of the prism was reported to be 17.9 mm. The image plane was placed at 2m. This design form has recently been pushed to support an 80 degrees diagonal field of view [90].
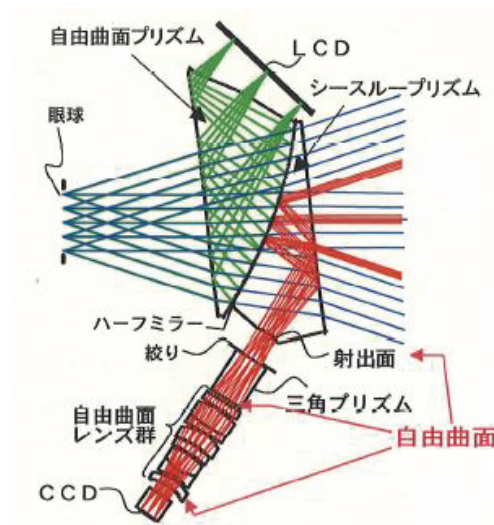
Figure 2.5 Canon video see-through system. (Adapted from [224])

Table 2.3 Mid field-of-view designs (Between 40 and 60 Degrees)

Chen. Helmet visor display employing reflective, refractive and diffractive optical components.
US 5,526,183
June 11, 1996
(a)

Chen. Wide spectral bandwidth virtual image display system.
US 5,436,763.
July 25, 1995.
(b)

Chen. Ultra-wide field of view, broad spectral band visor display optical system.
US 5,499,139
March 12, 1996.
(c)

Takeyama. Image display apparatus.
US 6,342,871
January 29, 2002
(d)

Togino. Visual display apparatus.
US 5,436,765
July 25, 1995
(e)

Becker. Head Mounted Display for Miniature
Video Display System.
US 5,003,300
March 26, 1991
(f)

Erfle. Ocular.

US 1,478,704

Dec 25, 1923

(g)

2.7 Applications of Head-Worn Displays: Wide Field of View Designs (>60 Degrees)

Wide FOV designs are potentially applicable for applications requiring immersion such as simulation and training and computer games. Buchroeder [23] designed a 48.5 degrees 56.5 degrees per eye with a 15 degrees overlap thereby achieving a 102 degree image. The design is reported to have a 12 mm exit pupil and a 55 mm eye relief. In terms of collimation, dipvergence is 15 arcminutes, divergence is 40 arcminutes, and convergence is 25 arcminutes. Distortion of the design was specified to be ±1%. Color correction across 486–656nm was reported to be less than 1 arcminute for axial color and less than 3 arcminutes for lateral color. The final design is an achromatized version of the layout shown in Fig. 2.6. The fiber optic faceplate shown in Fig. 2.6 was eliminated in the final design in exchange of a field flattening single lens element.
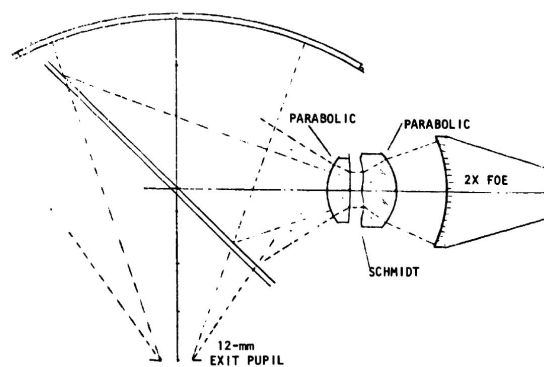
Figure 2.6 Visually coupled airborne systems simulator (VCASS) . (Adapted from [23])

A classic and compact design in this regime is the Pancake window [104]. In the original system a single curved, spherical beamsplitting mirror was used as the image-forming element. More compact pancake windows have since been designed using holographic combiners. The Pancake window is based on polarization optics and the original version has a very low light throughput at about 1%. Using cholesteric liquid crystals, the light throughput of the original Pancake window has been improved to 20% [14]. The HWDs, based on cholesteric liquid crystal type pancake windows, have been built in tiled configurations, providing a FOV about 100 degrees horizontal by 30 degrees vertical. Later versions of the tiled pancake windows are reported to reach fields of view of about 50 degrees by 150 degrees with 4 arcminutes resolution [121].

Using nodal aberration theory, one can predict that the astigmatism and coma from the tilted combiner can be corrected by a rotationally symmetric, but tilted, optical system. This idea was employed in the system shown in Fig. 2.7 to achieve a high FOV (100°x50°) rotationally symmetric system with a 50 mm eye relief and a 15 mm exit pupil size. Rotational symmetry leads to ease of fabrication. Writings of Thompson [193-195] and Rodgers [154] provide fundamentals towards understanding nodal aberration theory and its application to tilted and decentered systems.
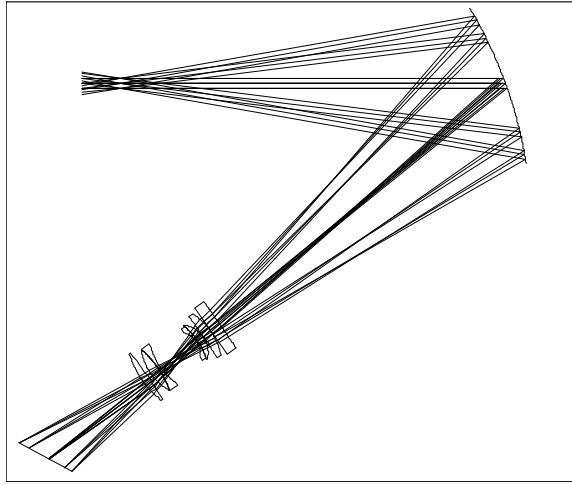
Figure 2.7 (Adapted from [177])

Rodgers [153] discusses the benefits of using freeform surfaces in order to reduce astigmatism from oblique rays. Astigmatism will be reduced dramatically, for example, for a single surface, broken into local patches, in the case where each patch is symmetric from the point of view of the chief ray of the beam hitting the patch. As ray-trace codes adopt tools and metrics to evaluate the degree of difficulty for manufacturing a surface, we can expect the newer HWD designs to utilize more free-form optics. The magnitude of the sag of a non-rotationally symmetric surface has been proposed as a metric to quantify the degree of fabrication difficulty [67]. The authors note that each process would have a unique cost function, thus, until general and flexible metrics are in place characterizing different processes, the authors chose not to implement this single metric in their ray-trace code for the time being.

Huxford designed a 120°x67° FOV display for a low-cost driving simulation application [89]. The system has a 15 mm eye relief, 20 mm exit pupil, ≤0.25% distortion, ±4 D focus

46

adjustment, and a pixel-limited resolution of less than 4 arcminutes. The optical layout of this system is shown in Fig. 2.8. Huxford was using an FLCOS as the microdisplay. The eyepiece in this design was based on the Pancake window. The polarizers in the system were utilizing wire-grid technology. The eyepiece had >27% distortion. In order to achieve 50 lp/mm resolution across the FOV and the weight target, the author was using a combination of plastic and glass materials. The relay is composed of six lenses, four polymer lenses each having an aspheric surface and two glass lenses with spherical surfaces. The strong aspheres close to the screen help with distortion correction. The authors also designed an all polymer relay and they report that the hybrid relay performed 33% better in terms of the transverse color correction.
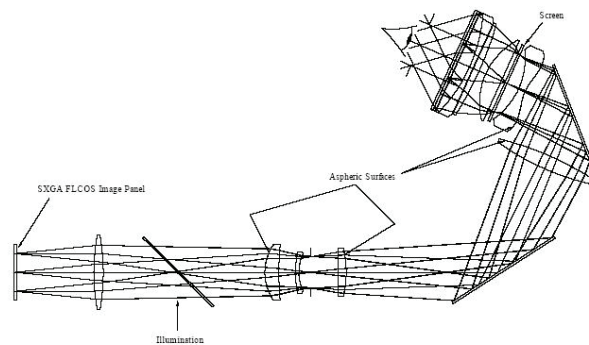
Figure 2.8 An example wide field of view system (120°x67°)

CHAPTER 3 FUNDAMENTALS OF OPTICAL DESIGN PERTINENT TO THE DESIGN OF HEAD-WORN DISPLAYS

In this chapter we review some of the optical design fundamentals and relevant theory that are applied in the consequent chapters. As this thesis is primarily concerned with the design of off-axis freeform magnifier systems, we provide a review of diffractive optical elements, design of aspheric surfaces, and freeform surfaces because these types of optical elements are useful in the design of compact off-axis magnifiers. One research question in this thesis will be to establish the optical performance, eyebox diameter, and field of view limits in particular off-axis magnifier systems, where the design constraints such as tilt angles of the elements, eye clearance constraint, etc. isolate the major optimization variable to be the shape of the surface. One way to investigate such limits is to couple a general framework for shape representation to optimization and analysis tools from aberration theory. The aspheric shape profile generation problem, in the case of an aplanatic system, can be solved by noniterative techniques. The Wasserman-Wolf technique serves as one of the cornerstone ideas in noniterative aspheric profile generation, for stigmatic imaging of a point to a point, and will be reviewed. However, the presence of multiple fields and constraints leads us to connect optimization techniques with a surface representation. Therefore, we rely on optimization theory while finding the weights of the linear combinations describing the surface shape. We will summarize the industry standard optimization algorithm called Levenberg-Marquardt also known as damped least squares, and the Karush-Kuhn-Tucker saddle point system, which is the generalized Lagrange multiplier technique for handling constrained optimization problems. Our surface representation approach

will be described in detail in chapter 5. In terms of performance analysis, we will review nodal aberration theory (a.k.a. vector aberration theory), which provides a basis for analysis in off-axis systems, especially through the full field displays. Specifically, nodal aberration theory provides a way to connect rotationally symmetric systems with rotationally non-symmetric systems by establishing an optical axis ray in rotationally non-symmetric systems. Therefore, aberration theory of rotationally symmetric systems remains relevant in the analysis of rotationally non-symmetric systems and will be briefly reviewed.

3.1 Diffractive Optical Elements

Diffractive optical elements are referred to under several names such as binary optics (discrete phase) or kinoforms (smoothly varying phase). Diffractive optical elements (DOE) provide the optical designers with a) dispersion characteristics opposite to that of refractive materials that enable their use in *color correction* and b) the implementation of *arbitrary phase profiles*. Unlike reflective or refractive elements, DOEs diffract light into various orders and usually only one order is utilized in an optical design while the other orders end up as background intensity. The designer's ability to direct as much light as possible into the design order is measured via the diffraction efficiency. Typical losses in the diffraction efficiency of a DOE are caused by wavelength detuning or fabrication errors in the desired profile. DOEs may be fabricated by diamond turning or photolithography. Designers need to be aware of the minimum feature sizes that can be achieved during fabrication to use DOEs successfully. In our work, we have been fabricating diffractive optical elements using the diamond turning process

while limiting the minimum feature sizes to the order of 10 microns. At the visible wavelengths, the minimum feature size constraining the grating period in the diamond turning process is about 20 times larger than the mid-wavelength (~0.5μm), therefore the scalar approximation to the Maxwell's equations is valid. The systems we are concerned with do not have a high-NA (NA>0.6) in order to require that the diffraction integral accounts for polarization effects [115].

Diffraction gratings can be categorized into amplitude and phase gratings. Amplitude gratings transmit a certain percentage of light and absorb or reflect the rest. Theoretically, phase gratings can direct the entire incident light for a particular wavelength into one diffraction order. A desired phase variation can be imposed on the incident light by either modulating the refractive index or by varying the thickness of the substrate.

Geometrical optics is inadequate in describing the intensity distribution at an observation plane of a light wave passing through a narrow slit with physical dimensions comparable to the size of the wavelength. Diffracted orders from a slit are illustrated in Fig 3.1.
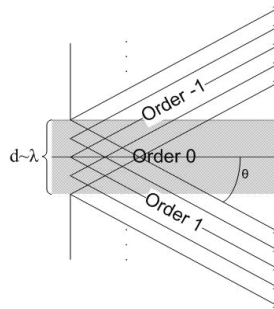
Figure 3.1 Diffraction from a slit with size comparable to that of a wavelength. Only three orders (m=0 and m=+-1) are shown.

Huygens introduced the idea of breaking down a wave radiated by a source into secondary point sources that radiate spherical waves and the superposition of waves radiated by the secondary sources determines the propagating wavefront. Kirchhoff formalized the idea of a Huygens wavelet in his integral. Rayleigh-Sommerfeld removed the boundary condition inconsistencies of the Kirchoff integral [19]. The geometry for the Rayleigh-Sommerfeld integral is illustrated in Fig. 3.2. In the context of diffractive optical elements, the Rayleigh-Sommerfeld integral allows us to compute the far field intensity pattern of a diffractive transmission function and can be written as [72]

$$a(x_o, y_o, z_o) = \frac{1}{j\lambda} \iint_A \gamma \, a(x_1, y_1, 0) \frac{e^{jkr_0}}{r_o} dx_1 dy_1 \qquad (3.1)$$

where $a(x_1, y_1, z = 0)$ is the field amplitude at the aperture plane and $a(x_0, y_0, z_0)$ is the diffracted field amplitude at some positive plane z_0, r_0 is the distance between the aperture origin and the observation point, γ is the obliquity factor [72].
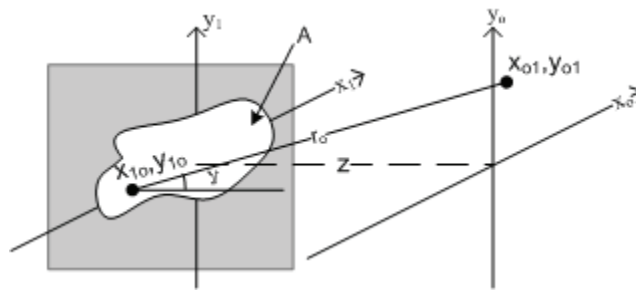
Figure 3.2 Rayleigh-Sommerfeld diffraction integral geometry.

Under Fraunhofer conditions, the far field intensity pattern can be written as a Fourier transform of the aperture function. This property is used to calculate the diffraction efficiency of

diffractive optical elements in the discussion below. As Swanson points out [188], it is possible to draw an analogy between refractive and diffractive elements. The importance of relating refraction and diffraction is that it will help us understand 1) how arbitrary phase profiles can be accomplished using diffractive elements and 2) the relation of diffraction efficiency in the case of a blazed grating and an arbitrary phase profile. One can view both a diffractive element and a refractive element as wavefront converters. For example, a refractive lens converts a plane wave input to a focusing spherical wave. The same effect can be achieved with a Fresnel zone plate through diffraction. Fresnel zones are illustrated in Fig. 3.3. Fresnel zones divide a wavefront into equal area zones with radii r+λ/2, r+λ, r+3λ/2 and so on. The contributions from each neighboring zone are out of phase. Therefore, at some observation plane away from the Fresnel zone plate, we have the following on-axis field amplitude

$$a(x_0 = 0, y_0 = 0) = |a_1| - |a_2| + |a_3| + ... \quad , \tag{3.2}$$

where a_x denotes the on-axis amplitude contributed by points in a given zone x. Neglecting the obliquity factor, the amplitude of a given zone is equal to the average of the two adjacent zones

$$\frac{1}{2}|a_{n-1}| - |a_n| + \frac{1}{2}|a_{n+1}| \approx 0 \tag{3.3}$$

$$A(x_0, y_0) = \frac{1}{2}|A_1| + \overbrace{\frac{1}{2}|A_1| - |A_2| + \frac{1}{2}|A_3|}^{0} + \overbrace{\frac{1}{2}|A_3| - |A_4| + \frac{1}{2}|A_5|}^{0} + ... = \frac{1}{2}|A_1| \tag{3.4}$$

the on-axis field amplitude is half of the contribution by the first zone. By blocking the even or odd zones, it is possible to increase the on-axis irradiance. Therefore, a Fresnel zone plate is an example of diffractive lens (in the form of an amplitude grating) that can create an analogous focusing effect to that of a refractive lens.
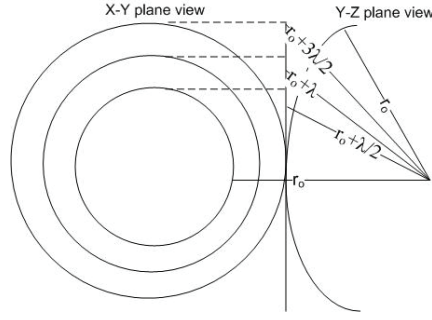
Figure 3.3 Illustration of the Fresnel zones.

Another refractive-diffractive analogy can be made between a prism and a blazed grating. A prism deflects rays by refraction and a grating deflects rays into the diffracted orders by diffraction. The phase of a prism is shown in Fig. 3.4(a). We use a nonlinear limiter such as the mod function to map the linear phase of the refractive prism into a 2π periodic phase and the phase after the mod function is shown in Fig. 3.4(b).

(a)   (b)
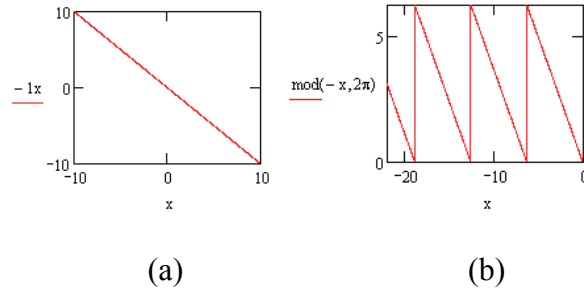
Figure 3.4 (a) Linear phase of a refractive prism (b) Equivalent diffractive phase

The etching or diamond turning depth can be calculated by setting the OPD equal to a wavelength that results in

$$d = \frac{\lambda}{n-1} \tag{3.5}$$

where d is the depth of the diffractive profile, λ is the wavelength of operation and n is the refractive index of the diffractive substrate. In the visible part of the electromagnetic spectrum, the DOE profile depth is on the order of 1 micrometer, assuming a plastic substrate with a refractive index of about 1.5.

We will next describe the transmission function for the phase grating shown in Fig 3.5 and analyze the diffraction efficiency. The diffracted orders and the grating geometry are shown in Fig 3.5.
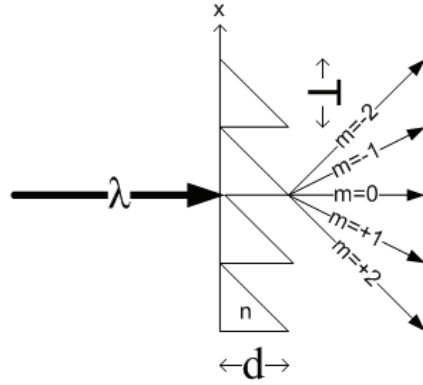
Figure 3.5 Geometry of a phase grating.

The transmittance of the surface relief grating is described as [188]

$$t(x) = \sum_{m=-\infty}^{\infty} \delta(x - mT) * rect(\frac{x}{T}) \exp(i2\pi\beta x) \tag{3.6}$$

$$F(f) = \sum_{m=-\infty}^{\infty} \delta(f - \frac{m}{T}) \frac{\sin(\pi T(\beta - f))}{\pi T(\beta - f)} \tag{3.7}$$

The amplitude for the mth diffraction order is given by

$$a_m = \frac{\sin(\pi T(\beta - f))}{\pi T(\beta - f)}$$

(3.8)

The diffraction efficiency is given as a ratio of the output to the input intensity and yields

$$\eta_m = \left[\frac{\sin(\pi T(\beta - f))}{\pi T(\beta - f)} \right]^2$$

(3.9)

As Swanson points out, the diffraction efficiency of an arbitrary phase profile is identical to the linear phase profile (equation 3.8). The steps in establishing the equivalence between the diffraction efficiency of an arbitrary phase profile and a linear phase grating involves writing the diffractive phase as a modulo of the refractive phase. Therefore, the diffractive phase is periodic as a function of the refractive phase allowing one to write it as a Fourier series. The expression for the orders of an arbitrary phase profile is detailed in the Appendix. Arbitrary phase profiles are typically implemented and optimized in optical design codes as a rotationally symmetric polynomial. The expressions for diffraction efficiency can be extended to include number of phase levels during fabrication [188].

Buralli and Morris provide a way to calculate the impact of diffraction efficiency on the MTF [24]. Buralli and Morris show that the polychromatic integrated diffraction efficiency acts as a scale factor for the entire modulation transfer function. The polychromatic integrated diffraction efficiency is defined by Buralli and Morris as

$$\eta_{int,poly} = \frac{\int_{\lambda_{min}}^{\lambda_{max}} \eta_{int}(\lambda)d\lambda}{\lambda_{max} - \lambda_{min}} \tag{3.10}$$

where $\eta_{int}(\lambda)$ is a pupil-averaged integrated efficiency value as a function of wavelength.

We proceed to summarize the steps for understanding color correction using diffractive optical elements. Achromatic color correction can be achieved by requiring optical powers at two wavelengths to be equal. The condition for axial color correction is [97]

$$\frac{\phi_1}{V_1} + \frac{\phi_2}{V_2} = 0 \tag{3.11}$$

Therefore, it is important to understand the dispersive and the first order optical power properties of diffractive optical elements in order to understand their color correction properties, which is discussed next.

The Abbe number is defined to be

$$V = \frac{n_d - 1}{n_f - n_c} \tag{3.12}$$

Modeling a diffractive optical element as a Fresnel zone plate yields the expression

$$f_o = \frac{r_m^2}{m\lambda_o} \tag{3.13}$$

where r_m is the radius of the mth zone that is a half wavelength away from the focal length, m the is the zone number, λ_o is the wavelength, and f_o is the focal length. Therefore,

$$f(\lambda) = \left(\frac{\lambda_o}{\lambda}\right) f_o \qquad (3.14)$$

The optical power is defined to be

$$\phi(\lambda) = \frac{1}{f(\lambda)} = (n-1)c \qquad (3.15)$$

Solving for n and using (3.16) in conjunction with (3.17) yields

$$n = \frac{1}{f(\lambda) \cdot c} + 1 = \frac{\lambda}{\lambda_o f_o c} + 1 \qquad (3.16)$$

Using (3.15) in conjunction with the definition of the Abbe number yields

$$V = \frac{\lambda_d}{\lambda_f - \lambda_c} = -3.452 \qquad (3.17)$$

Where C, d, and F are the spectroscopic lines at 656.3nm, 587.6nm and 486.1nm respectively. We note that the Abbe number of a diffractive optical element does not depend on the surface profile.

In terms of the optical power, in the case of a refractive lens, the optical path difference between a plane wave and the wave propagating in a lens, using an approximate sag expression, is given by

$$OPD = (n-1)\frac{r^2}{2R} = r^2\frac{\phi}{2} \qquad (3.18)$$

where is r is the radial distance on the lens, R is the radius of curvature, and ϕ is the optical power. Radially symmetric diffractive optical element profiles are defined as a polynomial in optical design codes as

$$OPD = c_1 r^2 + c_2 r^4 + c_3 r^6 + ... \qquad (3.19)$$

Therefore, the first coefficient in a diffractive lens defined with a polynomial sum is related to the focal length of the DOE as

$$c_1 = \frac{\phi}{2} \qquad (3.20)$$

3.1.1 A Comparative Analysis of Doublets versus Single-Layer Diffractive Optical Elements in Eyepiece or Magnifier Design

Eyepieces constitute an important subsystem in several optical systems such as head-worn displays, telescopes and microscopes. Several applications demand high performance, large field of view and compact physical size eyepiece designs, which make eyepiece design a challenging problem. Dominant aberrations in eyepieces are axial and lateral color, distortion, astigmatism and field curvature.

Mainstream color correction techniques for eyepieces include the use of doublets [56] and diffractive optical elements (DOE) [210]. Color correction level is typically categorized by the number of different wavelengths brought to a common focus. The terms achromatic, apochromatic and superapochromatic are used in reference to bringing two, three and four wavelengths, respectively, to a common focus. Herzberger states that since the dispersion of glass can be represented by an equation containing four constants, the superapochromat is practically corrected for all colors [82].

Glass choice is a key parameter to color correction using doublets. Doublets can be cemented with either crown first type where the high dispersion material is in the first element or a flint first type where the low dispersion material is in the first element, split doublet where there is an air space between the two elements providing an extra optimization variable, or Gauss type with two meniscus elements. Typically cemented doublets are preferred over split doublets given that they allow looser tolerances and do not require a sub-cell assembly that may impact

compactness. For the cemented type, crown first type is preferred as it provides more robustness and looser tolerances [97]. Optical power is set by the design requirements and once the proper glasses are chosen the only optimization variable left is the shape of the doublet in the case of cemented doublets [97]. Furthermore, the Seidel coefficient for Petzval curvature depends solely on the refractive index and the optical power of each element, and the Seidel coefficient for axial color depends solely on the dispersion of the material and the optical power, thus the two aberrations are linked through material choice. Hence, simultaneous correction of Petzval curvature and axial color may be challenged provided the limited choice of glasses. The human visual system has the ability to refocus and some amount of field curvature (defocus as a function of field) can be tolerated. Specifically, based on experiments of Burton and Haig, the human visual system can tolerate up to 0.3+-0.019 waves of peak to valley defocus wavefront aberration with a 90% threshold of discriminability. Converting the wavefront defocus aberration to visual space yields an aberration tolerance of approximately 0.5 arcminutes. This number is computed using the relation between the wavefront aberration and the longitudinal defocus, and then converting the longitudinal defocus amount to the transverse amount in visual space.

Diffractive elements can have an Abbe value of approximately -3.5, whereas optical glasses and plastics have an Abbe value between 25 to 75 that makes them complementary to the dispersion characteristics of optical glasses and plastics at least within the visible region of the spectrum. Such complementary dispersion characteristics are desirable in the design of color corrected eyepieces. Furthermore, Stone and George have shown that one can shift the power distribution between the diffractive and the refractive elements while maintaining constant overall power and the resulting Abbe number of the combined hybrid can span all real numbers except a small interval around zero. The ability of having an arbitrary Abbe number while

holding the total optical power constant can help eliminate the use of exotic or expensive glasses in certain applications. The drawback in utilizing a DOE is the potential contrast loss caused by light diffracted into undesired orders. Emerging multi-layer diffractive optics can potentially help minimize such contrast loss [130]. Diffraction efficiencies of multi-layer DOEs have been reported to be considerably higher (broadband efficiencies approaching 100%) than that of single-layer DOEs across the visible spectrum. Canon's implementation of this technology in their products makes it plausible that such technology can be practical or cost-effective.

Advances in fabrication techniques enable the implementation of diffractive optical elements on the surface of aspheric substrates. Several techniques are available for the fabrication of diffractive optical elements including diamond turning and photolithographic techniques. The minimum feature size imposed by fabrication limitations is an important parameter in the design of DOEs. Feature sizes as small as 0.1 μm have been fabricated [105]. Once a master element is fabricated it is possible to replicate the master element using techniques such as cast-and-cure replication, injection or compression molding, or etching into a substrate [126].

In spite of cost and challenges in fabrication and testing, diffractive-refractive hybrids have been shown to yield an overall increase in performance while providing a considerable decrease in weight for eyepiece designs. To-date several optical designers have successfully used diffractive optical elements in their systems [47, 58, 88, 100, 124, 211]. Cox et al. used DOEs in a head-mounted display design having a 15 mm exit pupil and 20 degree field of view across the visible band and they reported a 10-fold decrease in weight in the DOE-based system when compared against a doublet-based solution [47]. They achieved 93.5% diffraction efficiency in their design, and the diffraction inefficiency in this system caused a few percent less MTF (at

~4 cycles/mm) than the baseline MTF. Missig and Morris designed hybrid elements with more than 60 degrees FOV and reported a 70% weight reduction compared with the Erfle eyepiece as well as a 25% reduction in distortion [124]. Knapp et. al. compared an eyepiece using a DOE to an eyepiece using aspheric elements [100]. The development criterion for the lenses was a 60 degree full FOV, 24 mm focal length, 10% distortion, 12 mm exit pupil, and a 20 mm eye clearance over the 540-558nm spectral band. Aberration tolerances were ±2 diopters on the focus range, ±0.25 diopters on the field curvature, <0.5D for astigmatism. The authors drew the conclusion that the DOE yields a better optical design from a chromatic, field curvature, and MTF standpoint for their set of specific application requirements. Chen was one of the early adopters of diffractive technology and applied it to the design of eyepieces [43]. In the context of related work, to our knowledge, impact of the eye clearance parameter on the aberrations has not been quantified previously. The study of the impact of doublets versus diffractive optical elements on the eye clearance parameter is studied in Section 3.1.2.

3.1.2 Comparison of Doublets versus Diffractive Optical Elements (DOE) in Eyepiece Design

The comparison is motivated by the design of an eyepiece lens for an augmented reality application [28]. Given four different values of eye clearance (17 mm, 20 mm, 23 mm and 26 mm), we designed and optimized a doublet based system as well as a corresponding diffractive optical element based system. We define eye clearance to be measured from the exit pupil surface to the edge of the first surface assuming that the surface has a negative curvature. All lenses in this study have a negative curvature[1] for their first surface.

The starting point of the doublet resembles an Erfle eyepiece with a doublet, singlet and a doublet. Driven by the application domain, we are using the Erfle eyepiece in telecentric mode with a large back focal length (~30 mm), which is not a typical configuration for the Erfle eyepiece. However, such geometry is similar to that of LCOS illuminated eyepieces, thus the findings will be applicable to those eyepieces as well. As required by our application, both systems have a prism between their last surfaces and the image plane. For brevity, we shall refer to the system having the diffractive optical element as the DOE-based system and the system resembling the Erfle eyepiece as the doublet-based system. We will also use the term "both systems" when referring to both simultaneously.

---

[1] Our sign convention is defined based on a vector with the tail at the vertex of the surface and the head pointing towards the center of curvature. If the vertex to center of curvature vector is pointing in the positive direction within the Cartesian convention, the curvature is taken as positive.

Both systems were optimized for a 31 mm focal length, 9 mm exit pupil, photopic wavelengths (513nm-608nm), and a 40 degrees full field of view. Both systems were operating in telecentric mode. The image distance was set to 2 m for both systems. In the DOE-based system, the DOE combined with an aspheric surface was located immediately prior to the prism surface, i.e., as close to the image plane as possible. It is desirable to have the aspheric surface as close as possible to the image plane in order to help correct for field aberrations such as distortion. Furthermore, combining the DOE and aspheric surfaces allow their fabrication simultaneously at no additional cost. Both systems included a 2.5 mm image clearance as a potential compensator for tolerancing and a 1 mm distance from the closest point of their last surface to the x-cube to ease the optomechanical design and assembly.

The optimization strategy included varying the glasses (except the glass of the prism that was fixed to an STIH-4 glass), curvatures of the surfaces, the aspheric and DOE coefficients up to the 10th order, and the spacing between the elements while ensuring at least 0.5 mm between the lenses to ease the tolerance on the optomechanics and assembly. After optimization, virtual glasses were fit to a subset of the following set of glasses: STIH-4, SLAM-2, SLAM-3, SLAM-7, SLAL-10, PBH-10, PBH-1, SBAH-28, and SBAL14. Each DOE-based system has one type of glass for the lenses. Curvature variations were constrained, when applicable, to avoid highly curved surfaces that lead to higher-order aberrations, intolerant lenses, and difficulties in fabrication. Fields were weighted heuristically as we proceeded through the design in order to bring the polychromatic MTF curves together and the wavelengths were weighted (1-2-1) to take into account the spectral response of the human visual system.

Simulations for each lens in comparison include the spherical aberration, coma, astigmatism, distortion, axial color, lateral color, and the MTF. The aberration

magnitudes have been extracted from real-rays. The longitudinal spherical aberration was measured as the distance along the optical axis between the paraxial image point at mid-wavelength and the intersection point of a real marginal ray at mid-wavelength with the optical axis. The paraxial image plane is the location where a paraxial ray at a small (e.g., 0.001) fractional pupil height at mid-wavelength crosses the optical-axis. The longitudinal spherical aberration is converted to a transverse blur by tracing rays that approximate the caustic. The transverse projection of the caustic on the best focus plane in visual space is reported as the transverse blur caused by spherical aberration. Tangential coma was calculated at the maximum field by finding the difference in vertical height between the location where the marginal rays and the chief ray intersect. Astigmatism was extracted as the maximum difference in the sagittal and tangential focus values projected on to the optical axis as a function of the height in the field measured at the green wavelength. From the astigmatic curves, the mid value of the medial focus curve was computed as a potential best focus for the system from which we estimated the impact of astigmatism and field curvature on visual performance. Distortion was measured as the ratio of the difference between the real ray height and the paraxial normalized by the paraxial ray height at the maximum field angle.

Axial color was measured by the difference in focus between the on-axis values for the green and blue wavelengths in millimeters and extracted through paraxial ray traces. Conventionally, one would report the axial color as the difference in focus for paraxial rays between the red and blue colors. However, we are interested in estimating the transverse blur caused by axial color and our paraxial image plane is placed on the mid-wavelength paraxial focus, therefore, we are reporting the difference in green and blue focus. The transverse blur caused by longitudinal axial color is estimated by projecting the longitudinal component to the

paraxial image plane. For various values in the field of view, lateral color was measured as the vertical translation of the chief ray image point at the short wavelength and the chief ray image point at the long wavelength. Lateral color was computed within the tangential field of view in one degree increments and we report the maximum lateral color value occurring within the field of view.

In the case of a DOE-based solution, we also computed the minimum feature size and the diffraction efficiency across the spectrum. The minimum grating period across the DOE-based solutions is 30 times greater than the mid-wavelength; therefore, the diffractive structures considered here fell well within the validity of the scalar approximation regime.   Additionally, we computed a quantity called the polychromatic integrated diffraction efficiency defined by Buralli and Morris as in equation 3.12. Buralli and Morris show that the polychromatic integrated diffraction efficiency acts as a scale factor for the entire transfer function. The impact of variation of incidence angles on the diffraction efficiency has been calculated.

A total of 6 lenses were designed and optimized for this study. The rationale for the number of lenses stems from the fact that the doublet-based system is unable to support an eye clearance distance greater than 20 mm while maintaining the desired performance criteria, yielding only two possible solutions in the context of this study. The DOE-based system supports almost the full range of eye clearance between 17-26 mm, documented in 3 mm increments. The performance of the DOE-based system starts to deteriorate for eye clearance distances higher than 80% of the effective focal length. We investigated the impact of the aspheric coefficients on the MTF at the 20 cycles/mm frequency. We found that a 10th order for the aspheric (or 4 coefficients) was optimal for this application, increasing the order did not yield a significant improvement in performance and decreasing the order did not take full advantage of the surface.

In addition to the aberration and fabrication characteristics mentioned above, documentation for each lens consisted of the optical layout and the MTF plot and the field curves. Table 3.1 and 3.2 contain the optical layouts, MTF plots, and field curves comparison for the 17 mm and 20 mm eye clearance. The MTFs are plotted out to 30 cycles/mm as set by the Nyquist limit. Table 3.3 lists the optical layouts, MTF plots and field curves for DOE-based systems with 23 mm and 26 mm eye clearance. MTFs were plotted up to the maximum spatial frequency of 30 cycles/mm as limited by the microdisplay for this desired application.

Table 3.4 quantifies and plots the aberrations for each system in visual space in units of arcminutes. We observe that the DOE-based system provides up to an order of magnitude better correction for both axial and lateral color. The doublet-based system is about 2 times better corrected for spherical aberration than the DOE-based system. Doublet-based system is one order of magnitude better-corrected for coma than the DOE-based system. The DOE-based system provides better astigmatism correction by about a factor of 2. Both systems suffer from field curvature owing to mostly positive elements resulting in a non-zero Petzval sum. The doublet-based system provides better field curvature correction, and this can be explained by the lack of negative elements in the DOE-based system. Distortion level comparison at the 17 mm eye clearance involves design choices. Inspection of the layout in Table 1 leads us to believe that it should be possible to reduce the element count by removing the third lens that looks like a plane parallel plate. The main trade-off in two versus three element solution at 17 mm of eye clearance is distortion, the three element solution offers about half of the distortion of a two element solution. The three element DOE-based solution provides improved distortion correction by a factor of 2 compared to the doublet-based solution at 17 mm. This finding may be explained given that the aspheric substrate under the DOE is not sufficiently close to the image plane to

produce better distortion correction. The MTF and field curves for the two element solution are shown in Figure 3.6. The distortion magnitudes for both systems are about the same at 20 mm eye clearance.

(a) (b)
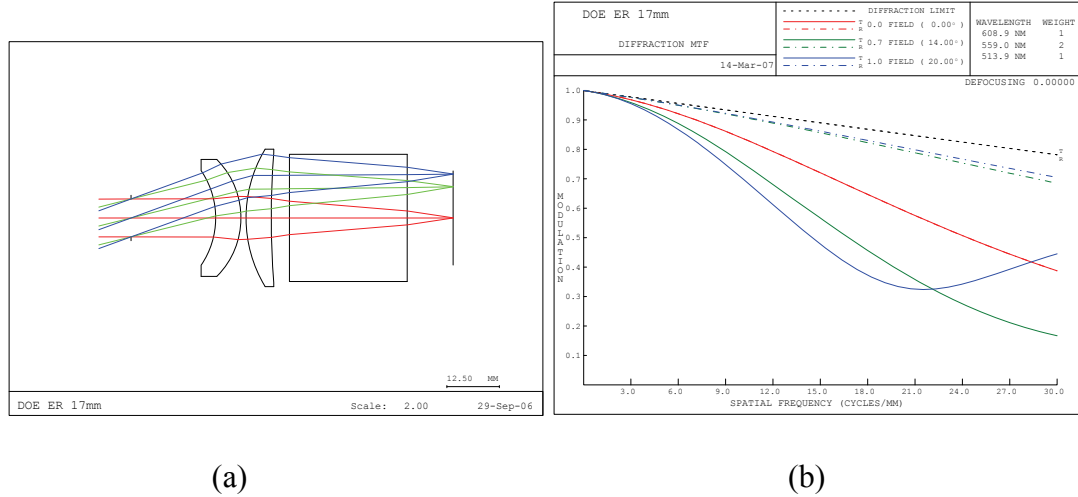
Figure 3.6 (a) Optical Layout of the 2-lens DOE solution at 17 mm eye clearance with a 9 mm pupil (b) MTF of the 2-lens DOE solution at 17 mm eye clearance evaluated at a 3 mm pupil

The diffraction efficiency across wavelength is shown in Fig. 3.7. An integrated polychromatic efficiency of $\eta_{int,\ poly}$ = 0.984 was found across all designs. Due to the telecentricity of the lens, the variation of the incidence angles on the diffractive surface were not large enough to cause a considerable drop (>1%) in the diffraction efficiency. The minimum feature sizes of the diffractive optical elements have been computed to be 22.5 μm, 16.3 μm, 15.2 μm, 15.8 μm, respectively for the DOE designs having 17 mm, 20 mm, 23 mm and 26 mm eye clearance specifications. According to our experience, DOEs with larger than 10 μm minimum feature

sizes can be fabricated without much difficulty. Table 3.5 lists a sample prescription for the

23 mm eye clearance DOE-based design.
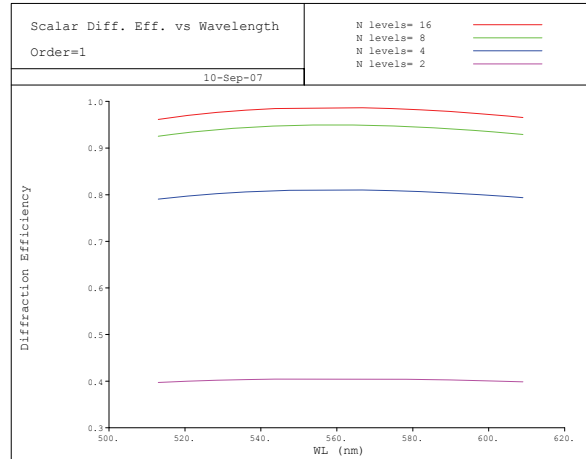
Figure 3.7 Scalar diffraction efficiency across the spectrum for 16 level masks for the design

wavelength of 558.98nm across 513-608nm.

Table 3.1 Optical layout, MTF and field curve comparison for 17 mm eye clearance
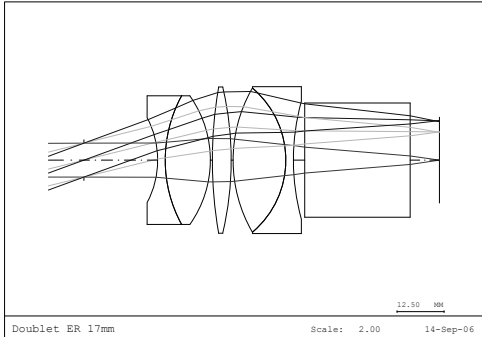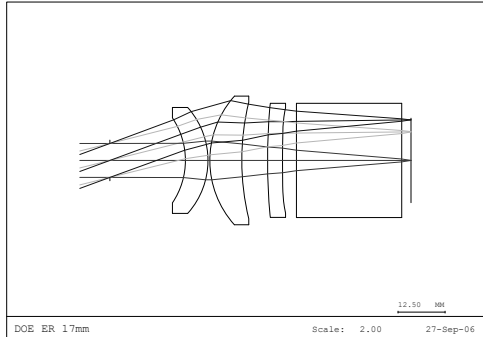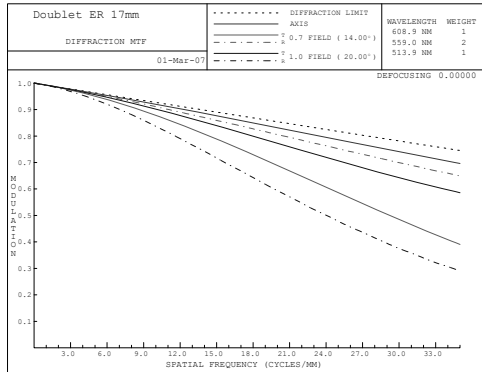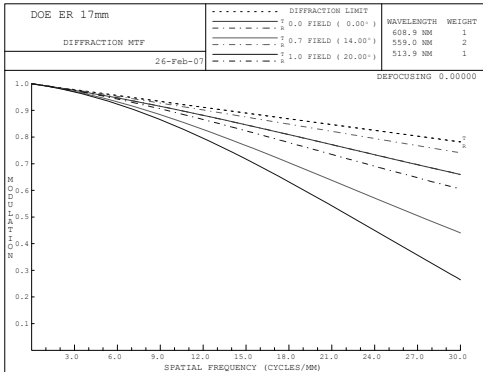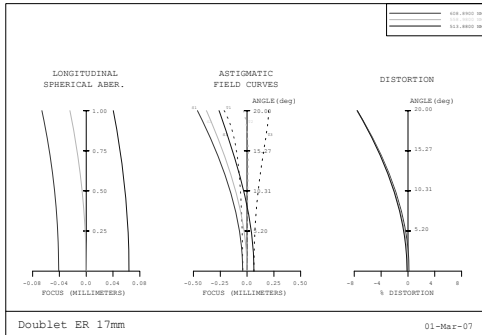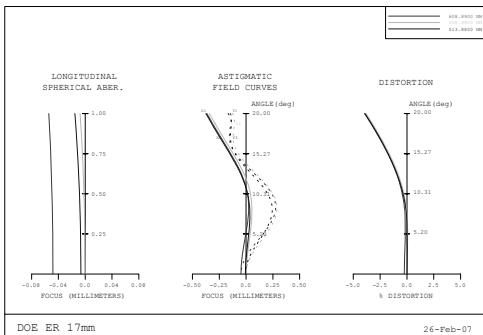
Doublet Based System	Diffractive Optical Element Based System
17 mm Eye Clearance	
Optical Layout (9 mm Exit Pupil)	**Optical Layout (9 mm Exit Pupil)**
Modulation Transfer Function (3 mm pupil):	**Modulation Transfer Function (3 mm pupil):**
Field Curves (3 mm pupil):	**Field Curves (3 mm pupil):**

Table 3.2 Optical layout, MTF and field curve comparison for 20 mm eye clearance
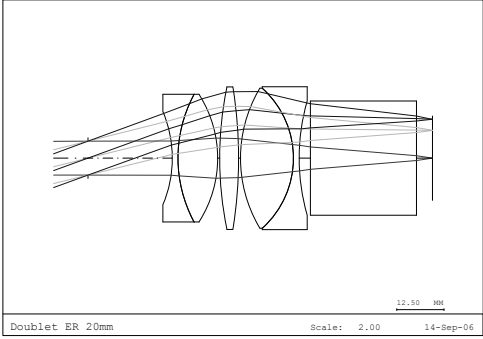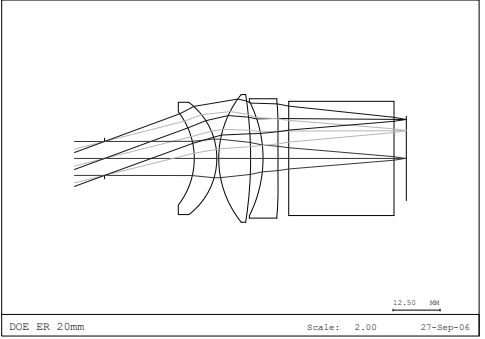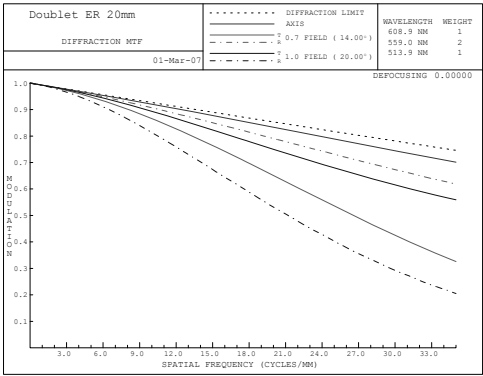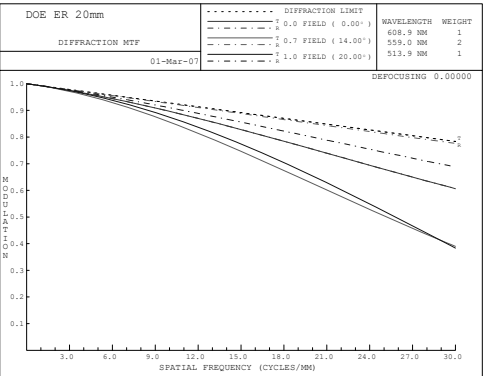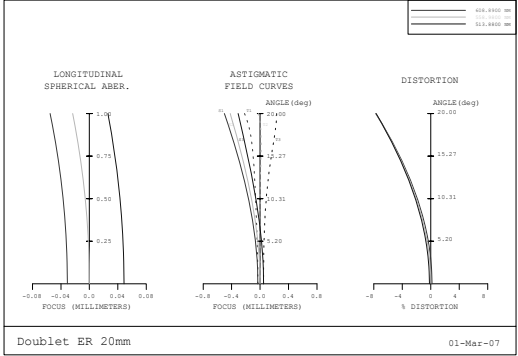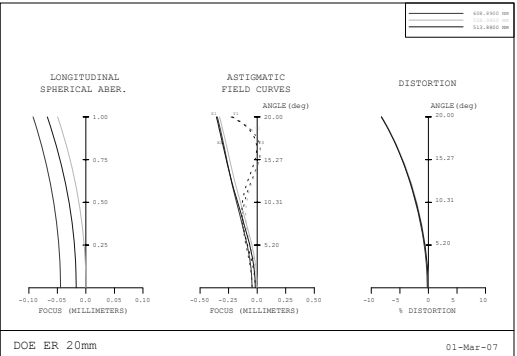
Doublet Based System	Diffractive Optical Element Based System
20 mm Eye Clearance	
Optical Layout (9 mm pupil): 	**Optical Layout (9 mm pupil):**
Modulation Transfer Function (3 mm pupil): 	**Modulation Transfer Function (3 mm pupil):**
Field Curves (3 mm pupil): 	**Field Curves (3 mm pupil):**

Table 3.3 Optical layout, MTF and field curve comparison for 23mm and 26mm eye clearance
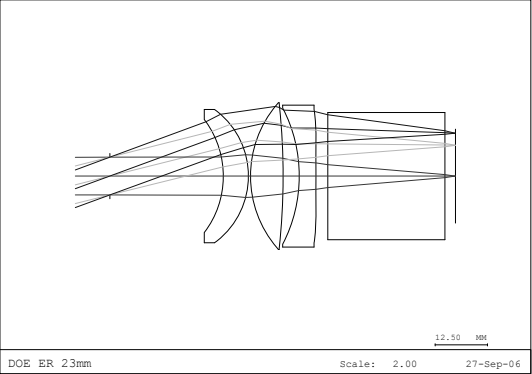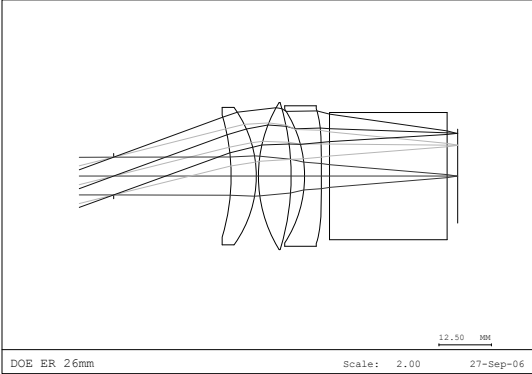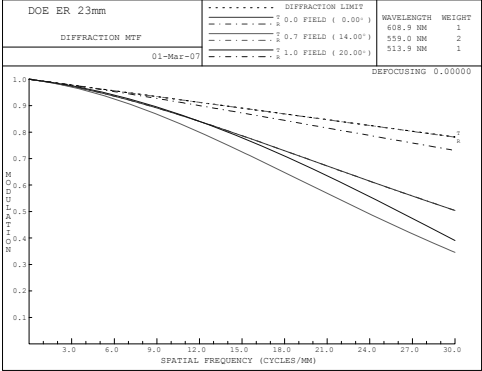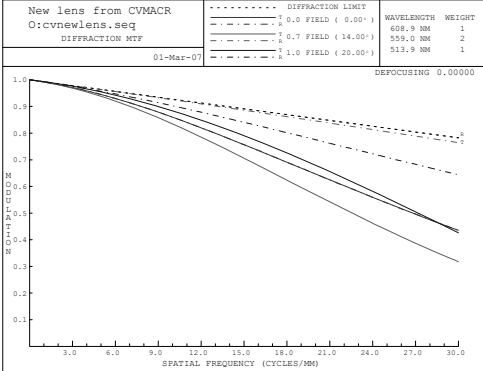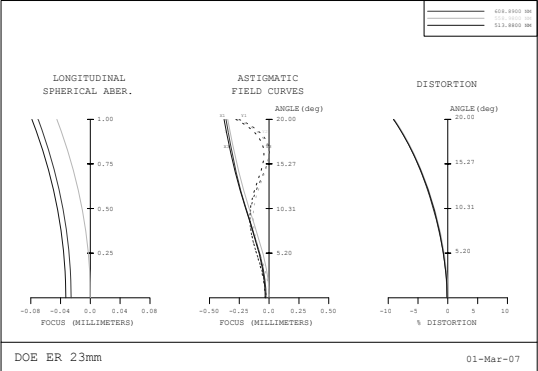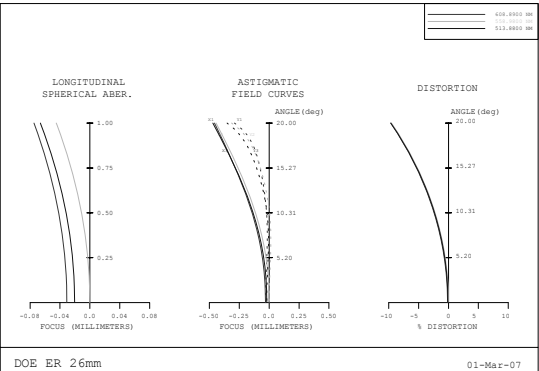
23 mm Eye Clearance DOE-based System	26 mm Eye Clearance DOE-based System
Optical Layout (9 mm Exit Pupil):	**Optical Layout (9 mm Exit Pupil):**
Modulation Transfer Function (3 mm pupil):	**Modulation Transfer Function (3 mm pupil):**
Field Curves (3 mm pupil):	**Field Curves (3 mm pupil):**

72

Table 3.4 Aberrations in the (a) doublet-based systems and (b) DOE-based systems as a function of eye clearance: (1) Spherical Aberration (2) Coma (3) Axial color (4) Astigmatism (5) Field curvature (6) Lateral color
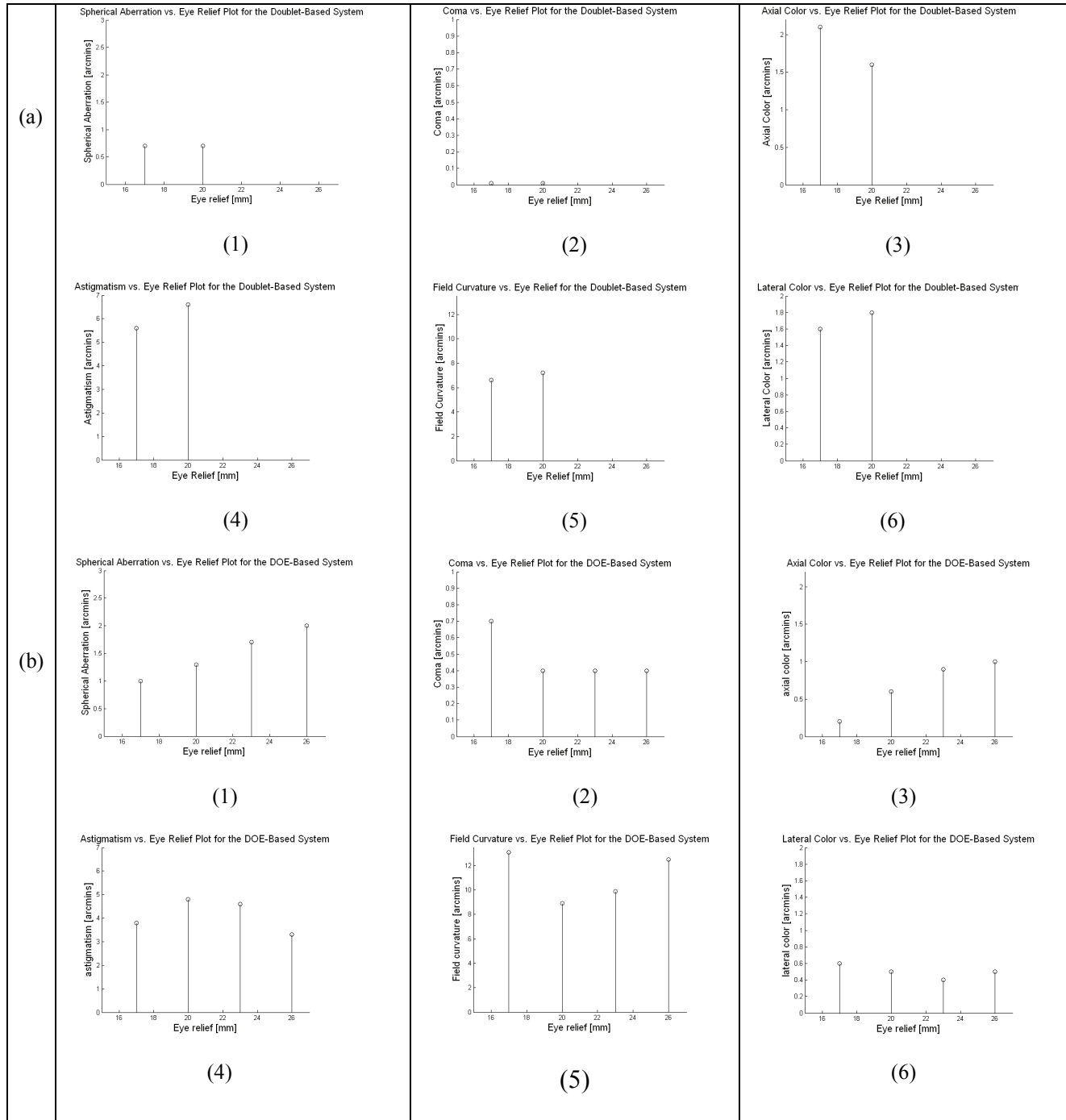
Table 3.5 Prescription of a sample lens.

DOE 23 mm

Surface #	Surface Type	Y Radius	Thickness	Glass	Y Semi-Aperture
Object	Sphere	Infinity	2000		
Stop	Sphere	Infinity	27.17		1.5
2	Sphere	-22.1	6	SLAM7	10.45
3	Sphere	-19.58	0.5		12.31
4	Sphere	26.18	7.8	SLAM7	14.1
5	Sphere	-160.25	3.85		13.58
6	Sphere	-35.22	4	SLAM7	12.83
7	Asphere	Infinity	2.85		12.71
8	Sphere	Infinity	28	755275	14
9	Sphere	Infinity	0		10.49
10	Sphere	Infinity	2.87		10.49
Image	Sphere	Infinity	-0.28		10.18

Aspheric coefficients: A: 1e-6 B: -1.205794e-8 C: -1.685598e-10 D: 3.307237e-13
Diffractive coefficients: C1: -0.002553 C2: 1.419239e-8 C3: 1.792873e-8 C3: -3.266823e-11

3.2 Aspheric Surfaces

Optical surfaces, either reflecting or refracting, are functions $f: \Re^2 \rightarrow \Re$ or surfaces of revolution where the base curve is a function $f: \Re \rightarrow \Re$. The term aspheric refers to surfaces that are not spheres or planes (i.e., spheres with infinite radius of curvature) [168]. Majority of optical designs to date have relied on spheres, conic sections and polynomials. The justification of spheres stems from the ease of manufacturing and testing. The choice of conic sections is easily

justified since conic sections are stigmatic surfaces under certain imaging configurations. The choice of polynomials from the point of view of surface description can be challenged. The advantage of using polynomials is that the wavefront aberration function is typically expanded as polynomials. Therefore, a polynomial surface description may link designers understanding of wavefront aberrations and the surface description. The disadvantage of using polynomials is twofold. First, several optical design applications require optimization to at least fine tune a starting point; optimization with polynomial matrices (e.g., Vandermonde matrices) is ill-conditioned. Second, polynomials are global functions. A local change requires changes in several terms of a polynomial series. In section 3.2.1 we will review the conic sections and in section 3.2.2 we will review the extensions proposed to conic sections. Section 3.2.3 will summarize the Wasserman-Wolf pair of differential equations that describe an aspheric profile. We will focus our summary on the surface representation aspects of aspheric surfaces. Other aspects of aspheric surfaces such as raytracing through aspheric surfaces can be found in [179].

In spite of fabrication, metrology, and alignment challenges, aspheres are becoming increasingly common in the design of optical systems, driven especially by the physical size and weight constraints in personal imaging applications such as cellular telephone cameras. Similar packaging constraints are shared in several other application domains such as head-worn displays, ophthalmic lenses, astronomy, and aerospace applications that drive the use of rotationally symmetric as well as rotationally non-symmetric aspheres in those applications. Fabrication techniques for free-form surfaces include diamond turning (slow and fast tool servos), flycutting and milling. Metrology techniques for free-form surfaces include profilometry and interferometry. Fiducial marks are sometimes employed during assembly for alignment to a common coordinate system.

In terms of aberration correction, aspheres do not correct for Petzval curvature, axial color, secondary color or spherochromatism. Aspheres can help correct spherical aberration, coma, astigmatism, and distortion. Shannon provides the aspheric contributions to the Seidel terms [173]. These coefficients are found by adding linearly the fourth order aspheric shape to the aberration coefficients of the base sphere. Therefore, the fourth order aspheric can be thought of directly as an aberration coefficient [173]. If the aspheric surface is placed at the stop, the aspheric surface will help correct spherical aberration without modifying the field dependent aberrations within the 3rd order approximation. If the aspheric surface is used away from the stop, the aspheric surface will help correct field dependent aberrations except the field curvature. The higher order rotationally symmetric polynomial coefficients help control the behavior of rays passing through the top or bottom edges of the aperture. The lower order rotationally symmetric polynomial coefficients help control the ray behavior in the central portion of a system.

3.2.1 Classical Aspheric Surfaces: Conic Sections and Polynomials

Conic sections arise from the intersection of a plane with a cone as illustrated in Fig. 3.8. Conic sections have been used in many designs to represent rotationally symmetric optical surfaces. Starting with the equation of a general quadric and solving for z using the quadratic formula yields the two intersections of a sphere, if we take the solution with the negative sign,

closer to the vertex, and rationalizing the expression with the square root, we obtain the equation

of sag as follows [97]

$$z = \frac{cr^2}{1 + \sqrt{1 - (1+k)c^2 r^2}} \tag{3.21}$$

where $r^2 = x^2 + y^2$ is the distance from the axis, c is the curvature of the surface, k is the conic

constant of the surface and $k = -e^2$, e is the eccentricity. Specifically, k=0 yields a sphere, -

1<k<0 is a prolate ellipsoid, k=-1 is a paraboloid, k<-1 is a hyperboloid, and k>0 is an oblate

ellipsoid.
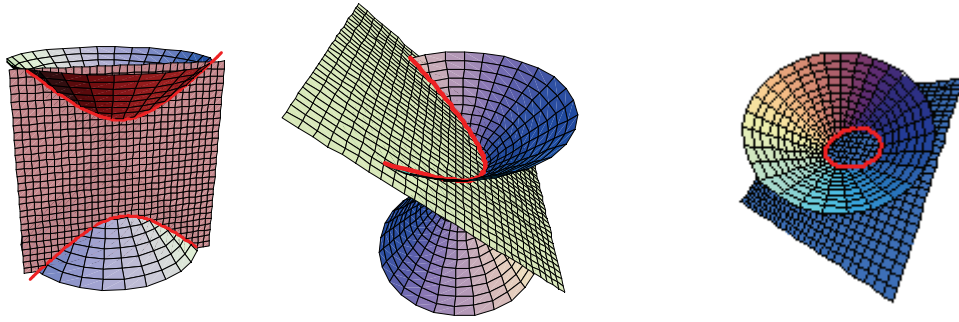
Figure 3.8 Conic sections generated in Mathematica. (Left) Hyperbola (Center) Parabola

(Right) Ellipse

Polynomial surface deformations are sometimes added to the base conic

$$z = \frac{cr^2}{1 + \sqrt{1 - (1+k)c^2 r^2}} + Ar^4 + Br^6 + Cr^8 + Dr^{10} \tag{3.22}$$

where *A,B,C,D* are the 4th, 6th, 8th, and 10th order polynomial coefficients, respectively. One advantage of adding a polynomial to the base conic for describing an asphere is that the asphere does not change the paraxial imaging properties of the surface.

Conics play an important role in certain imaging cases as they allow axially stigmatic imaging. A parabola perfectly conjugates an object at infinity to its focal point as shown in Fig 3.9(a). An ellipse perfectly images an object at its primary focus to its secondary focal point as shown in Fig 3.9(b). A hyperbola perfectly images a point placed at a focal point to the other focal point as shown in Fig 3.9(c). Axial stigmatism achieved with conic sections can be extended by imposing the sine condition, which results in new designs such as the Wolter X-ray telescope or the Ritchey-Chretien.

(a)   **(b)**   **(c)**
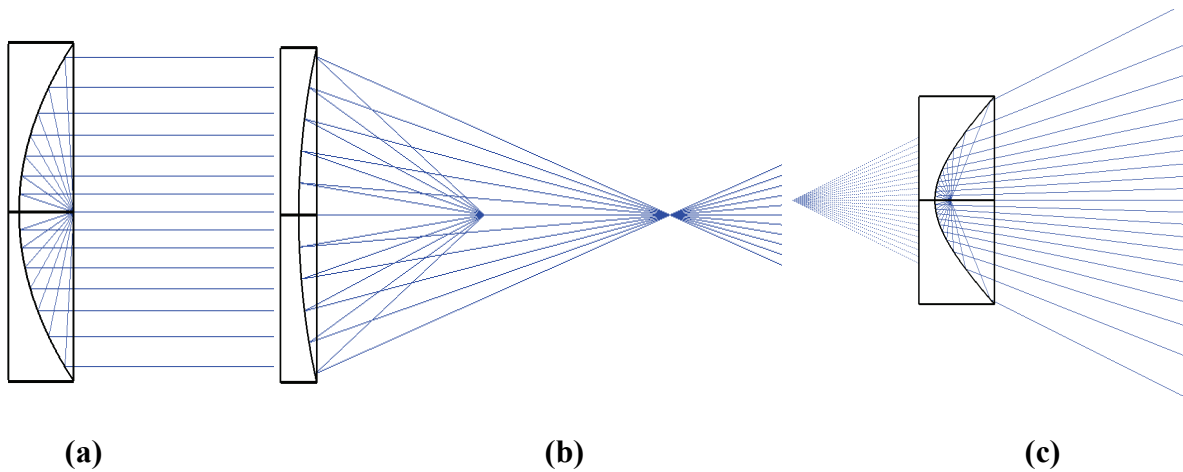
Figure 3.9 Illustration of stigmatic surfaces. (a) Parabola (b) Ellipse (c) Hyperbola.

3.2.2 Extensions to the Classical Aspheres

Most aspheric surfaces are surfaces of revolution of a base curve about the optical axis, for example, toroidal surfaces. A toric surface can be generated by rotating a circle about an axis. The axis of rotation lies in the same plane as the circle. We need the exclude the case where the axis passes through the center of the circle because we would get a sphere in that case. The parametric equations for the toric are given to be [212]

$$
\begin{aligned}
x &= (r_x - r_y + r_y \cos\theta)\sin\phi \\
y &= r_y \sin\theta \\
z &= rx - (r_x - r_y + r_y \cos\theta)\cos\phi
\end{aligned}
\tag{3.23}
$$
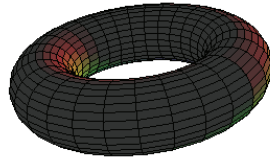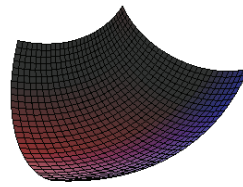
(a)   (b)

Figure 3.10 (a) 3D Plot of a ring toroid with r_x=4 and r_y=1. (b) Typical imaging applications would use portions of such a surface constants r_x and r_y determined by the application needs. Example shown with $r_x = 3, r_y = 1, \{\phi \mid 0 \le \phi \le 1\}, \{\theta \mid 0 \le \theta \le 2\}$.

Cylindrical surfaces are a special case of toroidal surfaces with $r_x \to \infty$. Toric and cylindrical surfaces find applications in astigmatism correction as they are able to provide two different curvatures in two different directions.

Historically, Schwarzschild (1905) designed telescopes composed of two aspheric mirrors to achieve aplanatism with a useful field of about 3° operating at f/3 [169]. In 1930 Schmidt designed the Schmidt camera that consisted of a spherical mirror with an aspheric plate at the center of curvature [166]. Schmidt camera allowed the imaging of larger fields of view of about 16° operating at f/1.75.

Wasserman-Wolf (1949) is a non-iterative technique generating a pair of differential equations for stigmatic imaging of a point object [207]. Coma correction is obtained by satisfying the sine condition. Essentially, Wasserman-Wolf generates a slope field representing the tangents on each point of each aspheric surface, the slope field is integrated using Runge-Kutta or Adam's method to get the surface profile. The original paper deals with the case of rotationally symmetric aspheres. Evelyn Vaskas (1957) extended the Wasserman-Wolf technique in a way that any number of optical surfaces can reside between the two aspheric surfaces [203]. David Knapp's thesis work (2002) generalized the Wasserman-Wolf technique by removing the axial symmetry assumption in the original work [99].

Rigler and Vogl (1971) proposed cubic splines as an alternative representation in the context of optics [149]. Stacy explored the ACCOS-V implementation of the Rigler and Vogl splines as general surfaces in the design of unobscured reflective optical systems [180]. The basic aim in Stacy's work was to compare spline-based designs to the classically designed Galileo narrow angle camera. Stacy found that the spline-based designs allowed wider field angles and higher transmission than the Galileo with lower but still acceptable image quality.

Chase discussed the application of parametric curves such as Non-uniform Rational B-Splines (NURBS) to the description of rotationally symmetric asphere design [42]. Chase (2002) compared the spherical aberration correction of a NURBS surface against an even and odd polynomial in a Cassegrain system. Davenport investigated NURBS as a tool for creating an incoherent uniform circular illuminance distribution on a target plane [52]. Benefits of freeform surfaces have been summarized by Rodgers and Thompson (2004) as ``greater control of the location in the field of nodes in the aberration field, and potentially the larger number of nodes in the field'' [153].

Dave Shafer (1978) proposed designing with two-axis aspheric surfaces, which is the superposition of two separate conventional aspheres with axes that are shifted relative to one another [172]. Shafer gave the example of a parabola added to a sphere with a laterally shifted axis. This can be seen as an early example of a linear combination of translated bases functions where the bases were polynomial and there were only two functions to work with.

Lawrence Mertz (1979) describes two procedures in his applied optics paper for generating aspheric profiles [122]. The first procedure corrects spherical aberration. A spherical primary reflects rays onto an aspheric secondary. The idea is to generate an osculating ellipse at each point along the aspheric secondary. The foci of the osculating ellipse are set by the location where the ray reflected from the primary crosses the optical axis. An ellipse will perfectly image rays from one of its foci to the other, and since the aspheric has an osculating ellipse for each ray height traced through the primary, procedure 1 of the Mertz paper will yield a stigmatic image. Procedure 2 of the Mertz paper discusses coma correction by satisfying the sine condition.

J. Michael Rodgers (1984) considered alternative aspheric representations to the standard method of adding a power series to a base conic [152]. Rodgers' thesis discussed nonpolynomial

basis functions added to a conic in order to yield perfect axial imagery. Examples of nonstandard aspheric functions are given in Table 2.1 of the Rodgers' thesis to be hyperbolic cosine, logarithm, secant, inverse sine, tangent and a Gaussian. Convergence properties of such alternative representations were studied. Rodgers' dissertation considered rotationally symmetric systems only. Rodgers' thesis studied global and compact representations for optical surfaces. Rodgers' thesis includes noniterative as well as iterative approaches to generating aspheric profiles. The noniterative surface generation technique was based on Wasserman-Wolf, which is summarized in Section 3.2.3. Imaging and energy redistribution examples of 2, 3, and 4 mirror reflective designs were given.

Greynolds implemented a superconic in 1986 and a subconic surface as an alternative aspheric representation. Superconics existed under the name of superquadratics proposed by Alan H. Barr since 1981 [10]. Barr defined superquadrics through a spherical product operator [10].

Scott Lerner (2003) introduced a novel explicit superconic surface [106]. Lerner also explored parametrically defined optical surfaces and implicitly defined optical surfaces. A truncated parametric Taylor surface and an xyz-polynomial were shown to be general surface descriptions. Representations were compared for ray tracing speed, optimization complexity, the ability to correct highly aspheric wavefronts, and the ability to represent steeply sloped surfaces. Specifically, Table 5.2 in [106] lists a superconic, explicit superconic, truncated parametric Taylor, and an xyz polynomial as the surfaces that were compared.

Greg Forbes (2007) recently proposed a sum of Jacobi polynomials to represent axisymmetrical aspheres [65]. Forbes emphasizes the use of Jacobi polynomials that are global and orthogonal. Forbes representation has the key property that the mean square slope of the

normal departure from a best-fit sphere is related to the sum of squares of individual coefficients of Jacobi polynomials. This property facilitates the enforcement of fabrication constraints.

Hicks formulates the free-form mirror design problem using tools from partial differential equations and vector fields [83].

3.2.3 Summary of the Wasserman-Wolf Aspheric Profile Design Equations
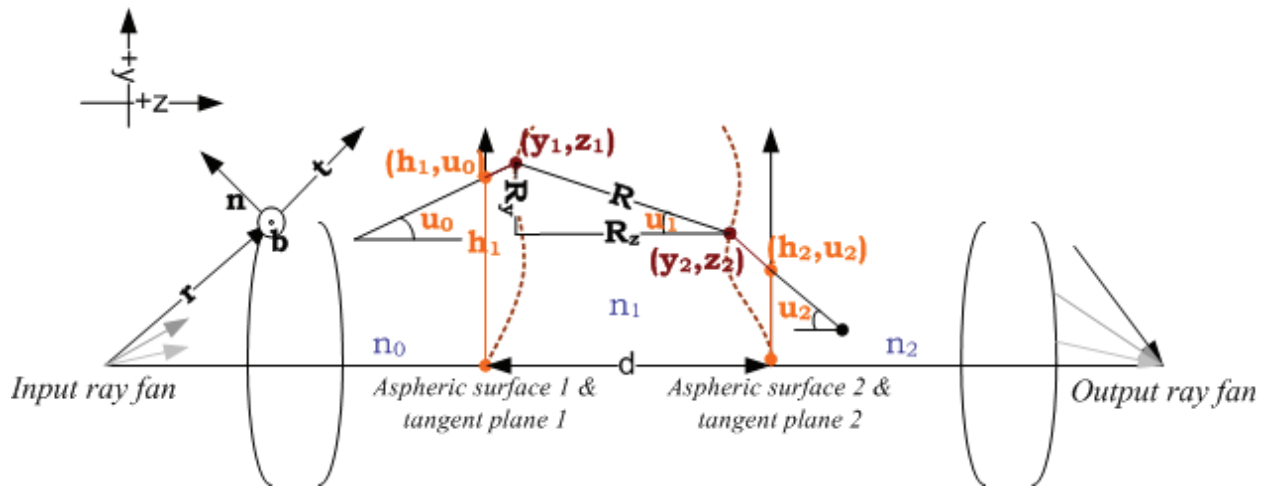
Figure 3.11 Geometry of the Wasserman-Wolf differential equation pair.

The idea in Wasserman-Wolf equations is to describe an aspheric profile pair that connects a given input ray fan to a given output ray fan [207]. The two aspheric profile vertices are separated by a distance d. From left to right, the refractive index before the aspheric surface is n_0, the aspheric system has the refractive index of n_1, and the refractive index after the aspheric surface is n_2.

The connection between the input and the output is established in two steps. First, the input ray is traced up to the tangent plane of the first aspheric surface. Knowing the height and the angle at the tangent plane of the first aspheric surface, the refraction of the ray into the asphere can be described using Snell's law. Any number of optical surfaces can reside between the input and the tangent plane of the first asphere. Second, through a backward raytrace, the height and angle of the ray at the tangent plane of the second aspheric surface can be determined. Similar to the forward raytrace, any number of surfaces can reside between the output and the tangent plane of the second asphere. The refraction of the ray at the tangent plane of the second asphere is described using Snell's law as well. From the geometry in Fig 3.11., Wasserman-Wolf defines the following quantities:

$$
\begin{aligned}
R_y &= y_1 - y_2 \\
R_z &= z_2 - z_1 \\
R &= \sqrt{R_y^2 + R_z^2} \\
\sin u_1 &= \frac{R_y}{R}; \cos u_1 = \frac{R_z}{R} \\
y_1 &= h_1 + z_1 \tan u_0
\end{aligned}
\tag{3.24}
$$

At the point of intersection of a ray with a surface, we get a local coordinate system, similar to a Frenet-Serret frame. This coordinate system has the axis along the gradient of the curve at the point (tangent vector, $\vec{t}$), the vector perpendicular to the tangent vector in the same plane (normal vector, $\vec{n}$), and the binormal vector ($\vec{b}$), which is the cross product of $\vec{t}$ and $\vec{n}$. Recall that the $\vec{t}$, $\vec{n}$, $\vec{b}$ vectors are unit vectors. Vector form of Snell's law states that

$$
n_1 (\vec{r_1} x \vec{n}) = n_2 (\vec{r_2} x \vec{n})
\tag{3.25}
$$

84

In the *y-z* plane, a ray has two components $\vec{r} = (r_y, r_z)$. The tangent vector for a surface represented parametrically and parametrized as a function of t has a tangent vector $\vec{t} = (\frac{dy}{dt}, \frac{dz}{dt})$.

Using the cross product relation between the $\vec{t}$, $\vec{n}$, $\vec{b}$ vectors in Snell's law and expanding the resultant vector triple product, recalling the perpendicularity relation between $\vec{t}$, $\vec{n}$, $\vec{b}$ vectors that cancels two of the dot products, we can rewrite Snell's law applied to the refraction at the first aspheric surface as

$$n_0 \left(r_{1y} \frac{dy}{dt} + r_{1z} \frac{dz}{dt} \right) = n_1 \left(r_{2y} \frac{dy}{dt} + r_{2z} \frac{dz}{dt} \right) \tag{3.26}$$

Recalling the definition of *y*, taking its derivative using chain rule and solving equation 3.26 for $\frac{dz}{dt}$ yields the Wasserman-Wolf pair of differential equations describing the aspheric sag for each surface

$$\frac{dz}{dt} = \left(\frac{dh}{dt} + z \frac{d \tan u}{dt} \right) \left[\frac{n_1 R \cos u_1 - n_0 \cos u_0}{n_0 \sin u_0 - n_1 R \sin u_0} - \tan u_0 \right]^{-1} \tag{3.27}$$

This pair of differential equations can be solved using Runge-Kutta or Adam's method as suggested in the original paper. The quantities $\frac{dh}{dt}, \frac{d \tan u}{dt}$ can be computed via numerical derivatives or otherwise. Once a sag table is generated, a spline or polynomial fit can be performed to the sag points to generate the final surface.

3.3 Nodal Aberration Theory (Vector Aberration Theory)

In this section, we will briefly review the definition of aberrations in rotationally symmetric systems following the Hopkins treatment [85]. We will summarize the work done by Shack and Hopkins on rewriting the wavefront aberration function in vectorial form. Introduction of the σ vector by Buchroeder [21] and the first-order theory by Shack and Thompson [194] allows us to calculate the center of the aberration field in tilted and decentered systems.
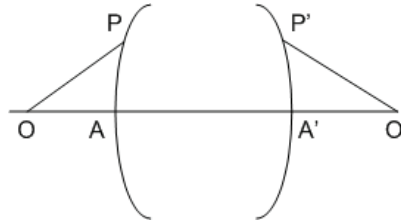
Figure 3.12 Geometry of paraxial imaging of a point through an optical system (adapted from [85]).

The optical system shown in Fig 3.12 is forming an image of a point O at O'. The first surface of the system is AP and the last surface is A'P'. OP is the ray entering the system and after propagating through the system, the ray OP emerges as P'O'. If OP and P'O' is sufficiently close to the optical axis OAA'O' then O' is said to be a paraxial image of the point O. The wavefront emerging at P'A' will be perpendicular to the rays P'O' and A'O', therefore, in the

paraxial approximation the emerging wavefront will be a sphere centered on O'. Difference in the optical paths from this spherical reference sphere is defined to be the aberration of the system as discussed in detail below.
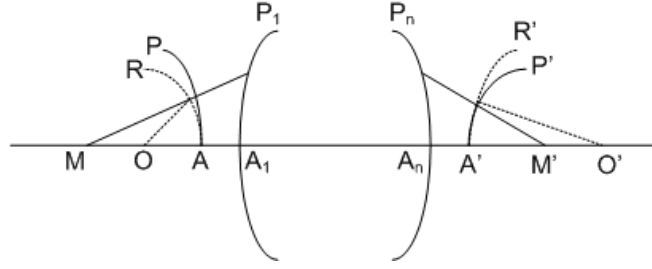
Figure 3.13 Geometry of aberrated image of a point (adapted from Hopkins [85])

Let $OAA_1..A_nA'O'$ and $MRPA_1..A_nR'P'M'$ be two rays passing through the system shown in Fig 3.13. O and O' are the centers of the reference spheres AR and A'R'. AP is the incident wavefront and A'P' is the emergent aberrated wavefront. The time it takes to travel between the wavefront AP and the wavefront A'P' is the same for each ray. Therefore, we can write the following optical path length relation between the two rays in the object space wavefront to two rays in an emergent wavefront as

$$[AA_1A_n...A']=[PP_1...P_nP'] \qquad (3.28)$$

Equivalently we can rewrite equation 3.28 as

$$-[RP]+[RPP_1..P_nR']+[R'P']= [AA_1A_n...A'] \qquad (3.29)$$

87

Let W = [RP], W'=[R'P'] and rearranging yields

$$W' - W = [AA_1 \ldots A_k A'] - [RP_1 .. P_k R'] \tag{3.30}$$

The wavefront aberration at point P and P' due to the system is defined to be the optical path difference along the optical axis ray and a given ray between the chosen reference spheres [85]. In a rotationally symmetric system, Hamilton chose the parameters of the wavefront aberration to be H^2, ρ and $H\rho\cos(\phi)$. Shack rewrote these parameters vectorially with a geometry shown in Fig. 3.14.
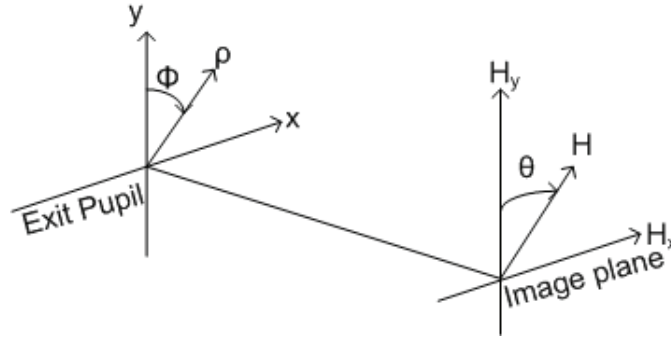
Figure 3.14 Coordinate systems for the field and aperture vector (adapted from [193])

Rewriting these parameters in vectorial form and Taylor expanding the wavefront yields

$$W(\vec{H}, \vec{\rho}) = \sum_{k,n,m} c_{2k+m,2n+m,m} \cdot (\vec{H} \cdot \vec{H})^k \cdot (\vec{\rho} \cdot \vec{\rho})^n \cdot (\vec{H} \cdot \vec{\rho})^m \tag{3.31}$$

To 3rd order equation 3.31 can be written as

$$W(\vec{H},\vec{\rho}) = \sum_j \left[c_{040j}(\vec{\rho}\cdot\vec{\rho})^2 + c_{131j}(\vec{H}\cdot\vec{\rho})(\vec{\rho}\cdot\vec{\rho}) + c_{222j}(\vec{H}\cdot\vec{\rho})(\vec{H}\cdot\vec{\rho}) + c_{220j}(\vec{H}\cdot\vec{H})(\vec{\rho}\cdot\vec{\rho}) + c_{311j}(\vec{H}\cdot\vec{H})(\vec{H}\cdot\vec{\rho}) \right] \quad (3.32)$$

The equation 3.32 assumes that there is an optical axis denoting rotational symmetry and connects the centers of the object, vertices of the elements, and the pupils. In the case of tilted and decentered systems the rotational symmetry is broken and the aberration field centers for each surface no longer coincide.

Nodal aberration theory introduces the sigma vector that describes the shift in the aberration field center caused by the tilts and decenters in a system. In a perturbed system, the aberration field is still the sum of individual surface contributions. The sigma vector is essentially the line joining the center of the pupil with the center of curvature for that surface, projected out to the object or image plane. The sigma vector is calculated for each surface in the system. The sigma vector modifies the field dependence in equation 3.32. The geometry for the sigma vector shifting the aberration field center is shown in Fig 3.15. Sigma vectors for rotationally symmetric aspheric (conic sections with an aspheric vertex) surfaces can be calculated in nodal aberration theory. A surface with an asphere would act as two surfaces since we would have a sigma vector for the spherical surface and another σ vector for the aspheric surface. Handling of free-form aspheric surfaces is ambiguous due to the lack of an aspheric vertex.

$$\vec{H} = \vec{H}_o - \vec{\sigma} \quad\quad\quad (3.33)$$
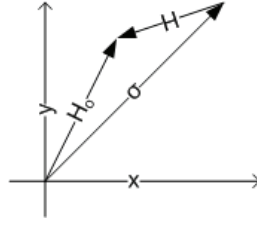
Figure 3.15 Sigma vector.

Vectorial the rotationally symmetric wavefront expansion is modified with the sigma vector and Thompson describes each component of the expansion for systems with tilted and decentered components.

Spherical aberration is constant across the field; therefore, it has no field dependence and is not modified by the sigma vector.

Coma in a tilted and decentered system is described as follows

$$W_{coma} = \sum_j c_{131j} \left[(\vec{H} - \vec{\sigma}_j) \cdot \vec{\rho} \right] (\vec{\rho} \cdot \vec{\rho}) \qquad (3.34)$$

Setting equation 3.34 to zero and solving for H locates the coma node for a tilted and decentered system. Tilted and decentered systems may exhibit on-axis coma, according to nodal aberration theory, there will still be a (node) point in the field that has no coma.

Astigmatism referenced to the medial surface in a system with tilted and decentered components is described as

$$W_{astigmatism} = \frac{1}{2} \sum_j W_{222j} \left[(\vec{H} - \vec{\sigma}_j) \cdot \vec{\rho} \right]^2 \qquad (3.35)$$

90

Setting equation 3.35 to zero and solving for H using the quadratic formula yields the *binodal*

astigmatism behavior. A graphical illustration of the binodal astigmatism is given in Fig 3.16.

Field curvature is described by

$$W_{\text{field curvature}} = \sum_j W_{220_j} \left[\left(\vec{H} - \vec{\sigma}_j \right) \cdot \left(\vec{H} - \vec{\sigma}_j \right) \right] \left(\vec{\rho} \cdot \vec{\rho} \right) \tag{3.36}$$

Distortion is described by

$$W_{distortion} = \sum_j W_{311_j} \left[\left(\vec{H} - \vec{\sigma}_j \right) \left(\vec{H} - \vec{\sigma}_j \right) \right] \left[\left(\vec{H} - \vec{\sigma}_j \right) \cdot \rho \right] \tag{3.37}$$
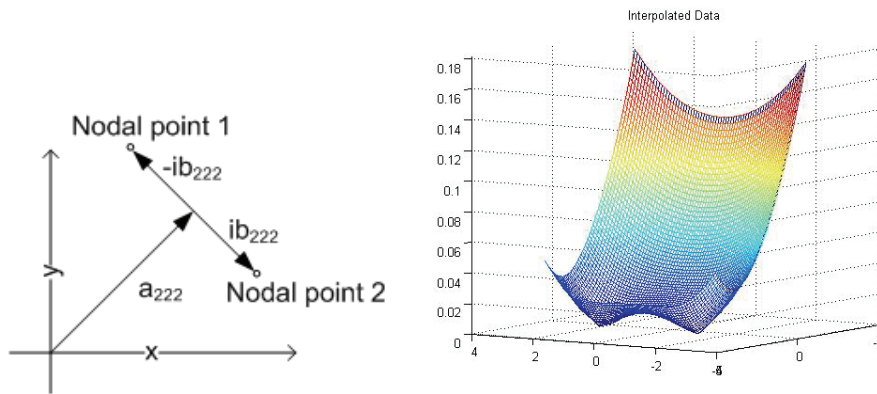
Figure 3.16 (a) Binodal astigmatism in a tilted and decentered system.(b) Magnitude of binodal

astigmatism across the field

Off-axis rotationally non-symmetric systems have been used in head-worn displays, laser beam expanders, unobscured telescopes and remote sensing applications. One key insight provided by nodal aberration from the point of view of optical system design is that nodal aberration theory allows a designer to transfer their knowledge from rotationally symmetric systems over to tilted and decentered systems.

One of the visualization tools provided by nodal aberration theory is the full field displays. Full field displays are based on real raytrace data. The full field display plots provide magnitude and orientation information. The length or circle size is proportional to the magnitude of the aberration of interest in a full field plot. In a system with pure astigmatism, the full field display for astigmatism aligns with the spot diagram. Full field displays are generated by tracing rays, calculating the OPD for each ray with respect to the chief ray and then fitting Zernike polynomials to the OPD map.

Thompson applies full field displays to the analysis of a triplet in Chapter 7 of his thesis [194]. The key quantity is emphasized to be the σW product. σW and $\sigma^2 W$ products appear when we multiply out the terms in equations 3.34, 3.35, 3.36 and 2.37. Large values of σ identify the dominant elements in the perturbation analysis. We remind that based on first-order theory, the σ vector is a function of the vertex displacement, tilt angle and the pupil location in the space of the element. Therefore, the magnitude of the σ vector can be controlled by changing the vertex displacements, tilt angles and the pupil location.

John Figoski [62] applied nodal aberration theory to find an alternate solution to a two mirror telescope problem. Figoski starts his design process with a centered 20 degrees field of view, f/4 reflecting inverse telephoto. The first step is to bias the aperture +1 unit. Figoski compares the centered design with the bias aperture across a 30x30 degrees field of view and

notes the bias aperture has gotten worse across the field. There is only one astigmatism node visible in the 30x30 degrees field map. The next step is to add a +20 degrees field bias and then optimize the design. The binodal behavior is recovered within the 30x30 degrees astigmatism field map after optimization. Figoski uses the field maps to provide insight into a second - 20 degrees bias solution that outperforms the +20 degrees bias solution.

John Rogers provides geometric interpretations of aberration correction in nodal aberration theory [154]. John Rogers applied the theory to illustrate how to collapse the binodal astigmatism into a symmetric field, restoring the field behavior in a tilted and decentered system similar to that of a rotationally symmetric system. Rogers restores symmetry in the field plots by targeted Zernike optimization constraints of symmetric astigmatism magnitude around the field center. The coma node is kept as close as possible to the location of the astigmatism nodes. The advantage of this approach is that further corrections can be achieved using rotationally symmetric field elements. It is noted that simultaneous correction of astigmatism and coma can be achieved when the untilted starting point system is monocentric.

Tatiana Jewell applied nodal aberration theory to the analysis of a plane parallel plate, prism, wedge, mirror, and a thin-lens [92].

3.4 Optimization Techniques in Optical Design

This section will focus on summarizing four main ideas from optimization theory: gradient descent, locally linear approximation of the merit function, damped least squares and generalized Lagrange multipliers, which is the Karush-Kuhn-Tucker system. The first two ideas, gradient descent and the locally linear approximation of the merit function, combine together in

damped least squares, which is the standard technique for optimization in raytrace codes today. Optical design problems include constraints, which are typically treated within the framework of Lagrange multipliers. A complete introduction to optimization techniques, compared to what is provided in this section, can be found in [69] and [64].

Usually the number of variables in a system is smaller than the number of problems we are trying to minimize and the constraints we are trying to fullfill, therefore, only an approximate solution may be found. Example constraints include the focal length, real ray based constraints, glass boundaries, packaging, weight and size constraints.

There is no guarantee using any technique that the solution found by means of optimization will be a global minimum. Each solution found through optimization will be sensitive to the starting point. Different optimization algorithms will typically yield different results even under the same starting point. In practice, each solution will need to be evaluated by an optical designer in many ways. Two examples of such an evaluation are tolerance analysis, and process specific fabrication feasibility.

Optical designers frequently make use of optimization algorithms to either improve starting points chosen by designer or to explore new solutions from scratch to optical design problems. An overview of methods in optical system design is provided by Hazra and Chatterjee [79] and shown in Fig 3.17. In this section, we summarize the "calculus based" techniques, which are based on derivatives, listed under the right branch in the classification tree given in Fig. 3.17.

Optimization aims to find the local minimum to the unconstrained optimization problem

$$\text{Minimize } f(\mathbf{x}), \; x \in \Re^n \tag{3.38}$$

where f(x) is referred to as the objective or the merit function. In the context of optical design, typical merit functions have several dimensions, and they are nonlinear. A simple approach we can apply to minimize the unconstrained optimization problem is gradient descent. Gradient descent requires numerical or analytical calculation of the derivatives of the function to be minimized. The gradient vector of a function points in the direction of maximum change. The idea in gradient descent is to follow some negative multiple of the gradient direction with the goal of reaching local minima.
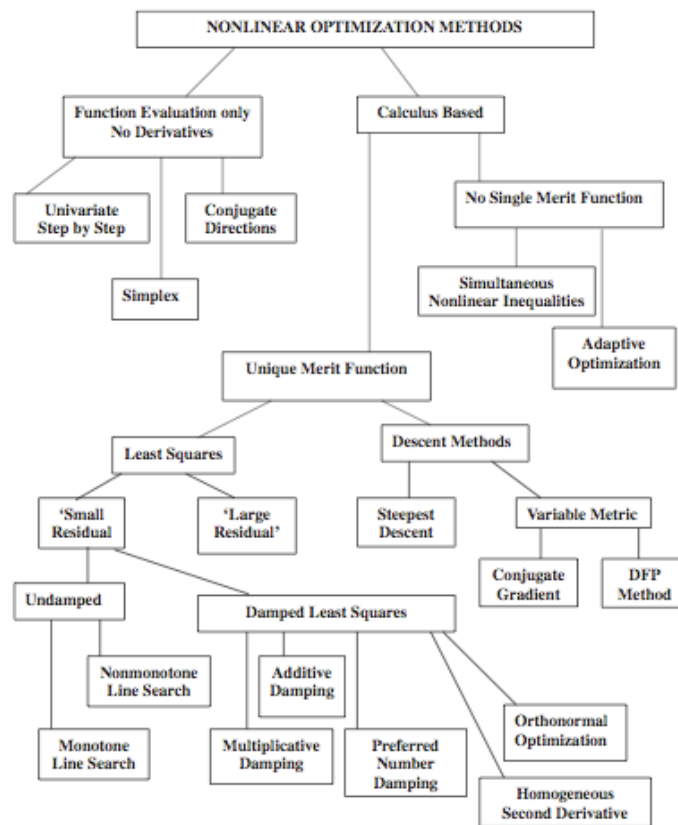
Figure 3.17 Overview of optimization algorithms in optical system design (adapted from [79]).

The gradient descent equation is

$$P_1 = P_o - \alpha \nabla \mathbf{f} \qquad (3.39)$$

where P_1 denotes the next step and P_0 denotes the starting point or the current position, α is a scalar that scales the length of our step size, ∇ denotes the gradient of a function that yields a vector as its output. Gradient descent will iterate until some convergence criteria is met.

(a)   (b)
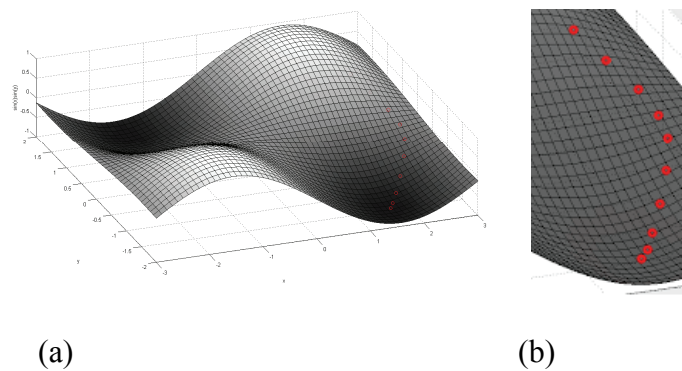
Figure 3.18 Illustration of gradient descent (α=0.5 in this example) (a) Plot of the y=sin(x)cos(x) function. The starting point and the descent path is marked with red circles. (b) Close up of the descent path.

Transverse error calculated on the image plane for typical imaging problems is a sum of squares type merit function. Therefore, let us assume that M(x) is the merit function we wish to minimize and is a sum of squares type merit function

96

$$M(\mathbf{x}) = \sum_{i=0}^{m} f_i^2 = f^T f \qquad (3.40)$$

where f is a particular aberration function and depends on the variables in the system, e.g., $f(x_1, x_2, x_3,..)$. Most optimization algorithms will assume a model that is a local approximation of the merit function. For example, starting with Taylor's theorem for two independent variables, and two aberration functions, and expanding the function with a linear model we have

$$f_1(x_1 + \Delta x_1, x_2 + \Delta x_2) = f_1(x_1, x_2) + \{\frac{df_1}{dx_1} \Delta x_1 + \frac{df_1}{dx_2} \Delta x_2\}$$
$$f_2(x_1 + \Delta x_1, x_2 + \Delta x_2) = f_2(x_1, x_2) + \{\frac{df_2}{dx_1} \Delta x_1 + \frac{df_2}{dx_2} \Delta x_2\} \qquad (3.41)$$

Rewriting in matrix form yields

$$f = f_o + \begin{bmatrix} \dfrac{df_1}{dx_1} & \dfrac{df_1}{dx_2} \\ \dfrac{df_2}{dx_1} & \dfrac{df_2}{dx_2} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = f_o + Js \qquad (3.42)$$

where $\mathbf{J}$ is the Jacobian matrix of partial derivatives and s is a vector. The gradient of the sum of squares merit function, using the chain rule and expressed in matrix form can be written as

$$g \equiv \frac{1}{2} \frac{dM(\mathbf{x})}{dx_i} = f_1 \frac{df_1}{dx_1} + f_2 \frac{df_2}{dx_1} + f_1 \frac{df_1}{dx_2} + f_2 \frac{df_2}{dx_2} Js \rightarrow g = J^T f = J^T (f_o + Js) = J^T f_o + J^T Js \quad (3.43)$$

Setting the gradient equal to zero and solving for s yields the update rule in a locally linear model

$$s = -(J^T J)^{-1} J^T f_o \qquad (3.44)$$

The main message we want to convey is that derivative based optimization techniques combined with a linear approximation to the merit function result in solving a least squares problem at each step of the iteration process.

Singular value decomposition (SVD) provides an alternative way of performing least squares. SVD of any matrix is written as

$$\mathbf{A} = \mathbf{U} \, \boldsymbol{\Sigma} \, \mathbf{V}^{\mathrm{T}} \tag{3.45}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices. $\boldsymbol{\Sigma}$ is a diagonal matrix with the singular values along the diagonal, the singular values are not the same as the eigenvalues. Geometrically, in the case of a quadratic form, $\mathbf{U}$ and $\mathbf{V}$ can be thought of as rotation matrices and $\boldsymbol{\Sigma}$ will stretch the axis to turn a circle into an ellipse.

Damped least squares is a popular algorithm in the design of imaging systems and will be summarized. Damped least squares combines the gradient descent idea with the locally linear approximation of the merit function in the following equation

$$\boldsymbol{s} = -\left(\boldsymbol{J}^{T}\boldsymbol{J} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{J}^{T} \boldsymbol{f} \tag{3.46}$$

Setting $\lambda \to 0$ in equation 3.46 yields equation 3.44. Setting λ to a large number yields a diagonally dominant matrix. Inverse of a diagonally dominant matrix is one over its diagonal elements. We can think of the term multiplying $J^{T}f$ as a scalar, which is identical to the α parameter in gradient descent discussed in equation 3.39. Therefore, setting λ to a large number

yields gradient descent. We can now understand how the parameter λ combines two ideas: gradient descent and the locally linear approximation. Users of damped least squares would need to decide how to choose or modify the λ. One potential strategy can be to monitor the merit function and increase or decrease the λ based on the changes in the merit function. For example, when closer to local minima we would switch to gradient descent, further away from local minima we can use the locally linear approximation. This is the damped least squares idea.

Optimization with constraints is treated in the framework of Lagrange multipliers. In optical design, the optimization problem is changed to

$$\text{Minimize } f(\mathbf{x}), \ x \in \mathfrak{R}^n \text{ subject to constraints} \tag{3.47}$$

One way to understand the Lagrange multiplier framework is through weighted least squares that extends ordinary least squares with a weighting matrix [184]. Weighted least squares aims to solve the problem

$$\boldsymbol{WAx} = \boldsymbol{Wb} \tag{3.48}$$

where $\mathbf{W}$ is the weighting matrix.

The normal equations for weighed least squares become

$$\boldsymbol{A^T CAx} = \boldsymbol{A^T Cb} \tag{3.49}$$

where $C = W^T W$.

The Karush-Kuhn-Tucker (KKT) equations for the weighted least squares problem become

$$
\begin{aligned}
e \quad + Au &= b \\
A^T Ce \quad &= 0
\end{aligned}
\quad \text{or in matrix form} \quad
\begin{bmatrix} I & A \\ A^T C & 0 \end{bmatrix}
\begin{bmatrix} e \\ u \end{bmatrix} =
\begin{bmatrix} b \\ 0 \end{bmatrix}
\tag{3.50}
$$

The first equation means that we are splitting a vector into its column space and null space projections. The first equation represents the primal problem. Column space is the subspace formed by all of the linear combinations of the columns of a matrix A. Null space contains the vectors perpendicular to the column space. In the case of weighted least squares, perpendicularity means a weighted inner product and vectors are no longer perpendicular with 90 degrees angles as they were in the case of ordinary least squares. The second equation expresses this new sense of perpendicularity. The second equation is the normal equation in the case of weighted least squares and it means that the columns in the null space of A are perpendicular to the column space in the weighted inner product sense. The second equation represents the dual problem. Both equations are combined in block matrix form. Symmetry in this matrix can be restored by introducing the variable $w = Ce$ and rewriting the block system as

$$
\begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix}
\begin{bmatrix} w \\ u \end{bmatrix} =
\begin{bmatrix} b \\ 0 \end{bmatrix}
\tag{3.51}
$$

In the weighted least squares example we have a homogeneous system of equations, i.e. $A^T w = 0$, which can be viewed as the constraint in that problem. However, in the most general form, we can have a non-homogeneous set of linear equations as the constraints, which would be $A^T w = f$. Therefore, the most general case of the KKT system for optimizing with constraints become [184]

$$\begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} b \\ f \end{bmatrix} \tag{3.52}$$

CHAPTER 4 SINGLE AND DUAL-ELEMENT OFF-AXIS MAGNIFIER DESIGN

The review of existing approaches to the optical design of head-worn displays provided in Chapter 2 indicates a research opportunity in applying free-form optics to this problem, which is explored in this Chapter. Specifically, we design and analyze one or two element conventional magnifiers using free-form optical surfaces described with polynomials as a function of two variables, and study the eyebox diameter, field of view and performance limits of these magnifiers. We are using the term free-form in reference to surfaces that are rotationally non-symmetric; examples of free-form surfaces include x-y polynomials and Zernike polynomials. The key benefits of single or dual-element free-form magnifier design approach are robustness, compactness, ease of manufacture compared to alternative approaches, conduciveness to mass-manufacture, low-cost, and light-weight.

The design forms presented in this Chapter are off-axis magnifiers that fold the optical axis around a human head while providing the necessary mechanical clearances. The symmetry around the ray connecting the centers of the image, object, pupils, and vertices of the elements is broken because of the asymmetries in the free-form mirror classifying these systems as off-axis. It is necessary to study the minimum fold angle in order to keep the incidence angles of the rays on the mirror as small as possible. In the case of a magnifier composed of the minimum number of optical elements, exploiting the degrees of freedom in free-form surfaces, such as the local anamorphisms, leads to compact off-axis magnifier designs as presented in this Chapter. Recent

advances in diamond turning technology facilitate the fabrication of rotationally non-symmetric free-form surfaces, and examples of fabricated prototypes will be shown in this Chapter.

We would like to note that the designs in this Chapter were considered high-risk at the time they were developed and these designs were developed using seed funds. There was no funding to characterize these systems in depth aside from predictions in simulation and user experiences. The detailed prescriptions can be found in a U.S. patent application [34].

4.1 Dual-Element Design

The optical layout is shown in Fig. 4.1 (a) and consists of an 8 mm diameter eyebox, a free-form mirror described with an *x-y* polynomial, a hybrid refractive/diffractive lens, a flat fold mirror, and a transmissive VGA resolution (640x480 pixels) microdisplay. The dual-element system shown in Fig. 4.1 (a) has a 20 degree full-field diagonal field of view and a 15mm eye clearance. Table 4.1 lists the key specifications for the dual-element magnifier design.

The surface normal along the vertex of the free-form mirror is tilted at about 34 degrees in the x-z plane with respect to the optical axis of the human eye. We found this angle to be the minimum tilt angle that can provide the necessary clearances around a human head based on a database of publicly available CAD models of human heads [1]. The distance between the center of the pupil and the vertex of the free-form mirror is 20mm. The distance between the vertex of the free-form mirror and the vertex of the hybrid lens is 29.5 mm. The distance between the spherical surface of the hybrid lens and the fold mirror is 9 mm, and the distance between the fold mirror and the microdisplay is 14 mm. Typically in a head-worn display, the pupil of the

103

human eye is the aperture stop of the system; therefore, the aperture stop and the exit pupil coincide in this system. Even though magnifiers do not form exit pupils, it is customary to optimize these systems across a finite pupil size that is larger than that of the human eye to accommodate natural head and eye movements.

Table 4.1 Specification table for the dual-element eyeglass display

Parameter	Achieved parameter in the final design
Full diagonal field of view [degrees]	20
Eye box diameter [mm]	8
Effective focal length [mm]	32
Eye clearance [mm]	>15
Spectral range	Photopic visible spectrum (450-650nm)
Spectral weights based on the photopic response of the human eye	2@635nm
	5@610nm
	10@555nm
	5@510nm
	2@490nm
Image quality	>20% at 33 cycles/mm
Distortion	<4%
Microdisplay parameters	Kopin VGA 640x480 pixels 11mm (.44") diagonal (9mm*6.8mm) 14.1 microns pixel size

A photopically weighted polychromatic MTF evaluated for a centered 8 mm pupil is plotted in Fig. 4.1 (b) for the on-axis field and the field points at the corner of the display. The performance of the MTF with 3mm pupils decentered up to 6 mm in various directions displayed similar performance to the 3mm centered pupil MTF. A red, blue, and green microdisplay pixel triplet is 14.1 μm in size, which yields a Nyquist frequency of ~35 cycles/mm. Thus the polychromatic MTF was plotted up to 35 cycles/mm. In the current prototype, the microdisplay has a 0.44" diagonal and contains 640x480 pixels. In visual space, the display provides 1.5

arcminutes resolution as limited by the pixel spacing on the microdisplay. It may be noted that a

calculation of the Nyquist frequency of the eye given a 2.5 μm cone spacing of the human visual
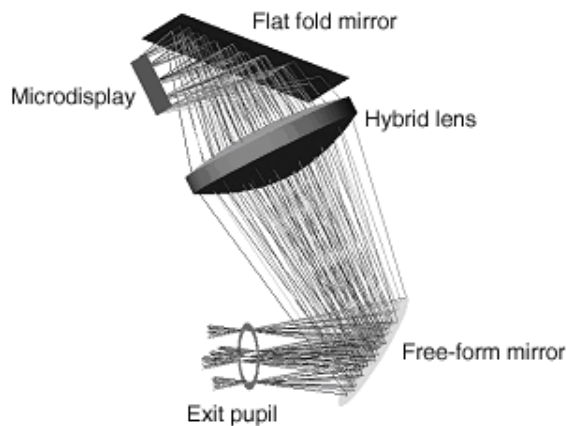
(a)
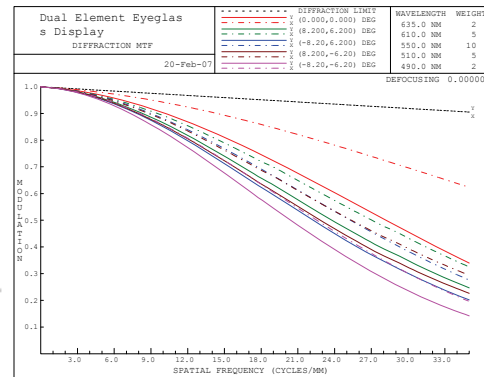
(b)
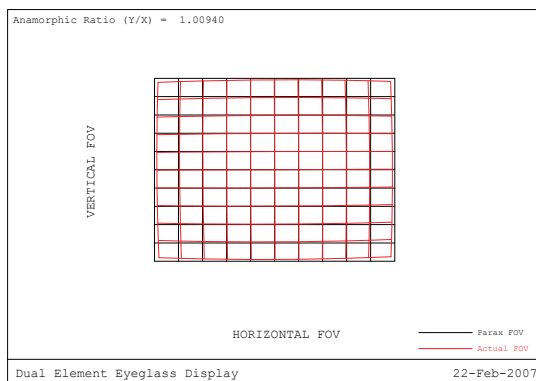
(c)
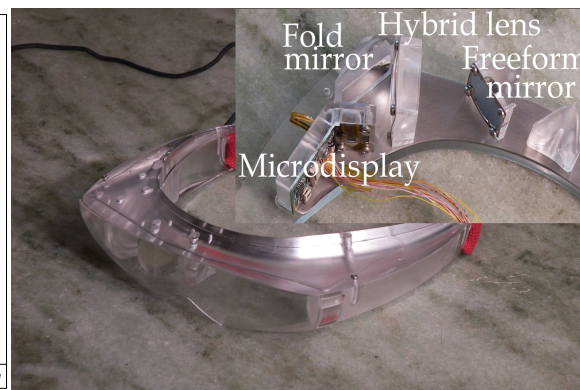
(d)

Figure 4.1 (a) Optical layout of the dual-element magnifier (b) Predicted modulation transfer

function of the dual-element solution evaluated with a 8 mm eye box (c) Distortion grid

comparing real and paraxial rays (<4% maximum distortion) (d) Photograph of the fabricated

and assembled near-eye magnifier

system yields a 1 arcminute resolution limit for the human visual system. The magnitude of maximum distortion occurs at (x=8.2°, y=-6.2°) in the field and was predicted by simulation to be -3.8%. Fig.4.1 (c) shows a plot of the real rays compared against the paraxial rays. Less than four percent maximum distortion has been predicted in simulation.

The optimization strategy was based on imposing a minimal number of constraints. Specifically, we optimized with a focal length, eye clearance and real ray based distortion constraints. The focal length was constrained to a target value of 32 mm based on a display size of .44" full diagonal and a 20 degree diagonal full field of view. The eye clearance was constrained to equal 20 mm. A set of three ray-based distortion constraints were used at the fields (x=0°,y=6.2°), (x=8.2°,y=0°), (x=-8.2°,y=0°) to constrain the image size to the size of the microdisplay. Optimization variables included the spacing between the elements, radius of the rear spherical surface on the hybrid plastic lens, radius and the aspheric coefficients on the hybrid plastic lens, rotationally symmetric diffractive phase polynomial coefficients on the hybrid lens, the x-y polynomial coefficients on the free-form mirror up to and including order 4, and the image plane tilt and defocus.

In order to assess manufacturability, the sag departure on the free-form mirror at 23 mm, the edge of the mirror where the maximum sag occurs, was computed to be 80 μm from the best fit sphere. We used a 3mm thick PMMAO material as the substrate for the free-form polynomial mirror. Plastic material was selected because plastics are lighter in weight than typical glasses. Another benefit of using plastic materials is to facilitate the fabrication of the free-form surface and the diffractive optical element. The profile of the fabricated x-y polynomial surface was described by the following polynomial

$$P(x,y) = \frac{cr^2}{1+\sqrt{1-(1+k)c^2 r^2}} + \sum_{j=2}^{10} C_j x^m y^n \text{ where } j = \frac{(m+n)^2 + m + 3n}{2} + 1 , \qquad (4.1)$$

where c is the vertex curvature, r^2 equal $x^2 + y^2$, k is the conic constant and C_j are the coefficients of the various $x^m y^n$ term. The x-y polynomial surface was fabricated with a Moore 350 UPL with c-axis (slow slide servo). The lens consists of a spherical surface and an aspheric surface with a diffractive optical element. It was found that the aspheric substrate improved the level of aberration correction. The lens was built out of Zeonex Z-ER48 material that has a 92% average transparency within 400-800 nm. Diamond turning can produce the aspheric substrate along with the diffractive optical element at no additional cost.

The diffractive was configured to operate in the +1 diffracted order. The focal length of the DOE was given by $-0.5/c_1$ where c_1 is the quadratic phase coefficient of the DOE. In our case, the DOE contributes about 5% percent of the optical power on the plastic lens. The diffractive element can be viewed as a Fresnel zone plate. It is possible to use the focal length of a Fresnel zone plate, the Abbe number, and the relation for optical power to show that the Abbe value of a diffractive optical element in the visible is about -3.5 as shown in Chapter 3. Therefore, diffractive optical elements exhibit complementary dispersion to that of optical glasses and plastics and they can be used in color correction. The diffractive optical element provides one order of magnitude better axial color correction for the system shown in Fig. 4.1 (a), predicted from the longitudinal spherical aberration plots for the two cases of with and without the diffractive optical element. The profile of the hybrid lens was defined by an aspheric component given by Z(h)

$$Z(h) = \frac{ch^2}{1+\sqrt{1-(1+k)c^2h^2}} + Ah^4 + Bh^6 + Ch^8 + Dh^{10} + Eh^{12} \qquad (4.2)$$

The coefficients *k, c, A, B, C, D, E* were initialized to be zero and were allowed to vary during optimization.

The phase function ϕ of the DOE was defined by

$$\phi = \frac{2\pi}{\lambda} \sum_{i=1}^{4} c_i r^{2i} \qquad (4.3)$$

where λ is the design wavelength (558 nm in our design), and c_i represent the polynomial coefficients describing the phase function. Coefficients c_i (i=1..4) were initialized to zero and were allowed to vary. The DOE and the aspheric surface were placed on the side towards the free-form mirror. One can place the DOE on either side of the hybrid lens and the performance is about the same for the cases of placing the DOE facing the free-form mirror versus the DOE facing the microdisplay. The diffractive element was fabricated with 198 rings and a minimum feature size of 19 microns. The step height at the design wavelength of 558 nm is 1 micron. The radius of the DOE element is 28 mm. The diffraction efficiency at the wavelength of 550 nm is predicted to be 98.7% and above 90% across 490-635 nm for 16 mask levels in the case of a lithography based process. The hybrid lens has been fabricated using diamond turning and we can expect the diffraction efficiency equivalent to at least a 16 layer process given that it is

possible to fabricate 256-level DOEs using diamond turning [107]. The hybrid lens was fabricated with a Precitech nanoform 350 2-axis diamond turning lathe.

The axial color was extracted from the longitudinal spherical aberration plot and was predicted to be approximately 20 μm (<2 pixels). The maximum lateral color was predicted to be 7.2 μm (about half a pixel) in the (x=-8.2°, y=-6.2°) field. The detailed lens prescription including the surface coefficients is provided in a U.S. Patent [34].

The assembled design is shown in Fig. 4.1 (d). A color and a black and white image formed by the dual-element off-axis magnifier are shown in Fig. 4.2 (a) and (b). In the first generation prototype, the whole assembly weighed approximately 250 grams including the outer shell, optics, optomechanical mounts, microdisplay and the microdisplay driver electronics circuit board.

(a)   (b)
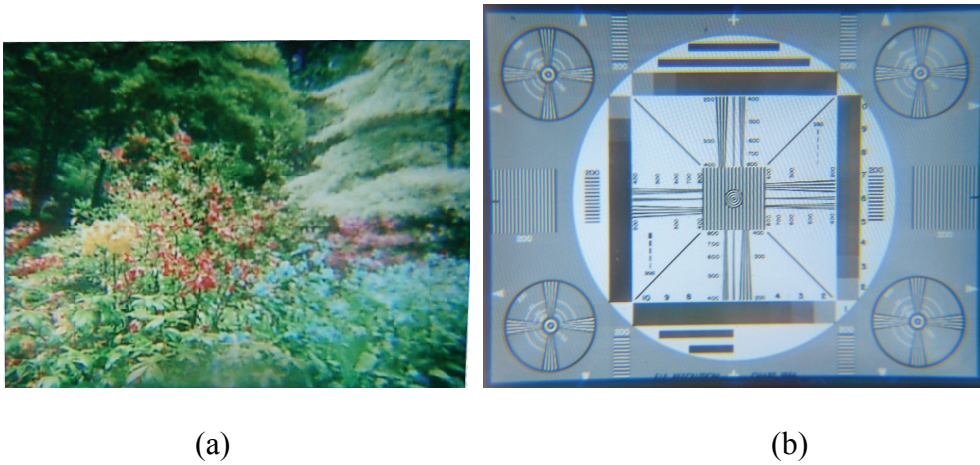
Figure 4.2 (a) Photograph through the fabricated dual-element system of a color target (b) Photograph through the fabricated dual-element system of a black and white target

The optomechanical design shown in Fig. 4.1 (d) has been redesigned to fit the eyeglass form factor. The computer aided design model for the Gen-II prototype is shown in Fig. 4.3(a) and the fabricated and assembled Gen-II prototype is shown in Fig. 4.3(b). The Gen-II eyeglass display prototype shown in Fig. 4.3(b) weighs 124 grams, a 50% improvement over the Gen-I prototype.

The use of a single free-form surface in our designs increases the feasibility of a zero power shell configuration. The see-through glass plate is shown in Fig. 4.4(a). Distortion of the scene as seen through the plate is simulated in Fig. 4.4 (b) and is found to be negligible (<1%).

(a)   (b)

Figure 4.3 (a) CAD model of the newer optomechanical design (b) Fabricated and assembled Gen-II eyeglass display
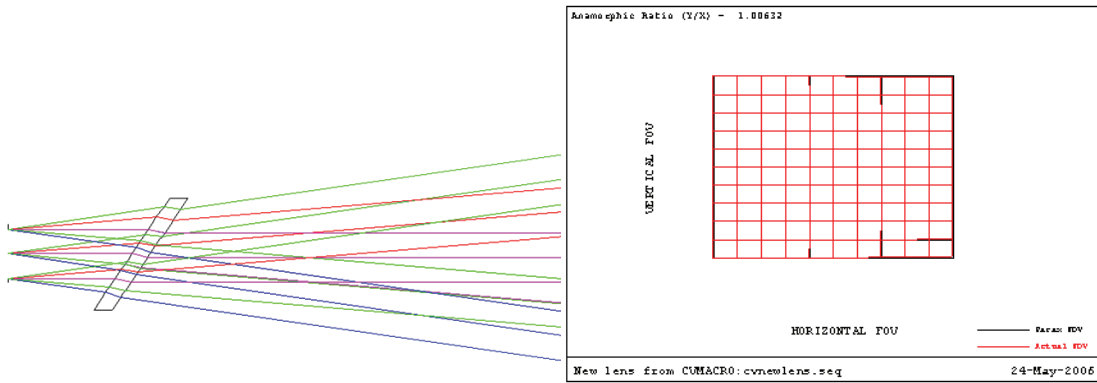
Figure 4.4 (a) Raytrace through the zero power shell configuration for see-through mode of operation (b) Predicted distortion of the scene as seen through the zero power shell configuration (<1%).

4.2 Single-Element Design

An ideal solution for an HWD would be a single surface mirror design. A single surface mirror does not have dispersion; therefore, color correction is not required. A single surface mirror can be made see-through by machining a similar surface profile on the opposite side of a substrate to form a zero power shell. From a system perspective, the optical system weight is reduced to a minimum. The layout of the single mirror magnifier is shown in Fig. 4.5(a). The system shown in Fig. 4.5 (a) has a 24 degree diagonal full field-of-view with an eye clearance of about 15 mm and a 3 mm eye box. The surface normal of the mirror along the vertex has a tilt angle of ~15° in the *x-z* plane with respect to the optical axis of the human eye. The system is full-color operating across the visible regime (400-700 nm). The surface profile of the mirror is described using an *x-y* polynomial. The fabricated mirror prototype is shown in Fig. 4.5(b). The predicted modulation transfer function plotted out to the Nyquist frequency is shown in Fig.

111

4.5(c). The Nyquist frequency for the modulation transfer function is based on a 15 µm pixel spacing and is therefore plotted out to 35 cycles/mm. The distortion grid is shown in Fig. 4.5(d); the system has a maximum predicted distortion of about 2.5%. The key system specifications are summarized in Table 4.2.

The fabricated mirrors have been measured with an interferometer as shown in Fig. 4.6 (a)(b)(c) shows the interferograms of the fabricated mirrors. These measurements were done at the optical shop and only a small part of the surface was measured as shown in Fig. 4.6 (a)(b)(c). The substrate material was aluminum. A surface roughness of 30nm RMS has been achieved as shown in Fig. 4.6(d).

Designs in this Chapter leveraged free-form optical surfaces described with using multivariate polynomials such as an *x-y* polynomial. Next Chapter will question the appropriateness of multivariate polynomials as free-form surface descriptors and investigate the potential of radial basis functions as an alternative descriptor in optical design.

Table 4.2 Specification table for the single-element eyeglass display

Parameter	Achieved parameter in the final design
Full diagonal field of view [degrees]	24
Eye box diameter [mm]	3
Effective focal length [mm]	14.25
Eye clearance [mm]	>15
Spectral range	Visible spectrum (400-700nm)
Image quality	*>10% at 33 cycles/mm*
Distortion	*<3%*
Microdisplay parameters	Kopin QVGA 320x240 pixels .25" diagonal (3.6mm*4.8mm) 14.1 microns pixel size
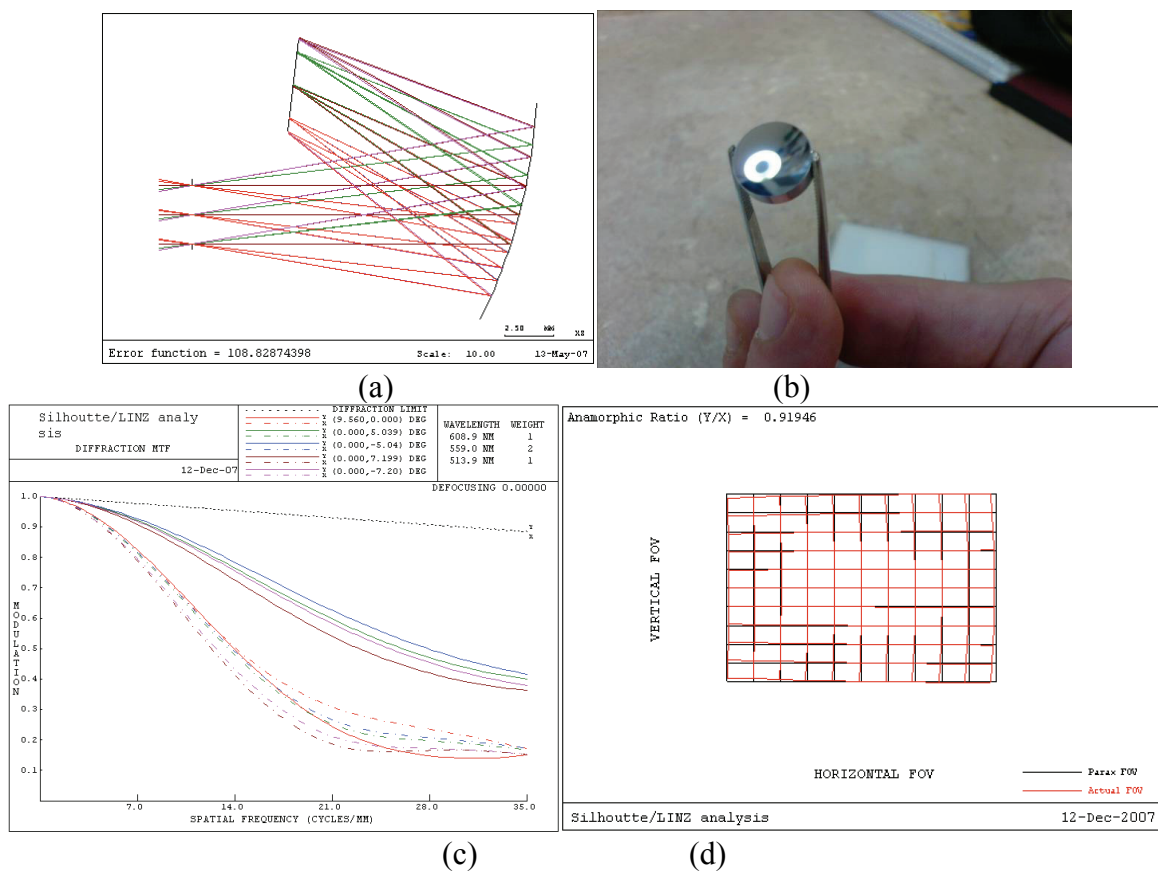
Figure 4.5 Single element magnifier (a) Optical layout (b) Fabricated mirror prototype (c)

Predicted modulation transfer function plotted out to 35 cycles/mm (d) Predicted distortion grid
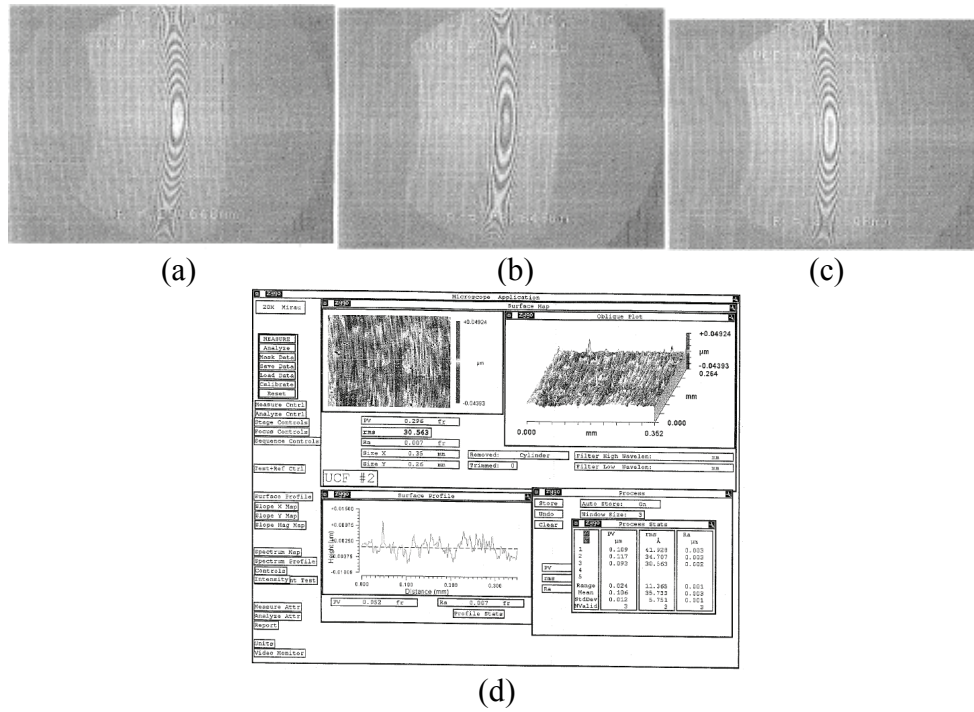
Figure 4.6 (a) Measured mirror #1 (b) Measured mirror #2 (c) Measured mirror #3 (d) Surface

profile measurement of mirror #2 showing a 30nm RMS surface roughness measurement on the

fabricated part

CHAPTER 5 MESHFREE APPROXIMATION METHODS FOR SURFACE REPRESENTATION IN OPTICAL DESIGN

In Chapter 4, we provided example head-worn display designs, where the sag of the free-form surfaces was described using polynomials with two variables (i.e., multivariate) such as an *x-y* polynomial. We use the term sag in reference to the departure of an optical surface from its tangent plane. We use the term free-form in reference to surfaces having no rotational symmetry. In this Chapter, we pursue alternatives to multivariate polynomials when describing the sag of optical free-form surfaces. We provide two reasons to motivate our research into alternatives to polynomials.

The majority of aspheric optical surfaces today are described by rotating a base curve around an axis. In 1D, there is a unique polynomial interpolant (e.g., base curve) to a set of distinct points [53]. For a rotationally non-symmetric surface, we can no longer rotate a base curve to describe a surface, the surface description task is now a 2D problem. First reason for seeking an alternative representation is based on the limitations of multivariate polynomials in 2D or higher as stated by the Mairhuber-Curtis theorem. The Mairhuber-Curtis fact will be quoted and interpreted in this chapter.

Second motivation for seeking alternatives to polynomials includes the need for descriptions that allow local anamorphisms from the point of view of the chief ray intersecting a local patch on a surface. Higher degree polynomials, depending on the point distribution, tend to oscillate when describing local changes, known as the Runge phenomenon [2]. Such oscillations

are undesirable for optical surfaces because these oscillations in the sag of a surface will degrade image quality.

In Section 5.1, radial basis functions are proposed as an alternative optical free-form surface representation. In Section 5.2, we compare the radial basis function framework to multivariate polynomials using the modulation transfer function criteria and assess the potential benefits in performance, eye box size and/or field of view expansion. In Section 5.3, we apply the radial basis function framework to the description of the free-form mirror in the dual-element head-worn display discussed in Chapter 4.

5.1 Surface Representation: Radial Basis Functions

As discussed in Section 3.2.1, stigmatic imaging of a point to a point can be achieved using conic sections for certain imaging geometries. A parabola will image an object at infinity to its focal point perfectly within the geometrical optics approximation. An ellipse will image perfectly, within a geometrical optics approximation, an object placed at one of its foci to the other focus.

It is possible to take an algebraic approach in finding analytical solutions to optical design problems. Interested readers may refer to the work of Andrew Rakich [147] and Korsh [103] for further pointers on algebraic approaches. Two aspheric surfaces can be designed for stigmatic imaging with the sine condition imposed by integrating the Wasserman-Wolf differential equations, which generate the aspheric surface profiles [207]. Wasserman-Wolf is a non-iterative technique that generates a pair of differential equations describing surfaces that will

result in a stigmatic image of a point object. The Wasserman-Wolf technique obtains coma correction by satisfying the sine condition. As reviewed in Section 3.2.3, essentially, by rewriting a vectorial version of Snell's law at each aspheric interface, Wasserman-Wolf generates a slope field representing the tangents on each point of each aspheric surface. The slope field is then integrated using a Runge-Kutta or Adams method to get the surface profile. The original paper deals with the case of rotationally symmetric aspheres.

The Wasserman-Wolf technique is no longer applicable in the case of imaging extended objects. We can conceptually break down optical surfaces into small areas with each surface patch having a surface normal. In the case of multiple fields, rays from different field points demand different surface normals from each local patch in order to image the field point of interest to its respective image location. Therefore, the problem turns into an optimization problem where we minimize the error for each field and balance the demands from each individual field.

Rotationally symmetric surfaces can be described by the revolving a 1D function about an axis. In the case of free form surfaces, the entire surface must now be described in two dimensions described, and mathematically we continue to think of an optical surface as a function, yet now in two dimensions. We will next introduce a unique perspective on optical surface representation by making a connection with the mathematics of scattered data approximation [213].

Problem (Scattered Data Approximation) We are given a set of *data sites* $\{x_1, x_2, ..x_n\}$ and a set of scalar *data values* $\{z_1, z_2, .., z_n\}$, we are looking for the function s that is approximately equal to the data values at each data site, i.e., $s(x_i) \approx z_i$. The terminology is illustrated in Fig. 5.1.

In optical terms, the data values are the sag of a surface. The data sites are determined by the intersection of the set of optimization rays with a free-form surface. The evaluation points are mostly of interest during the image quality analysis. Examples of image quality analysis metrics include but not limited to the Strehl ratio, the modulation transfer function, spot diagrams, and transverse ray aberrations. During the image quality analysis, a set of rays are traced through the system and the intersection of these rays with a free-form surface determines the evaluation points. The evaluation points not need to be coincident with the data sites.
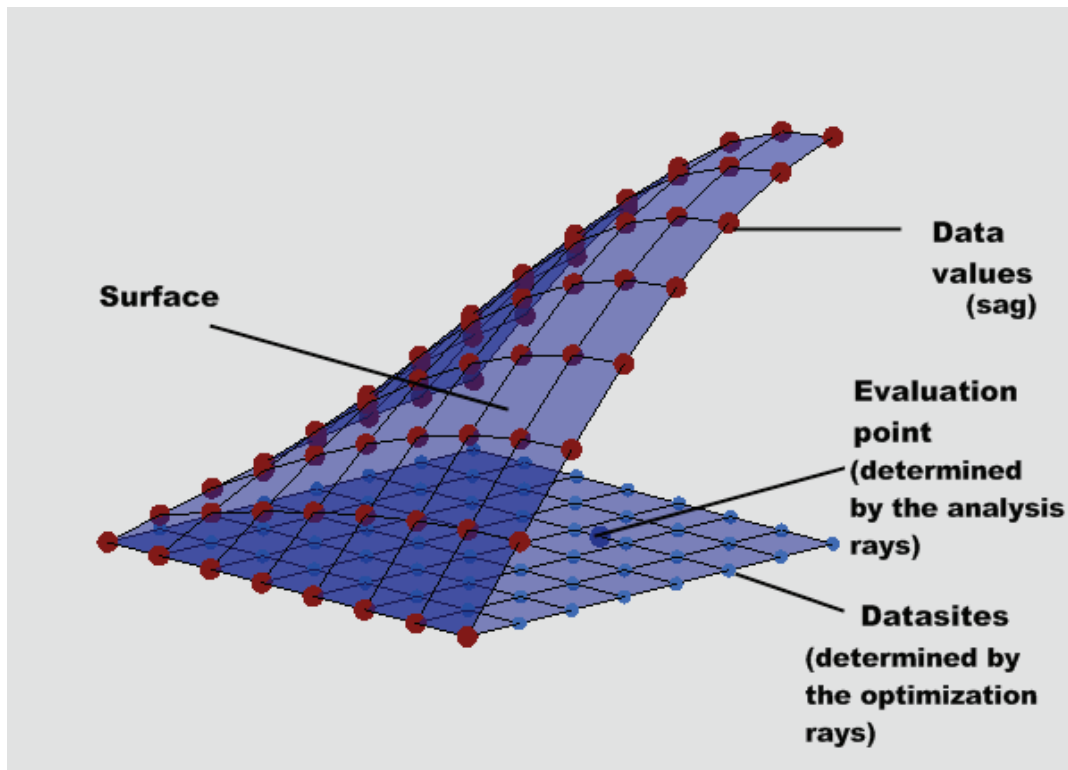
Figure 5.1 Illustration of the Function Approximation Terminology

Optical imaging problems with extended objects and aspheric surfaces often involve optimization in order to find an optimal solution to the surface shapes in the optical system in order to balance demands from multiple fields on the normals along the surface. The standard way, according to the International Standard Organization 10110 Part 12 [20], to describe aspheres in imaging optics has been to add polynomials to a base conic as

$$s(r) = \frac{cr^2}{1+\sqrt{1-(1+k)c^2r^2}} + \sum_{i=0}^{n} a_i r^{2i+4} \tag{5.1}$$

where c represents the curvature, r is $\sqrt{x^2+y^2}$, and k is the conic constant. In 2D polynomials such as x-y polynomials or Zernike polynomials are used.

In general, one could see the asphere being described as a function that can be approximated with a linear combination of basis functions

$$s(x,y) = \sum_{n} \phi_n(x,y) w_n \tag{5.2}$$

where the ϕ_n represent the basis functions and the w_n represent the weights. The three standard choices for the basis functions (ϕ_n) include polynomials, trigonometric functions such as a sine or a cosine function, and rational functions that are a ratio of two polynomial functions.

Polynomials have been a popular choice in optics and some of their properties in 1dimension and higher dimensions will be discussed next. It is known that in 1D there will be a unique polynomial of degree N-1 that will interpolate N distinct points. Interpolation using polynomials with equally spaced points suffers from the Runge phenomenon [2]. The Runge phenomenon (i.e., oscillations between points) can be observed by considering a set of equally

spaced sample points on the interval [-1,1] and fitting a polynomial to the Runge function [225].

Increasing the number of sample points and hence the polynomial order does not yield better

approximations with a uniform spacing of the interpolation points. In fact, this introduces huge

oscillations near the end points of the interval as shown in Fig. 5.2(a). Changing the point

distribution from uniformly spaced to Chebyshev distributed points (i.e., points equally spaced

on a circle and projected to an axis) will improve the fit to the Runge function. The Chebyshev

point distribution for 10 points (i.e., N=10) is illustrated in Fig. 5(c). Chebyshev points can be

generated using $x_j = \cos\left(\dfrac{j\pi}{N}\right), j = 0,1,...,N$ [196]. Increasing the polynomial order, with a

Chebyshev point distribution, helps achieving better fits in the case of the Runge function as

shown in Fig. 5.2(b).

We notice in Fig. 5(c) that the point distribution gets denser closer to the ends of the

interval. The polynomial fit in Fig. 5.2(b) is an order 9 fit. The polynomial fit shown in Fig.

5.2(a) was performed using a Vandermonde matrix [146] expressed as

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & \vdots & x_1^{n-1} \\ 1 & x_2 & \vdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \vdots & x_n^{n-1} \end{bmatrix}$$

(5.3)

The difficulty of choosing a basis in two or higher dimensions has been clearly stated by

the Mairhuber-Curtis fact [112], which says that a multivariate basis, polynomial or otherwise,

with a finite number of basis functions and a set of distinct interpolation points, may not lead to

an invertible interpolation matrix, if the basis functions are independent of the data. In the next

few pages, we expand on the implications of this statement. When we solve an interpolation

problem to describe an optical surface, we are often interested in a unique solution (e.g., optical

surface) and the uniqueness of a solution is related to the inverse matrix. That is why we will

refer to the invertibility of an interpolation matrix.
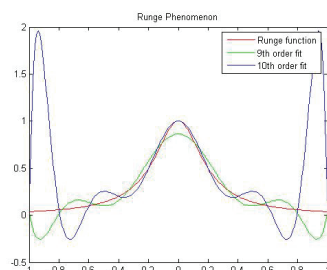
(a)
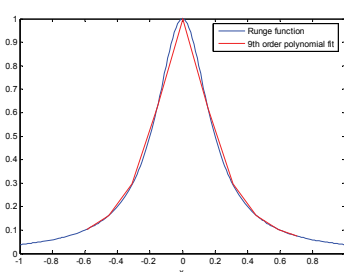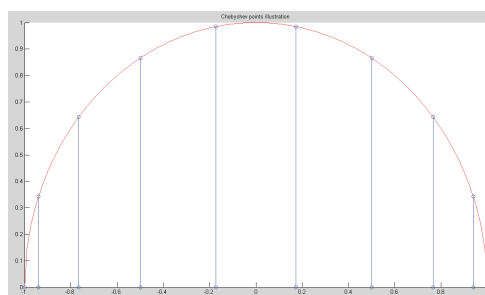
(b)   (c)

Figure 5.2 (a) Least squares polynomial fit to 9 and 10 equally-spaced samples from the Runge

function (b) Least squares based 9th order polynomial fit to a Chebyshev point distribution of the

Runge function (c) Illustration of 10 Chebyshev points.

First, we illustrate the Mairhuber-Curtis fact with an example. If we take an *x-y*

polynomial, such as $\{1, x, y, xy, x^2, y^2\}$ and take 6 distinct points, for example,

$\{(0,0), (1,-1), (1,1), (-1,1), (-1,-1), (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})\}$, the determinant of the interpolation matrix A will be

$$\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & \dfrac{\sqrt{2}}{2} & \dfrac{\sqrt{2}}{2} & \dfrac{1}{2} & \dfrac{1}{2} & \dfrac{1}{2} \end{vmatrix} = 0$$

We notice that the matrix *A* is singular because of the linear dependence of the last two columns of the interpolation matrix. This example illustrates how a polynomial basis that is a function of two variables and independent of the data can lead to a singular interpolation matrix that is not invertible.

In the rest of this chapter, we quote the Mairhuber-Curtis fact (i.e., generalization of the illustration given above); briefly discuss splines as one option to get around the Mairhuber-Curtis fact and point out the dependence of splines on an underlying *mesh* of control points; discuss radial basis functions as a second option to get around the Mairhuber-Curtis fact and without the need for an underlying mesh of control points; provide the condition (i.e., Bochner's theorem) that a radial basis function would need to satisfy such that the translates of a radial basis function lead to an invertible interpolation matrix; show examples of radial basis functions; illustrate the approximation of a paraboloid in order to provide some geometric (visual) intuition about how the RBF approach works; discuss the impact of the shape factor in the case of coincident and non-coincident basis centers and data sites.

Definition Let the finite dimensional linear function space $B \subseteq C(\Omega)$ have a basis $\{B_1,...B_N\}$. Then B is a Haar space on Ω if

$$|A| \neq 0$$

for any set of distinct $x_1, ..., x_N$ in Ω [59] and |.| denotes the determinant. Here A is the matrix with entries $A_{jk} = B_k(x_j)$.

Theorem (Maihuber-Curtis [59]) If $\Omega \subset R^s$, $s \geq 2$, contains an interior point, then there exists no Haar spaces of continuous functions except for one dimensional ones.

Therefore, in order to have a well-posed interpolation problem (i.e., invertible interpolation matrix), we are led to choose a basis that is a function of the data sites. As mentioned previously, one option for choosing a basis that takes the data into account, include breaking up the domain into piecewise segments and connecting each segment using a polynomial along with continuity constraints, referred to as the spline approach. One of our main motivations in discussing the spline approach is to get familiar with the concept of a *mesh* of control points so we can understand the term meshfree.

A simple approach to deriving the Bezier curve (spline) is the repeated application of linear interpolation as illustrated in Fig. 5.3.
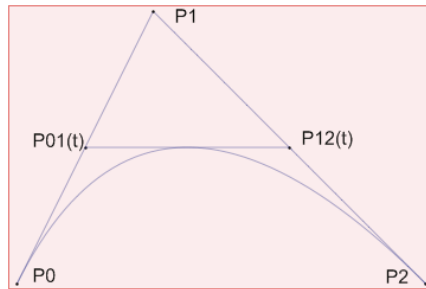
Figure 5.3 Example Bezier Curve

We describe the Bezier curve P(t) shown in Fig. 5.3 as

$$P(t) = (1-t) \, \mathbf{P_{01}}(t) + t \, \mathbf{P_{12}}(t) \tag{5.4}$$

where

$$\mathbf{P_{01}}(t) = (1-t) \, \mathbf{P_0} + t \, \mathbf{P_1} \tag{5.5}$$

$$\mathbf{P_{12}}(t) = (1-t) \, \mathbf{P_1} + t \, \mathbf{P_2} \tag{5.6}$$

Using equation 5.5 and 5.6 in equation 5.4 and rewriting in matrix form yields

$$\mathbf{P}(t) = \begin{bmatrix} t^2 & t & 1 \end{bmatrix} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} \mathbf{P_0} \\ \mathbf{P_1} \\ \mathbf{P_2} \end{bmatrix} \tag{5.7}$$

where $\{\mathbf{P_i}\}$ is the set of *control points*.

The Bezier curve can be extended to a surface using the Cartesian product. The Cartesian product is defined as $\mathbf{S}(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{m} f_i(u) g_j(v) \mathbf{P_{ij}}$ [142]. Applying the definition of the Cartesian product to the Bezier curve in (5.7) yields a bi-quadratic Bezier surface patch

$$\mathbf{S}(u,v) = \begin{bmatrix} u^2 & u & 1 \end{bmatrix} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} \mathbf{P_{00}} & \mathbf{P_{01}} & \mathbf{P_{02}} \\ \mathbf{P_{10}} & \mathbf{P_{11}} & \mathbf{P_{12}} \\ \mathbf{P_{20}} & \mathbf{P_{21}} & \mathbf{P_{22}} \end{bmatrix} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}^T \begin{bmatrix} v^2 \\ v \\ 1 \end{bmatrix} \tag{5.8}$$

The set of control points $\{\mathbf{P_{ij}}\}$ is called a control *mesh*. Generation of these meshes for irregular point sets in three dimensions is difficult and "virtually impossible for higher dimensions" [59]. The Bezier curve can be further generalized through weights and knots to a NURBS curve, however, the reliance on a mesh remains. The NURBS curve is essentially a rational basis [142].

Next we discuss a second option to get around the Mairhuber-Curtis fact, namely radial basis functions. Radial basis functions, unlike NURBS, do not rely on an underlying mesh of control points to generate a surface. A meshfree approach has the advantage of avoiding the computationally costly and cumbersome triangulation algorithms.

Definition [59] A function $\Phi : R^S \rightarrow R$ is called *radial* provided there exists a *univariate* function $\varphi : [0, \infty) \rightarrow R$ such that

$$\Phi(x) = \varphi(r), \text{ where } r = \|x\|$$

And $\|.\|$ is some norm on R^S.

The implication of a radial function is illustrated in Fig. 5.4. Figure 5.4 shows two vectors x_1 and x_2

$$x_1 = \begin{bmatrix} 0.5 \\ 0 \\ 0.5 \end{bmatrix} \text{ and } x_2 = \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

The 2-norms of the vectors $\|x_1\|$ and $\|x_2\|$ are equal. In this example, we have chosen the function $\varphi(r) = e^{-(\varepsilon r)^2}$. Since $\|x_1\|$ and $\|x_2\|$ are equal, r is equal to $\|x_1\|$ and $\|x_2\|$, yielding the same number after r is evaluated in the function $\varphi(.)$.

In the case where the function ϕ belongs to the set of radial basis functions, the surface s can be written as

$$s(x) = \sum_n \phi(\|x - c_n\|) w_n \ . \tag{5.9}$$

where x is an evaluation point, c_n are the centers of the basis functions, and w_n are the weight vectors. The translated basis functions in equation 5.9 take the data into account by shifting a
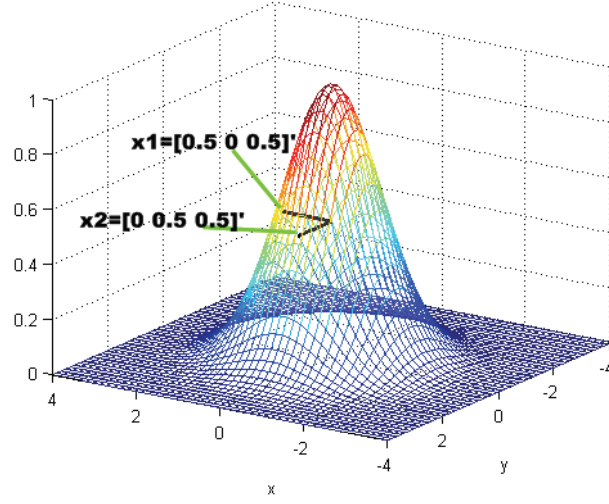
Figure 5.4 Illustration of a radial function

basic function to the data sites (and thus there are no problems with the implications of the Mairhuber-Curtis theorem). Note that we no longer have an index on ϕ because the interpolation matrix is generated by the shifts of the same function. The basis centers are placed within a domain $\Omega \subset R^s, s \geq 1$, and a discrete surface is approximated at the points of interest (i.e., evaluation points). The placement of the basis centers can be uniform, according to Chebyshev points, or even completely random. In our work, $\|\cdot\|$ denotes the 2-norm. Composing the basis functions with the 2-norm makes these basis functions spherically radial as illustrated in Fig. 5.4.

Using the approximation condition $s(x_i) \approx z_i$, where $i=1,2,\ldots,m$, at m data points and using n basis functions for $s(x)$ in equation 5.9, we can express the approximation at each point x_i as a dot product. It is possible to express the collection of m dot products with a matrix-vector multiplication and arrive at the following m by n linear system

$$\Phi_{mxn} \mathbf{w} = \mathbf{Z} \tag{5.10}$$

126

where $\mathbf{\Phi}$ is an *m×n* matrix, *w* is a length *n* vector of weights, and *Z* is the resulting discrete surface evaluated at the *m* evaluation points. Each column in the $\mathbf{\Phi}$ matrix corresponds to a radial basis function, and $\mathbf{\Phi}$ may be written as

$$\mathbf{\Phi} = \begin{bmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \dots & \phi_n(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_m) & \phi_1(\mathbf{x}_m) & \cdots & \phi_n(\mathbf{x}_m) \end{bmatrix}. \tag{5.11}$$

Therefore, the question now becomes the conditions on the basic function ϕ, translates of which generate the $\mathbf{\Phi}$ matrix. A popular point of view is to assume that $\mathbf{\Phi}$ is positive definite (and *m=n*). In this case, the connection between the choice of ϕ and a desired positive definite interpolation matrix $\mathbf{\Phi}$ is provided by the Bochner theorem.

Theorem (Bochner [59]) A function $\Phi \in C(R^s)$ is positive definite on R^s if and only if it is the Fourier transform of a finite non-negative Borel measure μ on , i.e.,

$$\Phi(x) = \hat{\mu}(x) = \frac{1}{\sqrt{(2\pi)^s}} \int_{R^s} e^{-ixy} d\mu(y), x \in R^s$$

The most important practical implication of the Bochner theorem is that the translated versions of a positive definite function ϕ indeed generate a positive definite $\mathbf{\Phi}$ matrix.

Definition The matrix A is positive definite if

$$x^T Ax > 0$$

for every nonzero *x* vector [185].

Positive definite matrices have positive eigenvalues; therefore, positive definite matrices are invertible. The theory developed by Micchelli, using the results of Bochner, covers the case where the basis centers and the data points coincide (i.e., symmetric and positive definite interpolation matrix $\boldsymbol{\Phi}$). There is not much theory available for the case when the data sites and the centers do not coincide [59]. The designs discussed in Section 5.2 and Section 5.3 did not require a user-defined error function that would trace rays only at the basis centers during optimization. In other words, the basis centers and the data points in those experiments are not coincident.

Examples of positive definite radial basis functions include the Gaussian, and the inverse multiquadric (i.e., $\beta > 0$ in Table 1). Thin-plate splines are conditionally positive definite. The mathematical expressions of these functions are given in Table 5.1.

Given a particular surface $\boldsymbol{Z}$, the choice of a radial basis function, the choice of basis centers and data points, and a shape factor in the case of a Gaussian or a multiquadric basic function, the weights needed to describe the optical surface can be found using a conventional least squares solution by

$$w = \left(\boldsymbol{\Phi}^T \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^T \boldsymbol{Z} \tag{5.12}$$

Figure 5.5 illustrates the approximation of a paraboloid using a linear combination of Gaussian radial basis functions.

We would like to note that except for the thin-plate spline, the Gaussian and the multiquadric functions contain a shape factor ε. In the case of the Gaussian, the shape factor is related to the variance as $\varepsilon^2 = \dfrac{1}{2\sigma^2}$. The impact of the choice of ε on the accuracy of a function

approximation, in the case of coincident basis centers and data points, is visualized in Fig. 5.6.

Figure 5.6 (a) is a plot of a parabolic function.
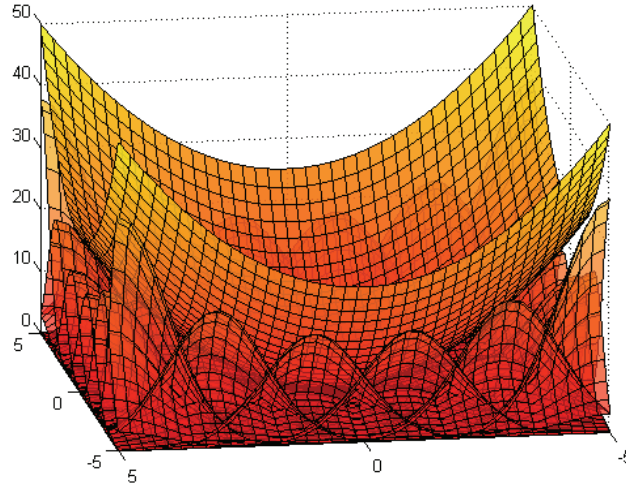
Figure 5.5 Illustration of a linear combination of Gaussian radial basis functions approximating a paraboloid. The 2D Gaussians are spaced uniformly with means centered on a 6x6 grid. Each 2D Gaussian has a variance of unity along the x and y dimensions (i.e., shape factor of 0.7071). The weighted Gaussian basis functions shown underneath the paraboloid are found through least squares.

A 6x6 uniformly distributed Gaussian RBF approximation of a paraboloid with $\varepsilon=0.1$ is shown in Fig. 5.6 (b). The RMS error is on the order of 10^{-9}. Therefore, a plot of it is not shown. A 6x6 uniformly distributed Gaussian RBF approximation with $\varepsilon=1$ is shown in Fig. 5.6 (c). The RMS error is on the order of 10^{-3} and is plotted in Fig. 5.6 (d). A 6x6 uniformly distributed Gaussian RBF approximation with $\varepsilon=10$ is shown in Fig. 5.6 (e). The RMS error is on the order of 10^{-2} and is plotted in Fig. 5.6 (f). We observe that when the basis functions become "too
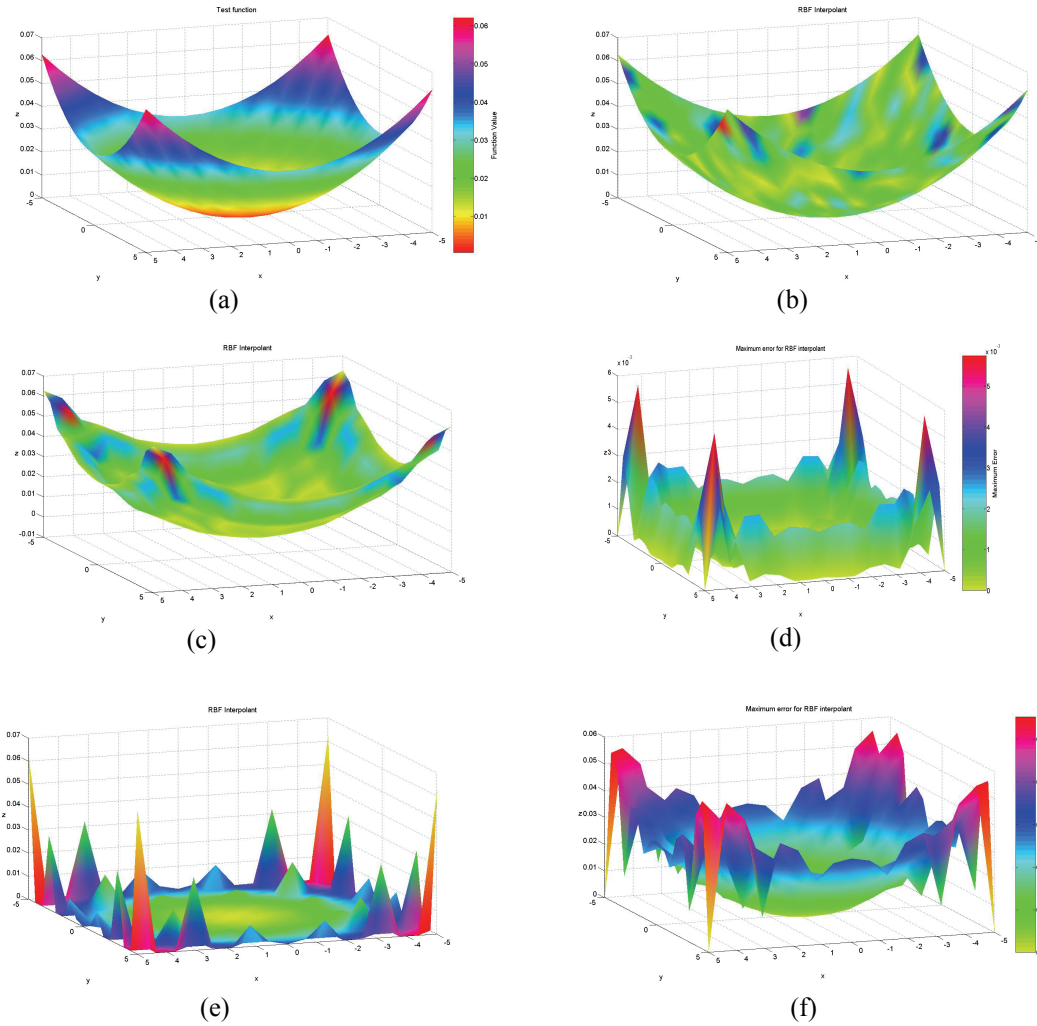
Figure 5.6 Impact of the shape factor on the accuracy of the approximation (a) Plot of the function (paraboloid) being approximated (b) Plot of the 6x6 uniformly distributed Gaussian RBF approximation with ε=0.1 (c) Plot of the 6x6 uniformly distributed Gaussian RBF approximation with ε=1 (d) Plot of the error $|f_{approx.} - f|$ (e) Plot of the 6x6 uniformly distributed Gaussian RBF approximation with ε=10 (f) Plot of the error $|f_{approx.} - f|$

peaky", the quality of the approximation is reduced. The optical design examples given in Section 5.3 used shape factors that were found through experimentation.
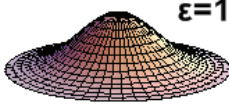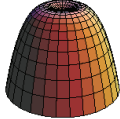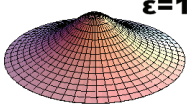
The shape factor plays an important role in the accuracy of an RBF based approximation scheme. The theoretical tools for choosing a shape factor (e.g., cross-validation) take the data values into account [150]. However, these data values are not known before our raytrace optimization procedure. This makes the application of techniques such as cross-validation challenging. Therefore, the link between a shape factor that yields maximal performance with respect to an optical performance criterion such as the RMS wavefront error or the modulation transfer function requires further research. Early experimentation suggests that it is possible to obtain a diffraction limited surface, in the case of different (not coincident) basis centers and data points, for the variances of 0.05, 0.1, 0.5, 1, 2, 3, 5, 10, 20, 30, 40 and 100. The high success rate for a relatively large range of variances could be either due to the simplicity of the paraboloid as a testcase or the differing centers and data points. Understanding the impact of the shape factor in the case of noncoincident basis centers and data sites requires further research.

A few natural model choices that come up in the application of radial basis functions to optical surfaces include the choice of the basic function, the number of basis functions to use in a particular problem, the spatial distribution of the basis centers, and the choice of the shape factor.

To our knowledge, there is no theoretical guidance on choosing the number of basis functions, type of the basis function, placement of the centers, for a particular optical design problem. The minimum number of basis functions that provides an adequate approximation is found through experiments and is application specific. In terms of the placement of the basis

centers, Chebyshev distribution of points is likely to be more desirable over a uniform distribution of points.

Table 5.1 Definition and plots of the Gaussian, thin-plate spline, and the multiquadric radial basis functions

Gaussian	Thin-plate spline (no shape factor)	Multiquadric
$\phi(x) = e^{-\varepsilon^2 \|x\|_2^2}$, $x \in R^s$	$\phi(x) = \| x \|^{2\beta} \log(\| x \|)$, $x \in R^s, \beta \in N$	$\phi(x) = (1 + \varepsilon^2 \| x \|^2)^{-\beta}$ $x \in R^s, \beta \in R \setminus N_0$

5.2 Radial Basis Functions: Optimization

In this section we will describe the optimization procedure that was developed to optimize RBF surfaces into the previously described optical designs. Our implementation and initial experimentation have been with the Gaussian radial basis function as the choice of basis. Gaussians possess several desirable properties from the point of view of optical design. First, Gaussians are smooth functions (C^∞) having derivatives of all orders providing a desirable property given that optical surfaces are smooth. In addition, smoothness is desirable from a fabrication point of view. Second, theoretically, Gaussians are not local functions. However, practically they can be considered local since the value of a Gaussian outside of 3σ is small

(<0.011, for a unit amplitude Gaussian). Third, the Fourier transform of a Gaussian is a Gaussian, which provides an analytical description for the power spectral density (PSD) of the surface.

The first step was to set the grid size (i.e., number of basis functions). For the magnifier example in Fig. 5.7, we have found a 17x17 grid to perform well when compared against relatively general surface representations such as x-y polynomials or Zernike polynomials using the MTF criteria. The second step was to initialize the starting point. Our raytrace software requires a base sphere in order to perform paraxial image calculations. We added a base conic to equation 5.9 so that paraxial calculations could be performed.

Given a number of basis functions (number of columns in the $\boldsymbol{\Phi}$ matrix), we divided the aperture into x_{num} pieces in the x-dimension and y_{num} pieces in the y-dimension. The number of columns in the $\boldsymbol{\Phi}$ matrix was set by the product of x_{num} and y_{num}. The number of rows in the $\boldsymbol{\Phi}$ matrix controlled the spatial resolution of the Gaussians and was set by the user. We make a rectangular aperture assumption in this case. However, the sum of basis representation accommodates any aperture shape. We divided the aperture diameter into x_{num} pieces in the x-dimension and uniformly placed each x-mean $1/x_{num}$ apart from each other. Similarly, we divided the aperture into y_{num} pieces in the y-dimension and placed each y-mean $1/y_{num}$ apart from each other. The variances in the covariance matrix were set to 1.

Unlike the function fitting example given in Fig. 5.5 where we have a surface $\boldsymbol{Z}$ (paraboloid) to fit, in the context of the optical design problem, the surface $\boldsymbol{Z}$ is unknown a priori, and the goal of an iterative optimizer is to adjust the weights $\boldsymbol{w}$ in equation 5.10 with the goal of reaching a minimum of the merit function given a starting point. The fifth step is an

optional step to represent a starting point only with the sum of basis, without the base conic. The intention with step 5 is to allow the exploration of alternative optimization techniques (for example using the MATLAB optimization toolbox) such as the trust region dogleg, Gauss-Newton or simplex. We remind that the addition of the base conic was only required for the paraxial calculations in the raytrace code. Alternative optimization environments could utilize step 5 as the preferred way to experiment with this surface representation. The sixth step is the construction or choice of the error function. We used the transverse error in the image plane, which is the sum of squares of the deviations of the rays from their respective reference wavelength chief rays, as our merit function. The seventh step was to choose an optimization technique and to optimize the error function. The results we achieved are based on the damped least squares algorithm.

A user defined surface type 1 has been implemented in C++ for Code V as a dynamically linked library to test the surface representation. A full description of user defined type 1 surfaces is provided in the Code V documentation. Code V interacts with the surface one point at a time meaning that Code V will ask the sag of the surface for a specific x, y and z point. Therefore, the Φ matrix reduces to a row vector and the sag calculation becomes a dot product operation with the weights.

We designed four systems to address the question of optimal shape for the off-axis magnifier problem. A 10th order anamorphic sphere, an x-y polynomial, Zernike polynomial using the first 66 terms in the standard Zernike expansion and a linear combination of Gaussians were compared. Each system under comparison had a >15 mm eye clearance, 3 mm eyebox, 24 degree diagonal full field of view (9.6 degree x semi-field and 7.2 degree y semi-field), and a ~15 degree mirror tilt angle. Each system was optimized with the minimum set of constraints

such as the effective focal length (14.25 mm), real ray based distortion constraints, and the field weights. Each system had 17 field points defined. The variables in each system included the surface coefficients, image plane defocus and tilt. The distance from the pupil to the mirror vertex was 16.9 mm. The distance from the vertex of the mirror to the image plane was kept around 13 mm since it is undesirable to have a microdisplay physically close to a human eye. The image plane has a rectangular aperture with a size of 4.8 mm by 3.6 mm in the x and y dimensions, respectively. The image plane needs to stay clear out of the ray path between the eye box and the mirror. 140 rays across the eye box were traced in each system during optimization. Figure 5.7 (a) shows the layout of the optimal off-axis magnifier. The MTF evaluated at λ=550nm (no dispersion), for the optimized linear combination of Gaussians surface is shown in Fig. 5.7 (b). An interferogram of the surface with the base sphere subtracted is shown in Fig. 5.7 (c). Distortion characteristics exhibit similar behavior with each design having a maximum of about 3 to 4%. Table 5.2 shows a comparison of the surface representation proposed and implemented in this thesis against an anamorphic asphere, Zernike polynomial using the first 66 terms in the standard Zernike expansion (Born and Wolf summarize the standard Zernike expansion [19]), an x-y polynomial up to and including 10th order (good balancing achieved with order 5) with the maximum distortion and the average MTF across 17 field points as the comparison metrics. The sum of local basis representation proposed in this thesis achieves the highest MTF performance averaged across 17 field points by 18.5% in the field with an acceptable level of maximum distortion, among the functions that were compared. Zernike optimization has been confirmed independently in Zemax and Code-V and the results agree to within 1.5% of the reported MTF value in Table 5.2.

(a)   (b)   (c)
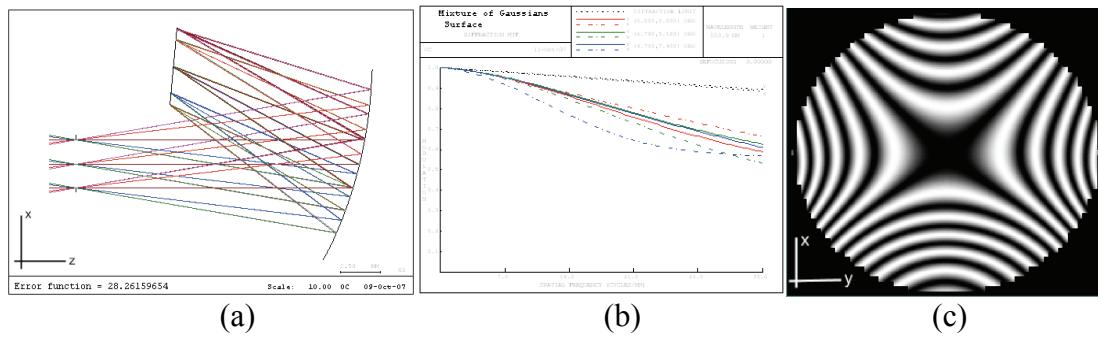
Figure 5.7 (a) Optical layout of the off-axis magnifier. (b) MTF evaluated on-axis, 0.7 in the

field, and at the maximum field plot up to 35 cycles/mm. (c) Interferogram of the surface

represented with a linear combination of Gaussian radial basis functions

Table 5.2 Comparison of the transverse error function value and 17 average tangential and

sagittal MTF values between an anamorphic asphere, x-y polynomial, Zernike polynomial, and a

linear combination of Gaussians surface type.

Surface Type	Average MTF	Max. Distortion
Anamorphic asphere	26.5%	3.8%
X-y polynomial	43.6%	2.65%
Zernike polynomial	42%	3.74%
Gaussian RBF	60.5%	3.6%

An alternative check on whether a 36 term Zernike can describe the radial basis function solution

shown in Fig. 5.7 is provided by designing an optical null surface. The layout for the design of

the 36 term standard Zernike based optical null surface is shown in Fig. 5.8 (a). The wavelength

in the null system shown in Fig. 5.8 is 5 µm. The aperture of the radial basis function based

136

mirror is 9.2mm in diameter. The aperture of the optical null surface described using a 36 term Zernike expansion is also 9.2mm. The axial difference between the optical null surface and the Gaussian RBF surface is set to be zero in the raytrace code, also the constant term (term #1) in the standard Zernike expansion is set to zero as well. The pupil map of the difference surface is shown in Fig. 5.8 (b). We observe in Fig. 5.8 that it is not possible to perfectly describe the Gaussian RBF surface using 36 terms in the standard Zernike expansion. There is an 8.6 wave peak to valley residual error in the difference surface shown in Fig. 5.8 (b).
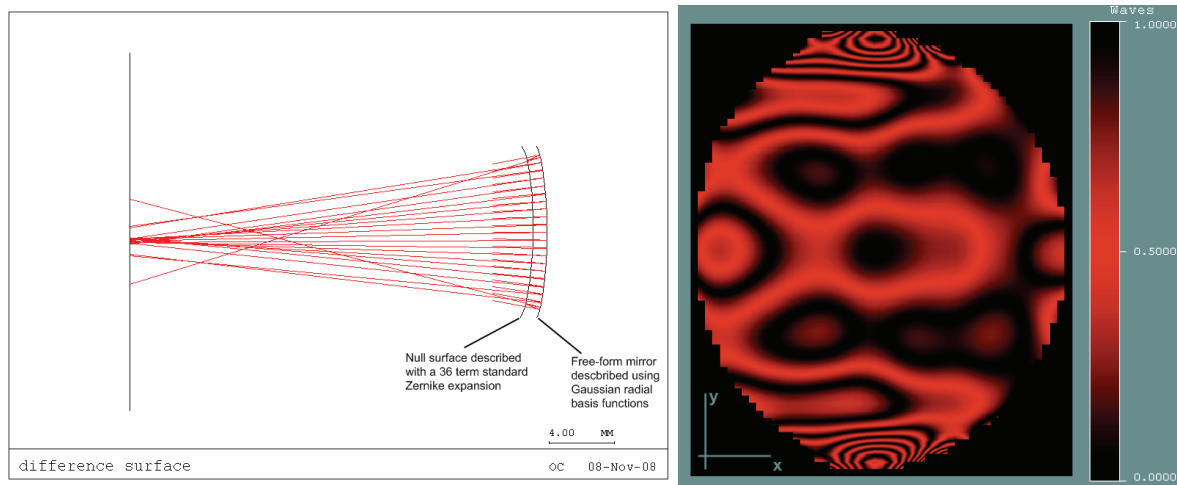
Figure 5.8 (a) The layout for the design of an optical null surface using 36 terms in the standard Zernike expansion (b) The pupil map of the difference surface ($\lambda = 5\mu$m with a 9.2mm aperture diameter)

5.3 Application of Radial Basis Functions to the Design of a Free-Form mirror in Dual Element Designs

Guided by the results given in section 5.2, in this Section we apply the radial basis function framework to the description of optical mirror shapes in a dual-element magnifier and study field of view [37] and eyebox [36] limits.

5.3.1 Field of View Limit with an 8 mm Eyebox

The optical layout for the dual-element magnifier used in this study is shown in Fig. 5.9 (a). From left to right, the system consists of an 8 mm eyebox, a free-form mirror described using Gaussian radial basis functions, the aspheric surface of the lens that forms the substrate for the diffractive optical element, the spherical surface of the lens, a flat fold mirror, and the tilted microdisplay. The free-form mirror is described using a linear combination of 225 2D Gaussian functions with variance equaling 2 in both the x and y directions. The diameter of the free-form mirror is approximately 22 mm. The free-form mirror is tilted at ~30 degrees. We found this angle to be the minimum angle that provides sufficient clearance in our previous work. The aspheric substrate of the lens is a 12th order rotationally symmetric polynomial. The diffractive optical element is described with an 8th order rotationally symmetric polynomial. The diffractive optical element is configured to operate in the 1st order and optimized at a wavelength of 558 nm. The optimization strategy was based on imposing a minimal number of constraints.

The variables included the weights of 225 Gaussians (15x15), the base curvature of the mirror, the base curvatures of both surfaces of the lens, the distance from the exit pupil (i.e., eye box) to the vertex of the mirror, the distance from the vertex of the mirror to the lens, 5 aspheric coefficients (4-12th order), 4 diffractive coefficients (2nd-8th order), and the image plane defocus. The thickness and material of the lens were not variables. The constraints included a focal length of about 27 mm, real ray based distortion constraints, real ray based eye clearance constraints, wavelength weights according to the photopic spectral response of the human eye, and the field weights. We took the design described in [32] as our starting point design. We changed the *x-y* polynomial surface describing the free-form mirror to a Gaussian radial basis function surface and initialized the *w* vector to 0 at the start of the optimization. The material for the lens is Zeonex Z-E48R and has a 92% average transmission across 400-800nm. The diffractive optical element was limited to a minimum feature size of 16 μm, which is feasible for fabrication. The performance of the system is shown in Fig. 5.9 (b). The cutoff frequency of 35 cycles/mm in the MTF plot is determined by the 15 μm pixel spacing on the microdisplay. The distortion grid is shown in Fig. 5.9 (c). The maximum distortion in this system occurs at (-10.1°,-7.6°) in the field with a magnitude of -4.2%. The lateral color is estimated to be 1 arcminute in visual space and the axial color is estimated to be about 2 arcminutes in visual space. The resolution of the human eye, for our purpose, is about 1 arcminute based on a 2.5 μm cone spacing. We achieved a 25 degree full field of view while meeting the requirements of an 8 mm eyebox, >15 mm eye clearance, and operation across the visual spectrum. Figure 5.9 (d) shows an interferogram of the free-form surface at the He-Ne wavelength of 632.8nm.

(a)

(b)

(c)

(d)
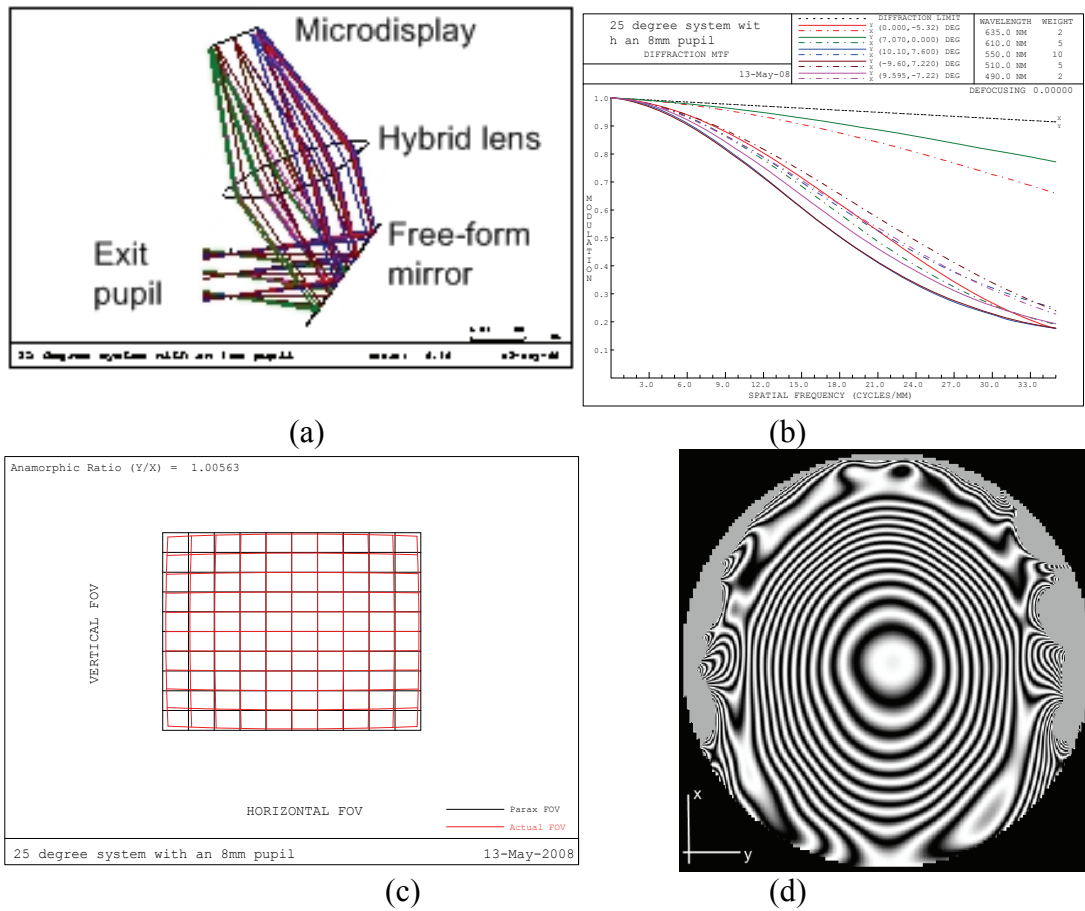
Figure 5.9 (a) Optical layout of the dual-element magnifier with a 25 degree field of view (b)

Modulation transfer function plotted out to 35 cycles/mm and evaluated at the full 8 mm eyebox

(c) Distortion grid (maximum distortion is 4.18%) (d) Interferogram of the free-from mirror

Full field displays from nodal aberration theory provide design insight into the aberration characteristics of tilted and decentered systems [194]. The full field displays are based on real raytrace data. The full field display plots can be used to provide magnitude and orientation information for individual aberration terms or for summary measures of image quality. The line length or circle size is proportional to the magnitude of the aberration of interest at the displayed grid point in the full field plot. To study the astigmatic aberration component in isolation, a generalized Coddington raytrace can be used in conjunction with the full-field display. In this

case, the orientation of the displayed images aligns with the spot diagram. Zernike coefficient based full field displays are generated by tracing rays, calculating the OPD for each ray with respect to the chief ray, and then fitting Zernike polynomials to the OPD map. Thompson applies full field displays to the analysis of a triplet in Chapter 7 of his thesis [194]. Figoski applied nodal aberration theory to find an alternate solution to a two mirror telescope problem [62]. Rogers illustrated an application of nodal aberration theory to guide the design of a head-worn display [154].

In Figure 5.10, we show field maps prior to targeted Zernike optimization for the system shown in Fig. 5.9 (a). Figure 5.10 (a) displays FRINGE Zernike terms Z5 and Z6 representing astigmatism and Fig. 5.10 (c) displays FRINGE Zernike polynomial terms Z7 and Z8 representing coma. We provide an interpolated surface in Fig. 5.10 (b) to the astigmatism magnitude given in Fig. 5.10 (a) in order to clearly display the binodal behavior with the nodes being close to ±2 degrees along the x-direction with a 0 degree y-field component. In Figure 5.10(c), the minimum coma magnitude is 0.003 waves (λ=550nm) occurring at (-7.97°, -7.6°) in the field. The RMS wavefront error across the field is displayed in Fig. 5.10 (d). The RMS wavefront error is quasi-constant with an average value of about 1.1 waves (λ=544.7nm) and a standard deviation of 0.6 waves across the entire field of view.

Figure 5.11 shows the fields maps after Zernike targeted optimization. We further optimized the system by adding a constraint of targeting Z5, Z6, Z7 and Z8 to 0 at field 1 with the intention of collapsing the binodal astigmatism and coma nodes to the center of the field. Figure 5.11 (a) displays Zernike terms Z5 and Z6 representing astigmatism. We can observe in Figure 5.11 (b) that the binodal behavior is eliminated and we now have quadratic astigmatism at 0.5 degrees in the x-field and 0.4 degrees in the y-field. The RMS wavefront error across the

field is displayed in Fig. 5.11 (c). The RMS wavefront error across the field after Zernike targeted optimization has an average of 1.1 wave (λ=544.7nm) across the field and 0.6 waves of standard deviation.

Figure 5.12 shows the fields maps after Zernike targeted optimization with the nodes at ($\pm6°,0°$). Figure 5.12 (a) displays Zernike terms Z5 and Z6 representing astigmatism. We can observe the binodal behavior in Figure 5.12 (b). The RMS wavefront error across the field is displayed in Fig. 5.12 (c). The RMS wavefront error across the field after Zernike targeted optimization has an average of 0.72 waves (λ=544.7nm) across the field and 0.15 waves of standard deviation.

Figure 5.13 summarizes the impact of moving the astigmatism nodes on the performance as measured with the MTF criteria. The MTF plots for the set of fields {(0°,0°),(0°,7.6°),(10.1°,0°),(-10.1°,0°),(9°,6.8°)} with binodal astigmatism nodes at ($\pm2°,0°$), (0.4°,0.5°), and ($\pm6°,0°$) are provided in Fig. 5.13 (a), Fig. 5.13 (b), and Fig. 5.13 (c), respectively.

Results indicate that separating the nodes further apart along the x-dimension in a rectangular image field yields better wavefront RMS performance across the field with well behaved mid-spatial frequencies in the MTF; collapsing the nodes yields worse wavefront RMS performance with worse mid-spatial frequency behavior in the MTF.

(a)

(b)

(c)

(d)
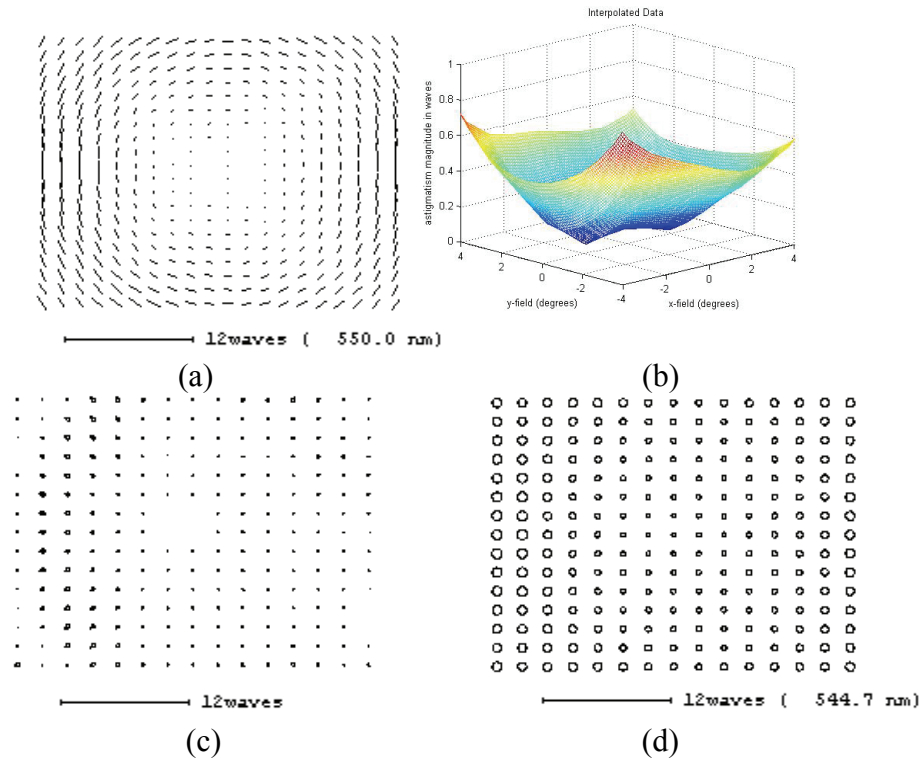
Figure 5.10 Full field displays before targeting specific Zernike terms in the field (a) Full field

display of the fringe Zernike Z5/Z6 components (astigmatism) in a 20x20 grid (b) Interpolated

plot of the astigmatism magnitude to clearly visualize the binodal behavior (c) Full field display

of the Fringe Z7/Z8 (coma) (d) Full field display of the RMS wavefront error (15x15 grid)
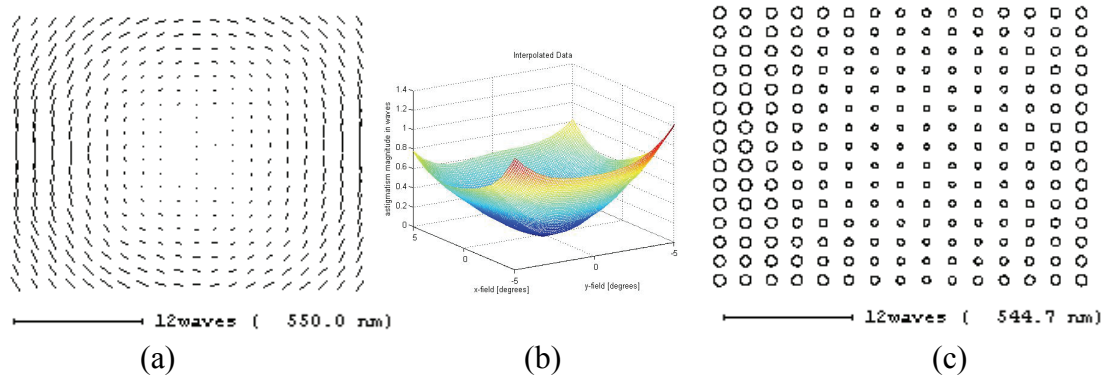
Figure 5.11 Full field displays with the astigmatism nodes collapsed at (0.5°, 0.4°) (a) Full field display of the fringe Zernike Z5/Z6 components (astigmatism) in a 20x20 grid (b) Interpolated plot of the astigmatism magnitude to clearly visualize the quadratic astigmatism behavior (c) Full field display of the RMS wavefront error (15x15 grid)
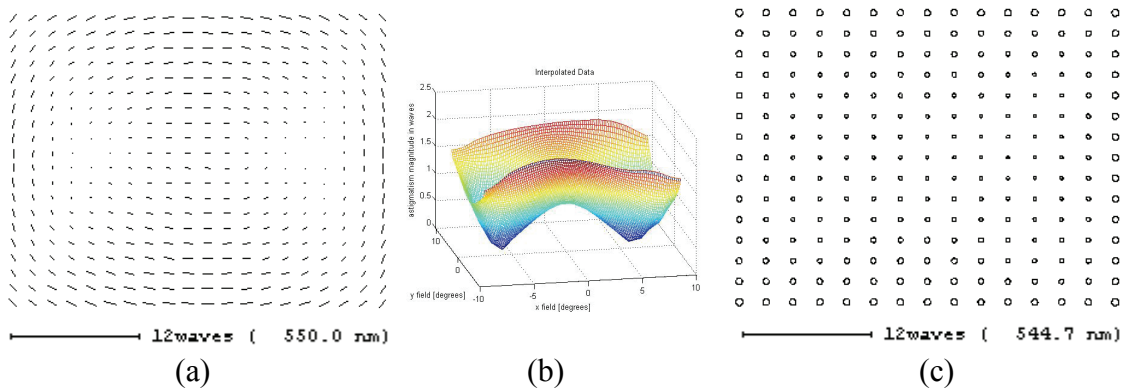
Figure 5.12 Full field displays with the astigmatism nodes at (±6, 0.4°) (a) Full field display of the fringe Zernike Z5/Z6 components (astigmatism) in a 20x20 grid (b) Interpolated plot of the astigmatism magnitude to clearly visualize the binodal astigmatism behavior (c) Full field display of the RMS wavefront error (15x15 grid)

(a)   (b)   (c)

Figure 5.13 Modulation transfer function plots for the set of fields {(0°,0°),(0°,7.6°), (10.1°,0°),(-

10.1°,0°),(9°,6.8°)} in the cases of (a) nodes at (±2°,0°) (b) nodes at (0.4°,0.5°) and (c) nodes at

(±6,0.4°)

5.3.2 Eyebox Size Limit with a 20 degree field of view

We used the design described in [32] as our starting point design. We changed the *x-y*

polynomial surface describing the free-form mirror to a Gaussian radial basis function surface

and initialized the **w** vector to 0 at the start of the optimization. The optimization policy and

constraints are the same as described in Section 5.3.1.

Figure 5.14 (a) shows the layout of the dual-element off-axis catadioptric magnifier

where the mirror is described with Gaussian radial basis functions. A red, blue, and green

microdisplay pixel triplet is 15 μm in size, which yields a Nyquist frequency of about

33 cycles/mm. The polychromatic MTF evaluated for a centered 12 mm eyebox is plotted in

Fig. 5.14 (b) for the performance limiting fields. In the case of a 3 mm eye pupil, the lowest

MTF value at the Nyquist frequency is about 50%. Thus the polychromatic MTF is plotted up to

35 cycles/mm. In this prototype, the microdisplay has a ~12 mm diagonal and contains 640**x**480

pixels. In visual space, the display provides 1.5 arcminutes resolution as limited by the pixel

spacing on the microdisplay. The maximum distortion occurs at (x=-8°, y=-6°) in the field and was measured in simulation to be -1.83%. Fig. 5.14 (c) shows the appearance of a rectilinear grid as viewed through the magnifier. Fig. 5.14 (d) shows an interferogram of the mirror surface compared to a flat reference wavefront. Based on a performance criteria of 20% light at the Nyquist frequency of the modulation transfer function evaluated at the full eyebox size, we found the eyebox size limit to be 12 mm while having a 20 degree field of view and 15.5 mm eye clearance.

(a)   (b)

(c)   (d)
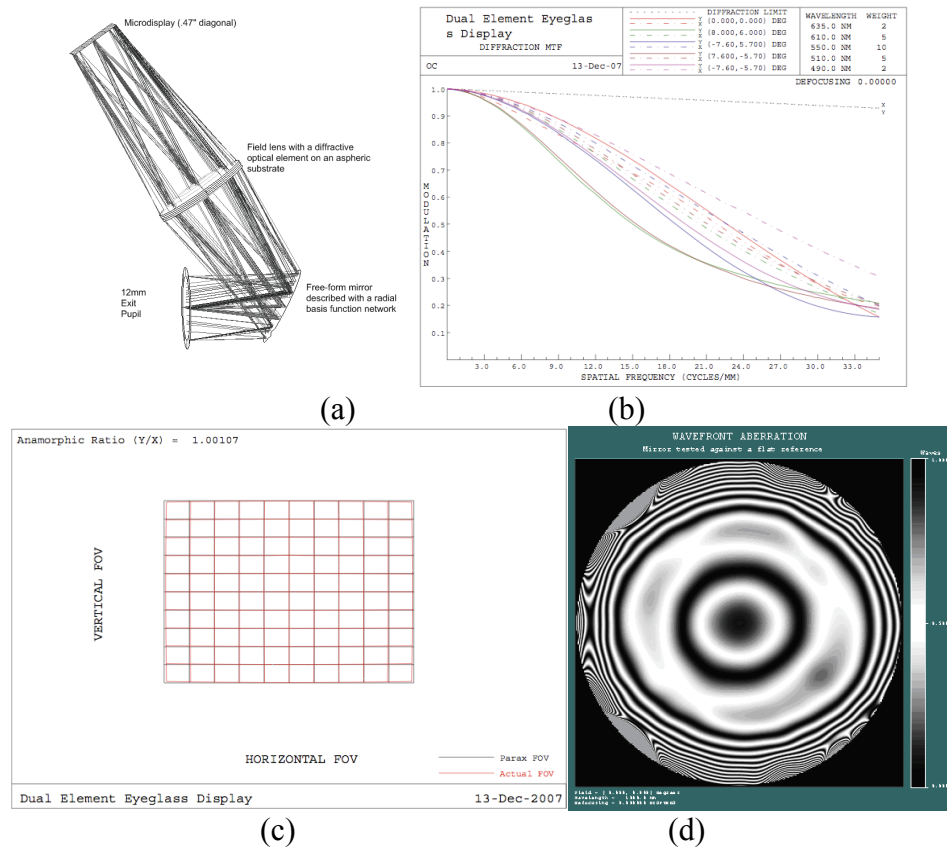
Figure 5.14 12 mm eyebox system with a 20 degree diagonal full FOV (a) Optical layout (b) Modulation transfer function plotted out to 35 cycles/mm (c) Distortion grid (d) Interferogram of the free-form mirror (632.8nm wavelength)

CHAPTER 6 PRELIMINARY DESIGN OF A MUTUAL OCCLUSION DISPLAY

In this Chapter, we summarize a preliminary approach to the optical design of a compact see-through mutual occlusion display. In Section 6.1, we introduce and define the problem of opaque overlays (i.e., occlusion). In Section 6.2, we discuss the preliminary optical design of a mutual occlusion display based on polarization optics. In Section 6.3, we summarize an initial experiment using a spatial light modulator to implement opaque overlays of computer generated imagery blocking the light reflected from the objects in the scene.

6.1 Introduction to the Problem of Opaque Overlays in Head-Worn Displays

Starting with Ivan Sutherland's original head-worn display, most optical see-through displays, even today, will combine the computer generated imagery with the real world using a beam splitter. Sutherland concludes in his 1968 paper that "showing "opaque" objects with hidden lines removed is beyond [their] present capability" [187]. He adds that "the three-dimensional objects shown by [their] equipment are transparent "wire frame" line drawings". It is well established that any beamsplitter will reflect some percentage of the incoming light and transmit the rest. The percentage of transmission and reflection can be adjusted through coatings on the surfaces. However, there will always be some amount of light that is transmitted from the

scene, which is the root of the occlusion problem in optical see-through displays. This transmitted light implies that it is optically impossible to overlay opaque objects for a display using a combination of an image source and a beamsplitter. Alternative mechanisms (to the beam splitter + image source combination) will be necessary to achieve the continuum between transparency and opaqueness of virtual objects. A virtual object needs to be opaque when it is occluding and it needs to be transparent when it is occluded by an object in the real world scene. Current computer graphics techniques and hardware allow for "hidden line removal" or visible surface determination, however, the display of "opaque" objects with optical see-through remains a problem.

Ian Robinson disclosed the concept for a system where a transmission type device was utilized to block/pass certain parts of the scene [187]. Soon after, in 1999, an article from Eric W. Tatham [190] illustrated results from a transmissive light blocking arrangement with no imaging optics. An active mask was proposed to modulate the content of a scene and was combined with the display. Tatham further pointed to some of the potential benefits of using a Digital Micromirror Device (DMD) in place of the transmissive mask, yet no optical layout was proposed. Kiyokawa and Billinghurst addressed the occlusion problem in their ELMO prototype, now in its fourth generation [98]. Vivid images of mutual occlusion were first demonstrated by the authors. The heart of their system is a transmissive spatial light modulator (SLM). ELMO-4 is based on a 320x240 transmissive LCD from Hunet. The LCD is reported to have a response time of 2ms. The ELMO-4 optical system contains 4 lenses, 2 prisms, and 3 mirrors per eye for the display component. The designers of the ELMO-4 prototype took special care to eliminate the viewpoint offset caused by the optics, which justifies the current size for the approach taken by the authors. The other components in the ELMO-4 are additional cameras located around the

148

optics to enable depth extraction of the scene, which will not be further addressed here. Reflective SLM approach using a DMD has been pursued by Uchida, Sato and Inokuchi [198]. According to our understanding (original paper is in Japanese), in this prototype, there are three optical paths. One optical path is for imaging the scene onto the DMD, another one is for imaging the microdisplay onto the DMD and the last one is for projection and/or eyepiece optics. This is a promising prototype due to the high contrast ratio of the DMD device. However, further work on the optical layout will be needed to combine the three separate optical paths into a single path and also to eliminate the potential viewpoint offset, if this approach is to be used in a head-worn display.

In this Chapter, we will discuss a preliminary optical see-through head-worn display that is capable of mutual occlusions. One difference between our system and the previous systems is that our system is based on a reflective SLM combined with a single optical path, suitable for use in head-worn displays. Our system has a resolution of 1280x1024, a switching speed within microseconds, which is due to the faster switching in ferroelectric liquid crystal compared to nematic liquid crystals (an order of magnitude faster than the transmissive masks), a 60Hz video rate, and importantly is optically compact (e.g., 2 lenses and a single x-cube prism per eye). A crude estimate of the weight of the proposed head-worn display yields less than 50 grams per eye. This estimate accounts for 17 grams for the X-cube prism, 5 grams for the lenses, and ~30 grams for the LCOS and the driver electronics. An F-LCOS (ferroelectric liquid crystal on silicon), which we propose to use as one implementation of the SLM panel, yields >1:200 contrast ratios. We expect a light throughput between 40-50% of what the scene would provide for the real world objects. For the virtual objects, we expect a 50% light throughput depending on whether the microdisplay used to display the images is polarized (e.g. LCD and F-LCOS

panels, the later being a possibility to be used also as a display) or unpolarized (e.g. OLEDs). Based on 700-1000fL brightness of current microdisplay technology, we can expect about 350fL brightness on the display of virtual objects. Finally, we shall show that the system can be designed distortion free while viewing the real scene.

A basic idea is that, depending on the optical properties of a material, in general the complex refractive index $\eta = n + i\kappa$ of an object, light is transmitted through a material, reflected, or it is absorbed. n is the real part of the refractive index and κ is the complex part of the refractive index that relates to the absorption coefficient α [cm^-1]. Occlusion is a result of light propagation and attenuation through different materials within a scene. Occlusion is a concept very familiar to us from daily life. For example, a tall person sitting in a theatre would block the light from the theatre stage for the people immediately in the row behind since light in the visible spectrum is not transmitted through people. We can say that the tall person is occluding the field of view of the people sitting in the immediate row behind. We present a visual example of occlusion in the context of augmented reality. Fig. 6.1 (a) illustrates a case where a virtual apple would be overlaid on a real scene with a conventional head-worn display. We can observe that the apple is semi-transparent as a result of the transmission through the beamsplitter. Fig. 6.1 (b) illustrates a virtual apple overlaid on a real scene through an occlusion supporting head-worn display.

(a)   (b)

Figure 6.1 (a) Concept of occlusion illustrated with a virtual apple as would be seen through a beam splitter based head-worn display (b) Concept of occlusion illustrated with a virtual apple as would be seen through an ideal occlusion based display.

We can observe that the apple occludes the cup behind it and is occluded by the cup in front of it, consistent with our daily experience of objects mutually occluding each other. Based on this ideal occlusion illustration shown in Fig. 6.1(b), we outline the issues to be addressed as follows and illustrate the steps that such a system would take in Fig. 6.2:

1) An optical system that can image a desired amount of the field of view, based on the application needs.

2) A depth sensing mechanism.

3) An algorithm to create a binary mask based on depth information.

4) An optical system that modulates the scene based on the mask.

5) An optical system that combines the modulated scene with the image display hardware.
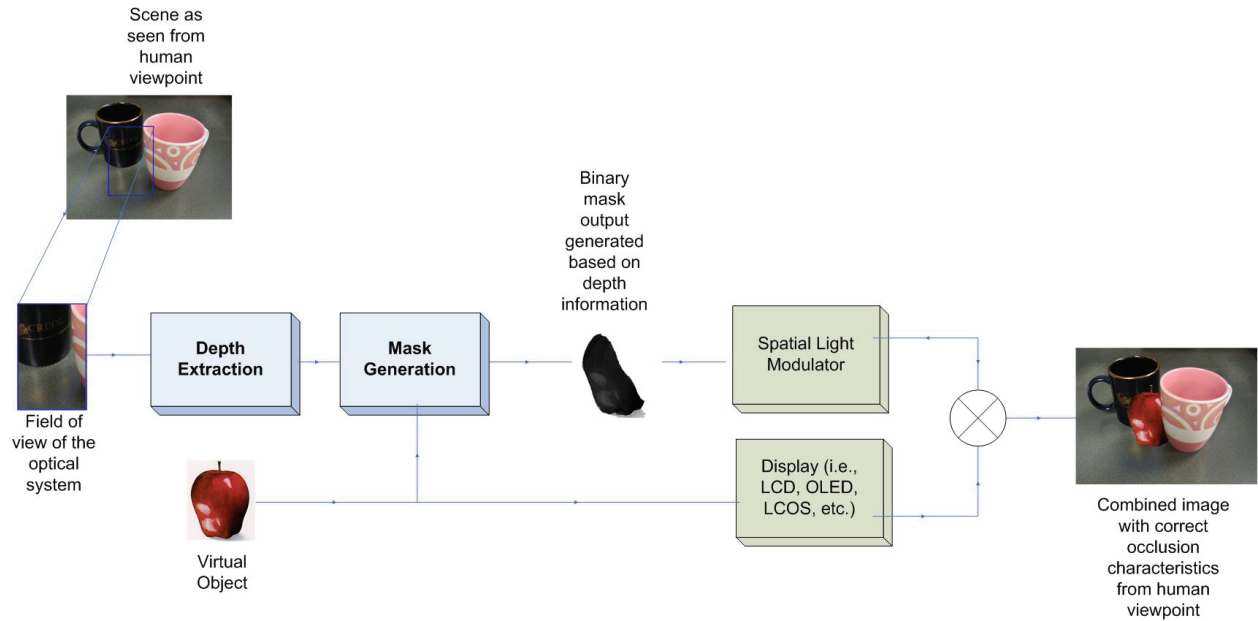
Figure 6.2 Block diagram of an ideal occlusion system

We envision the application scope of our head-worn display to include mobile outdoor scenarios. Mobility implies that the system should be compact and light weight for maximum comfort while wearing the display.

6.2 Preliminary Optical Design of an Occlusion Display

In this section, we present the first-order design details, preliminary lens design, and performance analysis of a compact optical system that can achieve mutual occlusions. As depicted graphically in Figure 6.3, our system consists of an objective lens, a polarizer, an x-cube prism, a reflective SLM (e.g. F-LCOS, DMD), a microdisplay as an image source, and an eyepiece.
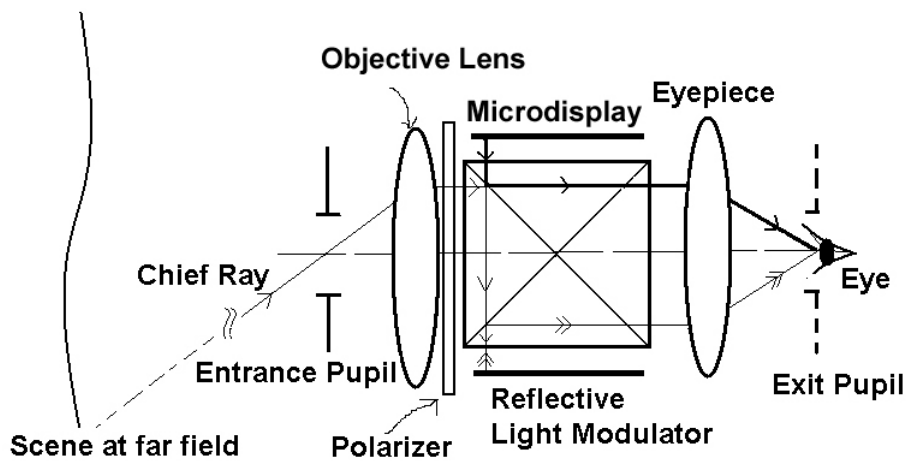
Figure 6.3 Optical Layout of the Mutual Occlusion System

The objective lens collects the light from the scene, in a telecentric manner, and delivers it to the SLM for modulating the content of the scene (pass or block light from the scene). A lens is telecentric in a given space if the chief rays are parallel to the optical axis. We will refer to the image on the SLM as an intermediary image since this image lies between the object and image spaces. The input light to the SLM needs to be linearly polarized; therefore, we use a polarizer after the objective lens. The x-cube prism directs the light imaged by the objective lens to the SLM. The SLM for this application can be modeled as a perfect flat mirror with a quarter wave-plate (i.e. an optical device that rotates polarization by 90 degrees in the case of double pass). The spatial light modulator allows for pixilated control of the illumination from the scene as collected by the objective lens. Once the scene is modulated according to the application needs, the eyepiece will take the modulated light output and image it back to the original scene. An x-cube prism directs the light from the SLM output to the eyepiece. If the image lies at the same plane as the object, the system is a called a Bravais system. In addition to the Bravais constraint, our system has the additional requirement that the objective should form a real intermediary

153

image. In the case of finite conjugates, the objective lens will image the scene beyond its focal point. In order to create a virtual image, the intermediary image will need to reside inside the focal length of the eyepiece. The final image should have unit magnification. The height of the intermediary image is fixed by the height of the SLM, therefore if the focal length of the eyepiece matches the focal length of the objective, the field of view for both the objective and the eyepiece will be exactly the same. Matching the field of view of the objective with the field of view of the eyepiece will ensure 1:1 magnification. An x-cube prism combines the modulated output from the scene with the output of the microdisplay. Compactness is achieved through the use of polarization optics. The system is capable of having very close to zero distortion.

The human eye is conjugated to the entrance pupil, which will cause a viewpoint offset on the order of 50 mm. A simple stereo depth resolution model estimates that over ~ 3.3m, this offset will not be perceived by the eye. This estimate uses a stereo depth resolution value of ½ arcminute for the human eye. When we consider the LCOS pixel size setting the acuity limit, the result shows that after ~6m this offset will not be perceived by the eye.

The entrance pupil must be placed at the focal point of the objective to force the chief rays in the intermediary image space to be parallel to the optical axis, in order to make the system telecentric in that space. Whenever the chief rays cross the optical axis, we get a pupil in that space. Therefore, after the telecentric intermediary space, the chief ray will focus down at the focal point of the eyepiece. This will form the exit pupil of the system. Ideally, to accommodate people wearing eyeglasses, the exit pupil should be located around 25 mm out from the last surface.

An exit pupil diameter of 9-12 mm is desirable, within which the pupil of the eye (i.e., 2-4 mm in diameter) will be located and naturally move within + -20 degrees. The larger the pupil of the instrument compared to the eye pupil, the larger the eye motions allowed. However this occurs at the expense of compactness. We propose a system with a 9 mm pupil to allow the 21 degrees natural eye motion.

First order layout of the system is similar to that of a Keplerian telescope operating at finite conjugates as shown in Fig. 6.4. The problem as specified here is an instance of the two component design problem. In our system, a meaningful object distance is slightly larger than an average arm's length, about 2 meters. The separation between the two lenses is constrained by the size of the x-cube prism. We also demand a telecentric system in the intermediary image space. Telecentricity is required mainly for efficient modulation on the liquid crystal based modulator as well as to reduce vignetting. We show that the same lens can be used as the objective and the eyepiece.
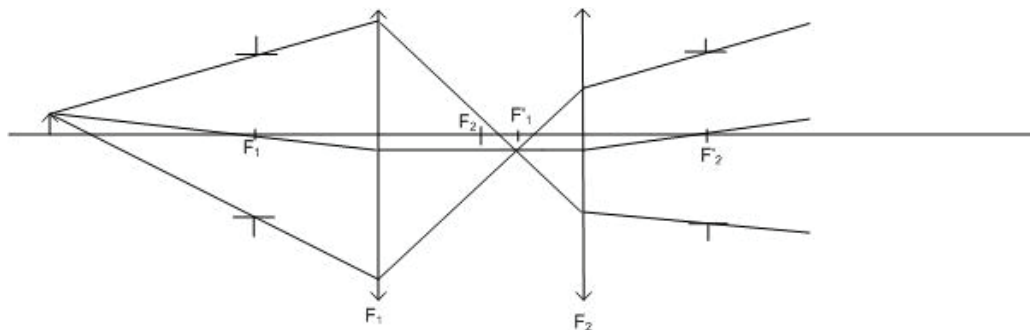
Figure 6.4 Geometry of a unit magnification Bravais system with a real intermediary image

In order to verify the image orientation, we need to consider the system shown in Fig. 6.4 to take into account the effect of reflection from a mirror on orientation (e.g., can't use the

unfolded system in this case). The diagram pertinent to verifying image orientation is shown in Fig 6.5 a-c. The object is indicated with an upright arrow and it is assumed to have an initial upright orientation. The object is first imaged through the objective lens and has an inverted orientation as indicated in orientation at step "1" with a solid line shown in Fig. 6.5 (a). Due to the polarizer, right after the lens, the light will be s-polarized, therefore, it will hit the s-reflect coating in the x-cube prism. The orientation upon reflection is shown in step "2" represented in Fig.6.5 (b) as a solid black line close to the SLM. The SLM will reflect the image and change the polarization, assuming the pixel is "turned on". Caused by this change of polarization, the light will now be p-polarized and therefore hit the p- reflect coating on the x-cube and will be directed towards the eye as shown in Fig. 6 (c). The orientation after the p-reflect mirror is shown in step "3" of Fig. 6c, the final step in the analysis. We can clearly verify that the final image will have an upright orientation. However, unfortunately, due to 3 reflections (odd number), there will be a left to right reversal of the original scene, which remains as one of the challenges to overcome with this layout. A fiber bundle could possibly rotate the image left-to-right orienting the final image correctly.

(a)

(b)

(c)
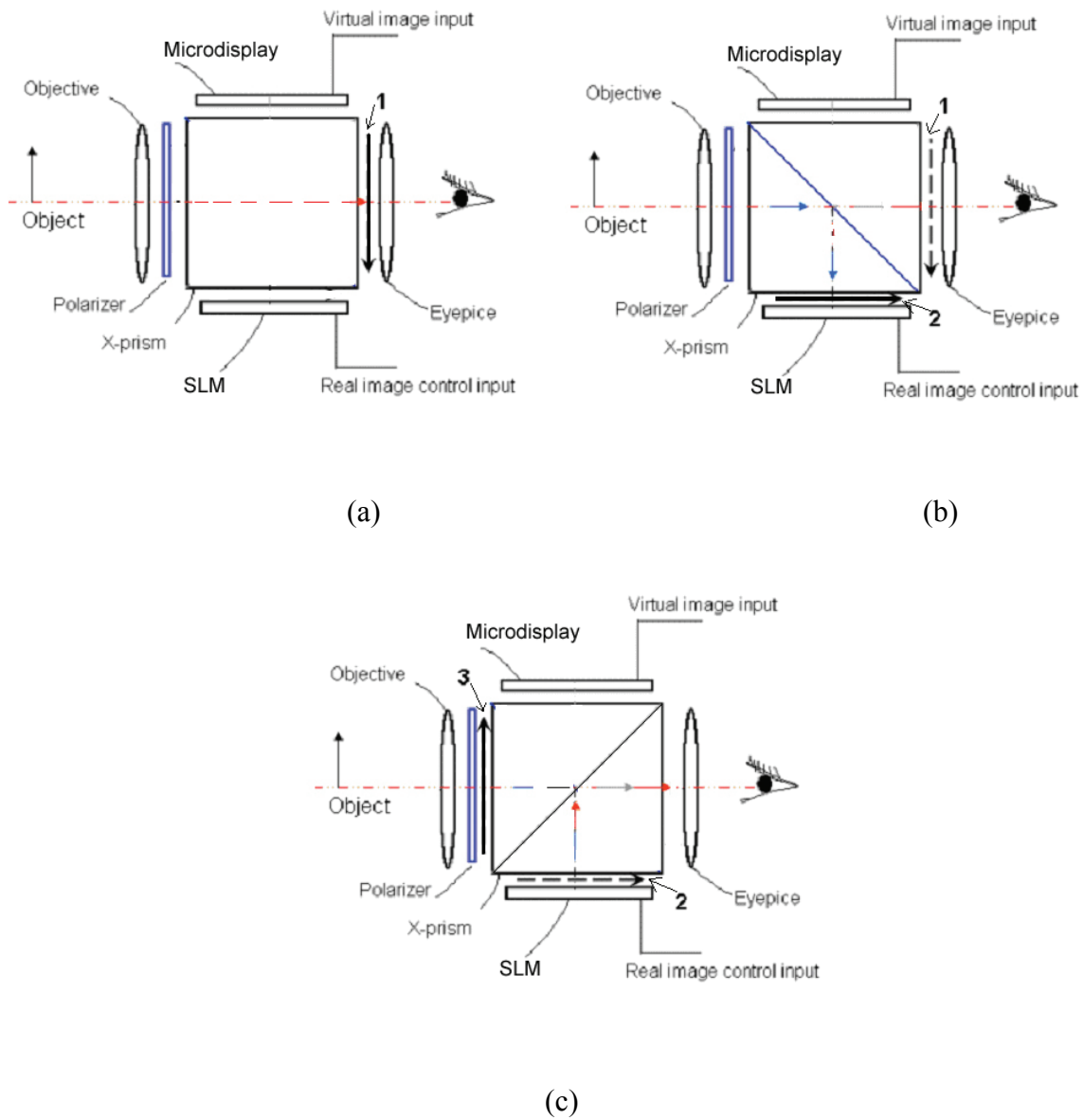
Figure 6.5 Image Orientation

The focal length of the first lens shown in Fig. 2 can be calculated if the object and image distances are known. A meaningful object distance for our system is 2m. The image distance depends on the size of the x-cube prism and can be estimated. The separation between the two lenses depends on the minimum size of the x-cube. The size of the x-cube prism depends on the f/number, the separation distance between the x-cube prism and the image plane, the height of

the marginal ray for the edge field (taking distortion into account), and the refractive index of the cube. Due to the telecentricity constraint in the intermediary space, ideally, the height of the chief ray at the exit face of the x-cube would be equivalent to the size of the detector. We can estimate the angle α' by applying the Lagrange invariant at a pupil plane and using the invariant at the modulator plane to solve for the angle α'. For large object distances, the angle α' for on-axis and maximum field angle will be approximately equal. Applying the Lagrange invariant at the pupil plane yields

$$H = -\theta_{1/2} \cdot EPD / 2 \tag{6.1}$$

At the image plane, the Lagrange invariant yields

$$H = u' \cdot h_{LCOS} \tag{6.2}$$

Based on this marginal angle, we can estimate the x-cube length as in equation 6.3, using the geometry shown in Fig. 6.6.

$$t = \frac{d_2 \cdot \tan(\alpha') + h_{LCOS}}{\frac{1}{2} - \tan(\theta_1)} \tag{6.3}$$

where d_2 is the distance between the exit face of the x-cube polarizing prism and the plane of the reflective spatial light modulator, α' is defined as $\dfrac{1}{2 \cdot F^{\#}_{working}}$, h_{LCOS} is the height of the marginal ray for the edge field, θ_1 is the marginal ray angle for the edge field inside the x-cube

polarizing prism. Using the Lagrange invariant, α' in the intermediary image plane can be calculated since we know the diagonal height of the LCOS in that plane.
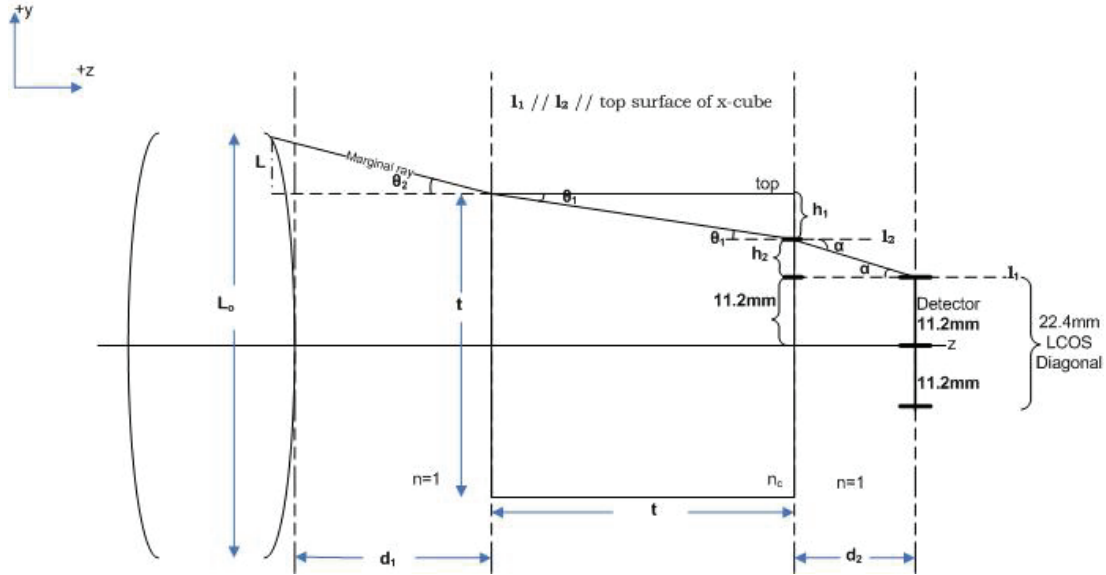
Figure 6.6 Geometry for the x-cube minimum size estimation

The parameters that can be set in this design are the field of view and the entrance pupil diameter. We start with a design choice of a 40 degree full field angle and a 9 mm exit pupil diameter. Based on equation 6.1, 6.2 and 6.3, we calculate the Lagrange invariant to be $\frac{\pi}{2}$, α' to be 0.14025, and the thickness of the x-cube to be ~28 mm. Assuming a refractive index of 1.7 for the glass used in the x-cube prism, the reduced thickness is calculated to be ~19 mm.

A minimum number for the focal length would be 22 mm for this system with 1.5 mm airspace added on each side to ease assembly tolerances. However, it would be advantageous to

consider adding more airspace than the minimum 1.5 mm on each side, simply to keep the system to reasonable angles of incidences to achieve good image quality. An upper limit on the additional airspace length is imposed by the maximum acceptable eyerelief distance. Therefore, we set the focal length of the objective to be 30 mm, resulting in an f/3.3 system. We should note that in a realistic system with thick lenses, the focal length is measured from the principal plane, which does not necessarily lie on the first surface of the system, and in our current implementation the principal planes lie within the lenses, therefore, we were able to achieve a 23 mm eye relief as measured from the edge of the first surface while maintaining telecentricity and the 30 mm focal length.

Finally, we check whether the telecentric system makes use of the maximum height on the optical modulator, by multiplying the field of view with the computed focal length, which yields an 11 mm height on the modulator. In our specific case, the light modulator has a height of 11.2 mm; therefore, a maximum height for the chief ray at the edge field angle of 11 mm is acceptable.

Due to +1 angular magnification at the pupils, we can conclude that the chief ray emerges from the system undeviated. We would expect no distortion and lateral color for a strict implementation of the telecentricity condition. Also, the sine condition is satisfied, if the system can be designed free of spherical aberration, we would expect the system to have no coma as well. Wetherell discusses the Gaussian analysis of afocal lenses, in the case of both finite and infinite conjugates in greater depth [214].

In Section 3.1.1, we compared the use of doublets versus diffractive optical elements on aspheric substrates in correcting chromatic aberrations and quantified the impact of eyerelief on

the design performance. Our study has shown that the diffractive optical element based system can support an eyerelief of up to 80% of its effective focal length. We have also seen that the diffractive optical element based systems yield better correction of both lateral color and astigmatism, the later being a main limiting aberration of large field of view eyepiece designs. Guided by these results, our starting point can be considered as a system resembling an Erfle eyepiece where one doublet is replaced by a singlet and the second doublet is replaced by a diffractive optical element. The optical layout of the objective lens is shown in Fig. 6.7 (a). The minimum feature size for the DOE is measured to be 11.5 μm, which can be considered feasible to fabricate. The performance of the lens, measured with the modulation transfer function, is shown in Fig. 6.7 (b). The modulation transfer function makes the astigmatic nature of the system apparent.
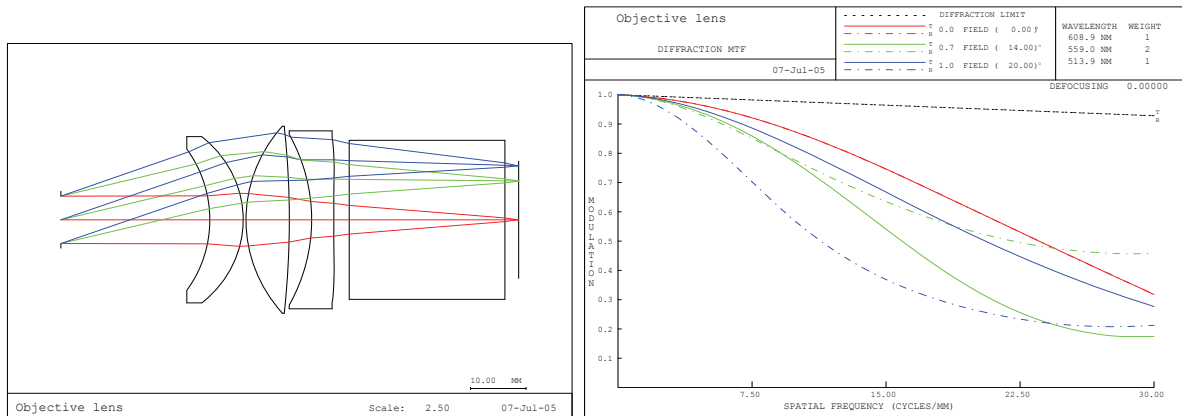
Figure 6.7 (a) Optical Layout of the Objective Lens (b) Performance of the Objective Lens

The system composed of the objective lens and the eyepiece is shown in Fig. 6.8. Currently, the system is modeled in transmission. Therefore, the ray trace indicates a negative

magnification. The performance of the combined system is measured to have an average 6 arcminutes of RMS spot size across the fields.
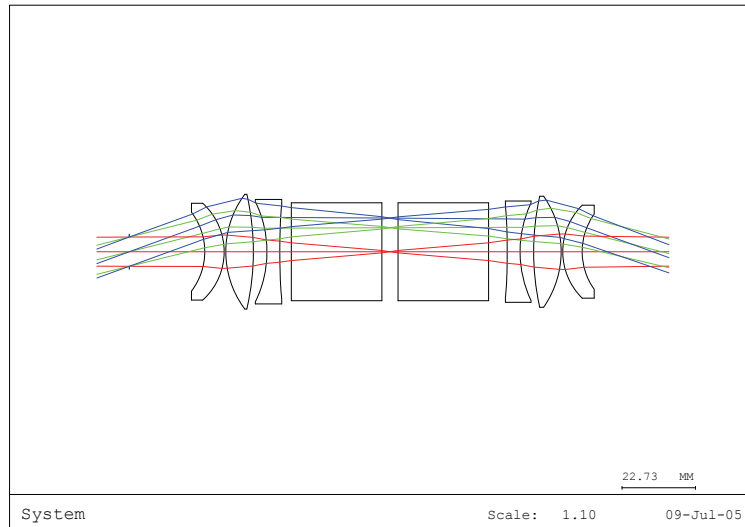
Figure 6.8 Optical Layout of the Combined System

We achieved our goal of close to zero distortion in the combined system as shown in Fig. 6.9. In the current implementation the eyepiece is a flipped about the y-axis version of the objective.
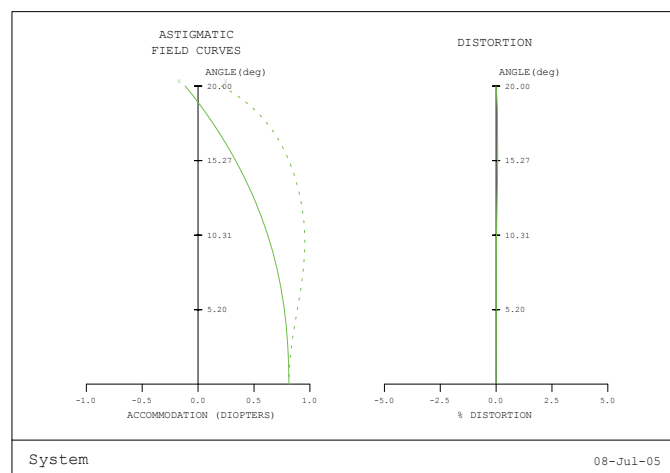
Figure 6.9 Field curves for the combined objective and the eyepiece system

6.3 Preliminary Experiments in Generating Opaque Overlays

Before building custom optics, using the design presented in Section 6.2, we instrumented a prototype, shown in Fig. 6.10, with commercially available components to demonstrate feasibility of the optical approach to occlusion. Listed in the order of light propagation path, starting at the left and propagating to the right, a white light source, a diffuser sheet, a transparency as the object, an achromatic lens (i.e. doublet), a Moxtek polarizing beam splitter, a liquid crystal optical shutter and an F-LCOS from CRL Opto connected to the driver electronics. Note that the F-LCOS was placed physically underneath the shutter. The user or a camera taking a picture of the modulated scene would look through the image formed by Lens 2. Lens 2 is acting as a weak magnifier.

The rationale for choosing the transparency instead of a view of the laboratory as seen through the optics is to increase the contrast ratio of the image formed by the objective. With an enclosure, the image formed by the objective will not compete with the ambient light within the laboratory and the laboratory scene can be used as the object instead of the transparency.

The achromat formed a fairly good image on the F-LCOS. It is clear however that a custom designed optical system as proposed in section 6.2 will provide an image quality to full specification across the full FOV. The quality of the image formed was acceptable to the human eye, while it was difficult to sharply focus on the image while taking a picture. The essential test is the light blocking and passing capability of the whole system.

Fig. 6.11 shows the original object, a Monet painting, that was used in the prototype setup. This image was copied from a book onto a transparency in color. We used the area around the child in the painting as the object. The transparency was fixed between a polarization preserving diffuser sheet from Microsharp and the light source. This diffusing sheet ensured that the object was illuminated uniformly, to the best of our ability within our preliminary setup. We used a white light source from American Optical. The object (i.e., Monet painting) was imaged through the achromatic lens of focal length of 50 mm. The output of the lens went to a polarizing beam splitter, fixed at a 45 degrees angle, then onto the liquid crystal shutter and finally to the F-LCOS.

We would like to highlight the nature of operation of our F-LCOS module since it has a direct impact on the results presented here. In addition, the nature of the F-LCOS operation will make clear the need for the liquid crystal shutter. The F-LCOS operates with a field sequential scheme.

For a typical application, this means cycling through the red, green and blue pixels. Each cycle consists of displaying one color (for example, red) and providing telecentric illumination that matches the color displayed at that moment. Additionally, the F-LCOS uses a 50:50 drive scheme. This means that only half the time allocated for a pixel is displayed on the display/SLM and the other half is used by driving the display/SLM with the negative of the pixel. In our application we are interested in blocking/passing parts of the scene.
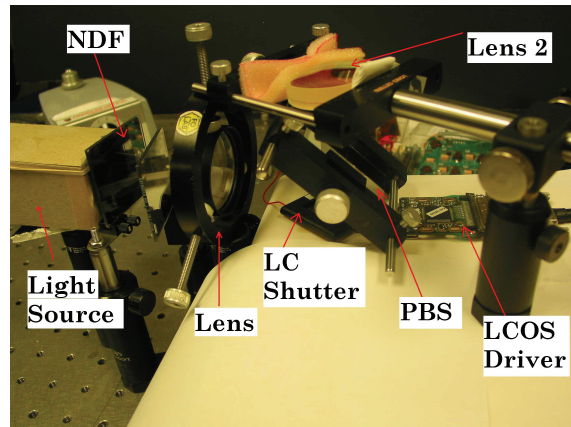
Figure 6.10 Experimental setup on the optical bench

Therefore the system operates in a black and white mode. In a 50:50 drive scheme, the display will show a "white" pixel for half the pixel cycle and a "black" pixel for the rest of the cycle. A 50:50 drive scheme is required to DC balance the liquid crystal to avoid charge migration. The 50:50 drive scheme also implies that a shutter will be necessary to block the negative image. Due to the persistence of the human visual system, the positive and negative cycles would be integrated in the eye, leading to a very low contrast (1:2) image without the shutter.

Depending on the orientation of the polarizers with respect to the F-LCOS module, one color will pass all (for example, white) the light and whereas the other color (for example, black) will block the light. The input light to the display must be linearly polarized. The simplest model for the F-LCOS is an electronically controllable quarter-wave plate laying on top of a perfectly flat mirror. If the pixel is turned off or "black" the orientation of the half-wave plate is aligned to the polarization, with no rotation of the polarization occurring. Thus the output light would be blocked by the analyzer. If the pixel is turned on or "white" the halfwave plate would be rotated

165

at 45 degrees (ideal) to the input polarization, and thus the resulting output polarization of the light would be at 90 degrees to the input. This light would then transmit through the analyzer. In practice, the display does not reach the 45 degrees level of performance, and thus the light is not fully rotated to 90 degrees at the output, leading to a loss in throughput.

In order to reduce vignetting, we are using two Moxtek PBSs side by side. The image is modulated on the F-LCOS, changes polarization, is reflected back, and is transmitted through the beam splitter. This is our basic setup. However, in order to take a picture of this setup, we are using a second lens to image the output of the F-LCOS. Fig. 6.12 (b) is a photograph of the optical image as would be seen through the head-worn display, with no modulation (no occlusion) on the original scene. For comparison purposes, Fig. 12a is a photoshop scaled version of the region of interest shown in Fig. 6.11; therefore, it looks slightly pixilated. In the basic setup, we are imaging a relatively small field of view and also lens 2 is hardly magnifying the image. The significance of the result is that we can form an optical image of the scene on the F-LCOS and modulate it for occlusion.

Figure 6.11 The 'scene' used as an object while testing our head-worn display prototype (area around the child is used in the prototype, note the rectangle that indicates this area)

Fig. 6.12 (c) shows the mask signal that will modulate the scene. Fig. 6.12 (d) shows the an image of the mask seen through the lens 2 on the F-LCOS superimposed on the image and in best focus we achieved (within the digital camera capability that we used to take the pictures). We can observe that head of the child is blocked according to the mask, which can have practically any shape and can be updated at video rates. This first result, which points to the promise of this new technology, also points to the need to further work on the engineering aspects of the system to improve the contrast ratio of the mask that appears to be scene illumination dependent. Furthermore, the line going across the middle of the image is the contact point of the polarizing beam splitters. The two PBSs were placed side by side to cope with the fact that we did not have a large enough beam splitter from the off the shelf components. Also, importantly in assessing the results, the reason that the background appears blurry is because the mask perceptually currently appears in a different plane than the modulated image of the scene. This was confirmed using parallax.

Due to a couple of millimeters discrepancy between the image and the mask map centimeters in visual space after the eyepiece (i.e. lens2 in Fig. 6.10), the camera cannot at this time capture both in focus at the same time. So we focused the camera on the image first to get Fig. 6.12 (b), and on the mask next to get the occlusion image shown in Fig. 6.12 (d). The images shown in Fig. 6.12 (b) and Fig 6.12 (d) were taken with an f/22 setting on the camera to try to increase depth of field to the maximum value available for the camera.
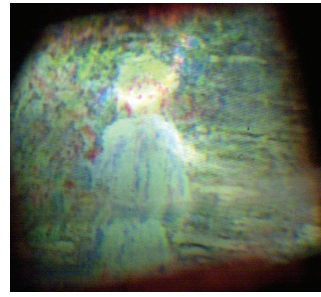
(a)   (b)
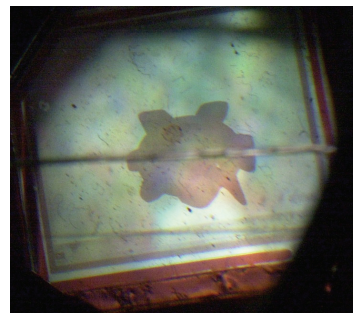
(c)   (d)

Figure 6.12 (a) Original FOV of the optics (b) The scene imaged on the F-LCOS through the

achromat lens (c) The mask image displayed on the F-LCOS module to occlude the face of the

child (d) The output of the 'scene' imaged onto the F-LCOS through the achromat lens and

modulated with the signal shown in Fig. 6.12 (c)

CHAPTER 7 CONCLUSION AND SUMMARY OF CONTRIBUTIONS

The research presented in this dissertation comprises three major contributions: 1) The development of compact head-worn display geometries leveraging free-form optics in the form of single and dual element off-axis designs that have potential for low-cost mass-manufacturing [27, 30, 32, 34]; 2) An alternative surface representation for free-form optical surfaces in imaging optics, based on radial basis functions [27, 29, 36], that have been shown to yield about a 20% increase in the MTF in the design of a single surface mirror; 3) The preliminary design of an optically-based occlusion display [28, 38].

Compact and lightweight optical designs achieving acceptable image quality, field of view, eye clearance, eyebox diameter, operating across the visible spectrum, are the key to the success of next generation head-worn displays. There have been several approaches in the design of head-worn displays including holographic optical elements and laser scanner systems. Our approach has been to investigate the field of view, eyebox size, and performance limits of single and dual-element magnifiers. The use of free-form optical surfaces enables compact head-worn display geometries that were unexploited prior to this thesis. The first part of this thesis in Chapter 4 reported on the design, fabrication and analysis of off-axis magnifier designs. The first design is catadioptric and consists of two elements and is discussed in Section 4.1. The lens utilizes a diffractive optical element and the mirror has a free-form surface described with an x-y polynomial. The use of a diffractive optical element for color correction was justified by a comparison of color correction using doublets as summarized in Section 3.1.2. The dual-element design has an 8 mm diameter eyebox diameter, 15 mm eye clearance, 20 degree diagonal full field of view, and was designed to operate across the visible spectrum between 450-650 nm. 20%

MTF at the Nyquist frequency with less than 4% distortion has been achieved in the dual-element head-worn display. An ideal solution for a head-worn display would be a single surface mirror design. A single surface mirror can be made see-through by machining the appropriate surface shape on the opposite side of a substrate to form a zero power shell. The second design was discussed in Section 4.2, and consisted of a single off-axis free-form mirror described with an x-y polynomial, which achieved a 3 mm diameter exit pupil, 15 mm eye relief, and a 24 degree diagonal full field of view. The second design achieved 10% MTF at the Nyquist frequency set by the pixel spacing of the microdisplay with less than 3% distortion. Both designs have been fabricated using diamond turning techniques.

The second part of this thesis, in Chapter 5, addressed the question of "what is the optimal surface shape for a single mirror constrained in an off-axis magnifier geometry with multiple fields?" The choice of polynomials from the point of view of surface description can be challenged. The limitations of using multivariate polynomials are stated in a theorem due to Mairhuber and Curtis from approximation theory. State-of-the-art description of free-form optical surfaces today relies dominantly on multivariate polynomials and cubic Bezier curves, which is a special case of the non-uniform rational basis splines (NURBS). The potential of radial basis functions in describing free-form optical surfaces had been known prior to this dissertation and proposed in a conference paper by Chan et al. [40]. However, the Chan et al. paper does not contain a single optical system that applies radial basis functions and investigates the consequences in an imaging system [40]. To our knowledge, this is the first instance of optical design examples, in the context of single and dual-element magnifiers, where the surfaces have been described using radial basis functions. We compared the polynomial descriptions to radial basis functions using the MTF criteria. The benefits of using radial basis functions for

surface description include, for example, the performance increase measured by the MTF, or the ability to increase the field of view or pupil size. Even though Zernike polynomials are a complete and orthogonal set of basis over the unit circle and they can be orthogonalized for rectangular or hexagonal pupils using Gram-Schmidt, taking practical considerations into account, such as optimization time and the maximum number of variables available in current raytrace codes, for the specific case of the single off-axis magnifier with a 3 mm diameter eyebox, 15 mm eye relief, 24 degree diagonal full field of view, we found the Gaussian radial basis functions to yield an 18.5% gain in the average MTF at 17 field points compared to a Zernike polynomial up to and including 10th order terms. The linear combination of radial basis function representation is not limited to circular apertures. In fact, any aperture shape can be accommodated using the radial basis function approach. Additionally, the radial basis functions do not rely on an underlying mesh.

A compact solution to the mutual occlusion problem as detailed in Chapter 6 remains a research challenge. A preliminary approach using polarization was discussed in Chapter 6. The preliminary design suggested in Chapter 6 suffers from the difficulties in manufacturing the x-cube prism as well as the incorrect reversal of the image in the left to right direction caused by an odd number of reflections. Image orientation problem could potentially be overcome using a fiber bundle to flip the image in the left to right direction.

Other contributions of the research reported in this dissertation include 4) A study of the pupil size, field of view, and performance limits of the specific off-axis magnifiers geometries considered in this thesis [37]; 5) A comparison of doublets versus diffractive optical elements in

eyepiece design [31]; 6) An application of full-field displays from nodal aberration theory to the analysis and optimization of the dual-element display [37].

Given the specific geometry for a dual-element magnifier discussed in Section 4.2, we estimate the eyebox diameter limit of this geometry to be 12 mm with a 20 degree diagonal full field of view. We estimate the field of view limit to be 25 degrees diagonal full field of view given an 8 mm eyebox diameter.

The impact of the eye clearance requirement on the performance of eyepieces utilizing doublets or diffractive optical elements on aspheric substrates was studied. The doublets were designed to be cemented on-axis elements. Specifically, four different values of eye clearance were implemented: 17, 20, 23, and 26 mm. For each value, axial and lateral color, spherical aberration, coma, astigmatism, field curvature, and distortion were compared. Each system under comparison was optimized for the same focal length, a 9 mm eyebox, photopic wavelengths (513-608 nm), and a 40 degree full field of view. It was found that the single-layer diffractive optical element supports an eye clearance of approximately 80% of the effective focal length, while the doublet-based solution drops below the desired specifications at approximately 65% of the effective focal length.

Visualization tools such as field map plots provided by nodal aberration theory have been applied during the analysis of the off- axis systems discussed in this thesis. Full-field displays are used for node placement within the field of view for the dual-element head-worn display. The judicious separation of the nodes along the x-direction in the field of view results in well-behaved MTF plots. This is in contrast to an expectation of achieving better performance through restoring symmetry via collapsing the nodes to yield quadratic astigmatism.

A contribution not discussed in this thesis include the efficiency analysis (i.e., number of spots per element, see [35] for the calculation of the number of spots) of three thousand one hundred and eighty eight (3188) optical systems sampled out of the Lensview® database [35].

7.1 Future Work

Unsolved problems in head-worn displays include compact head-worn displays capable of occlusions [98], multi-focal head-worn displays [161, 167, 208], and eye tracking integration in head-worn displays [202].

In terms of the surface representation aspects of this work, this thesis primarily explored the $R^2 \to R$ setting. The radial basis functions scale to surface design problems in higher dimensions; future work may explore applications of radial basis functions in different settings and application areas, for example, in illumination applications.

Our work was not optimized for numerical efficiency. Future work may further optimize numerical efficiency through using compactly supported radial basis functions such as the Wendland or the Wu function. As problem sizes get larger and the number of basis increases, future work may benefit from using the Fast Gauss Transform (FGT) [74]. Unfortunately, the benefits of the Fast Gauss Transform are not substantial for small problem sizes (<300 basis functions) compared to the naïve evaluation. This thesis explored primarily the Gaussians as the choice of the basis; future work may investigate multiquadrics and the thin-plate splines. There are tools, such as leave-one-out-cross validation, that assist in choosing optimal shape parameters

in the case where the target function is known beforehand. However, choosing an optimal shape factor for an unknown target function remains to be solved.

APPENDIX EQUIVALANCE OF A DIFFRACTIVE OPTICAL ELEMENT AND A REFRACTIVE OPTICAL ELEMENT

This appendix summarizes the equivalence between a diffractive optical element and a refractive lens, similar to the approach taken in [188]. To illustrate the equivalence, we will use the function $f(x) = -x^2 + x + 3$ plotted in Fig. A1 (left). This specific *f(x)* is chosen solely to visualize the steps and the arguments hold for an arbitrary phase function. The diffractive phase can be obtained by slicing the refractive phase into pieces of some height α. This can be achieved by applying the mod function to the refractive phase with a period p and multiplying it by α, we illustrate the analysis for *p equals 1*, however, we shall show that the arguments hold for any *p*. The choice of α will be informed by the fabrication constraints. For example, if we desire one wavelength of OPD, α can be approximately set to one micrometer, see equation 3.5. The mod function can be defined as

$$\mathrm{mod}(f(x), p) \equiv f(x) - p \left\lfloor \frac{f(x)}{p} \right\rfloor$$

where $\lfloor . \rfloor$ denotes the floor function. The mod function applied to $f(x) = -x^2 + x + 3$ with period p equals 1 is plotted in Fig. A1 (center). Fig. A1 (center) is the diffractive phase profile before multiplying with the scalar α. Plotting the diffractive phase profile, i.e., $\mathrm{mod}(f(x) = -x^2 + x + 3, 1)$ against the refractive phase, i.e., $f(x) = -x^2 + x + 3$, we observe that the diffractive phase is periodic in the refractive phase with a period of 1. Therefore, we can write the diffractive phase as a Fourier series in the refractive phase

$$e^{i2\pi\varphi'(x)} = \sum_{m=-\infty}^{\infty} c_m e^{+i2\pi m\varphi(x)}$$

176

where $c_m = \dfrac{1}{T}\displaystyle\int_t^{t+T} f(x)e^{-i\frac{2\pi}{T}mx}dx$ [68].

Note that once we multiply the diffractive phase with the scalar α, within a period, $\varphi'(x)$ equals $\alpha\varphi(x)$. The relationship $\varphi'(x)$ equals $\alpha\varphi(x)$ holds independent of the choice of a period in the mod function and justifies our choice of period 1 for the current argument. For a period T equals 1, and t equal $-\dfrac{1}{2}$, using the $\varphi'(x)$ equals $\alpha\varphi(x)$ relationship, the integral for the coefficients becomes

$$c_m = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(\varphi'(x)-m\varphi(x))}d\varphi(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(\alpha-m)\varphi(x)}d\varphi(x)$$

Using Euler's identity, the integral for the coefficients is evaluated to be

$$c_m = \mathrm{sinc}(\pi(\alpha-m))$$
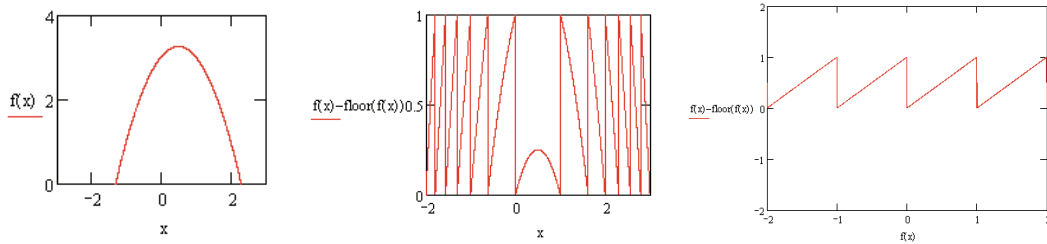
Fig. A1 (left) Plot of the $f(x)=-x^2+x+3$ function (center) The mod function applied to f(x) with a period 1 (right) Plot of the truncated mod function (i.e., diffractive phase) against the f(x) (i.e., refractive phase)

We observe for the first order (m=1), and when α equals 1, the diffractive and the refractive phases are equal.

REFERENCES

1.   "Computerized Anthropometric Research & Design Lab, Warfighter Interface Division, Human Effectiveness Directorate, Air Force Research Labs. http://www.hec.afrl.af.mil. Last visited on Sep 25, 2008.."

2.   M. B. Allen, and E. L. Isaacson, *Numerical Analysis for Applied Science* (Wiley-Interscience, 1997).

3.   T. Ando, K. Yamasaki, M. Okamoto, and E. Shimizu, "Head-mounted display using a holographic optical element," in *Practical Holography XII*, SPIE Vol. 3293, (SPIE, San Jose, CA, USA, 1998), pp. 183-189.

4.   Y. Argotti, L. Davis, V. Outters, and J. Rolland, "Dynamic superimposition of synthetic objects on rigid and simple-deformable real objects," Computers & Graphics **26**, 919-930 (2002).

5.   D. Armitage, I. Underwood, and S.-T. Wu, *Introduction to microdisplays* (Wiley, Chichester, England ; Hoboken, NJ, 2006).

6.   D. A. Atchison, and G. Smith, *Optics of the human eye* (Butterworth-Heinemann, Oxford ; Boston, 2000).

7.   T. M. Aye, K. Yu, I. Tengara, P. W. Peppler, and B. J. Pierce, "Compact HMD optics based on multiplexed aberration-compensated holographic optical elements," in *Helmet- and Head-Mounted Displays VI*, SPIE Vol. 4361, (SPIE, Orlando, FL, USA, 2001), pp. 89-97.

8.   R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," IEEE Computer Graphics and Applications **21**, 34-47 (2001).

9.   W. Barfield, and T. Caudell, *Fundamentals of wearable computers and augmented reality* (Lawrence Erlbaum Associates, Mahwah, NJ, 2001).

10.   A. H. Barr, "Superquadrics and Angle-Preserving Transformations," Computer Graphics and Applications, IEEE **1**, 11-23 (1981).

11.   M. Bauer, T. Heiber, G. Kortuem, and Z. Segall, "A collaborative wearable system with remote sensing," in *Second International Symposium on Wearable Computers* (1998), pp. 10-17.

12.   A. Becker, "Miniature video display system," U.S. Patent 4,934,773 (1990)

13.   S. Becker, C. Bowd, S. Shorter, K. King, and R. Patterson, "Occlusion contributes to temporal processing differences between crossed and uncrossed stereopsis in random-dot displays," Vision Research **39**, 331-339 (1999).

14.   A. L. Berman, and J. E. Melzer, "Optical collimating apparatus," U.S. Patent 4,859,031 (1989)

15.   M. Billinghurst, S. Weghorst, and T. Furness, III, "Wearable computers for three dimensional CSCW," in *First International Symposium on Wearable Computers* (1997), pp. 39-46.

16.   O. Bimber, and R. Raskar, *Spatial augmented reality : merging real and virtual worlds* (A K Peters, Wellesley, Mass., 2005).

17.   F. Biocca, and J. P. Rolland, "Virtual eyes can rearrange your body: adaptation to visual displacement in see-through head-mounted displays," Presence **7**, 262-278 (1998).

18.   A. N. Boev, K. N. Fountas, I. Karampelas, C. Boev, T. G. Machinis, C. Feltes, I. Okosun, V. Dimopoulos, and C. Troup, "Quantitative pupillometry: normative data in healthy pediatric volunteers," J. Neurosurg. **103**, 496-500 (2005).

19.   M. Born, and E. Wolf, *Principles of optics : electromagnetic theory of propagation, interference and diffraction of light* (Cambridge University Press, Cambridge, UK ; New York ;, 1997).

20.   B. Braunecker, R. Hentschel, and H. Tiziani, *Advanced optics using aspherical elements* (SPIE, 2008).

21.   R. Buchroeder,"Tilted component optical systems," Ph.D. thesis, University of Arizona (1976)

22.   R. A. Buchroeder, "Distortion correction for oblique projection onto a curved screen," in *Novel Optical Systems Design and Optimization*, SPIE Vol. 2537, (SPIE, San Diego, CA, USA, 1995), pp. 96-104.

23.   R. A. Buchroeder, G. W. Seeley, and D. Vukobratovich,"Design of a catadioptric vcass helmet-mounted display," Thesis Type, University

24.   D. A. Buralli, and G. M. Morris, "Effects of diffraction efficiency on the modulation transfer function of diffractive lenses," Appl. Opt. **31**, 4389-4396 (1992).

25.   D. Burr, "Motion smear," Nature **284**, 164-165 (1980).

26.   G. J. Burton, and N. D. Haig, "Effects of the Seidel aberrations on visual target discrimination," J. Opt. Soc. Am. A **1**, 373-385 (1984).

27.   O. Cakmakci, G. E. Fasshauer, H. Foroosh, K. P. Thompson, and J. P. Rolland, "Meshfree approximation methods for free-form surface representation in optical design with applications to head-worn displays," in *Novel Optical Systems Design and Optimization XI*, 7061, (SPIE, San Diego, CA, USA, 2008), pp. 70610D-70615.

28.   O. Cakmakci, Y. Ha, and J. Rolland, "Design of a compact optical see-through head-worn display with mutual occlusion capability," in *Novel Optical Systems Design and Optimization VIII* SPIE Vol. 5875, (SPIE, San Diego, CA, USA, 2005), pp. 58750H-58756.

29.   O. Cakmakci, B. Moore, H. Foroosh, and J. P. Rolland, "Optimal local shape description for rotationally non-symmetric optical surface design and analysis," Opt. Express **16**, 1583-1589 (2008).

30.   O. Cakmakci, A. Oranchak, and J. P. Rolland, "Dual-Element Off-Axis Eyeglass Based Display," in *International Optical Design*(Optical Society of America, 2006), p. TuC2.

31.   O. Cakmakci, and J. Rolland, "Comparative analysis of doublets versus single-layer diffractive optical elements in eyepiece or magnifier design," Appl. Opt. **46**, 8140-8148 (2007).

32.   O. Cakmakci, and J. Rolland, "Design and fabrication of a dual-element off-axis near-eye optical magnifier," Opt. Lett. **32**, 1363-1365 (2007).

33.   O. Cakmakci, and J. Rolland, "Head-Worn Displays: A Review," J. Display Technol. **2**, 199-216 (2006).

34.   O. Cakmakci, and J. P. Rolland, "Imaging systems for eyeglass-based display devices," U.S. Patent Application # 11/398,277 (2006)

35.   O. Cakmakci, J. P. Rolland, K. P. Thompson, and J. Rogers, "Design efficiency of 3188 optical designs," in *Current Developments in Lens Design and Optical Engineering IX*, 7060, (SPIE, San Diego, CA, USA, 2008), pp. 70600S-70610.

36.   O. Cakmakci, S. Vo, H. Foroosh, and J. Rolland, "Application of radial basis functions to shape description in a dual-element off-axis magnifier," Opt. Lett. **33**, 1237-1239 (2008).

37.   O. Cakmakci, S. Vo, K. P. Thompson, and J. P. Rolland, "Application of radial basis functions to shape description in a dual-element off-axis eyewear display: field of view limit.," Journal of Society of Information Display **16** (2008).

38.   O. Cakmakci, H. Yonggang, and J. P. Rolland, "A compact optical see-through head-worn display with occlusion support," in *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*(2004), pp. 16-25.

39.   C. E. Carr, S. J. Schwartz, and I. Rosenberg, "A wearable computer for support of astronaut extravehicular activity," in *Sixth International Symposium on Wearable Computers* (2002), pp. 23-30.

40.   A. K. Chan, C. K. Chui, and L. T. Guan, "Radial basis function approach to interpolation of large reflecting surfaces," in *Curves and Surfaces in Computer Vision and Graphics*, 1251, (SPIE, Santa Clara, CA, USA, 1990), pp. 62-72.

41.   W. N. Charman, "Optics of the Eye," in *Handbook of Optics*, M. Bass, ed. (McGraw-Hill, New York, 1995).

42.   H. Chase, "Optical design with rotationally symmetric NURBS," in *International Optical Design Conference 2002* 4832, (SPIE, Tucson, AZ, USA, 2002), pp. 10-24.

43.   W. Chen, "Biocular eyepiece optical system employing refractive and diffractive optical elements," U.S. Patent 5,151,823

44.   A. D. Cheok, W. Fong Siew, Y. Xubo, W. Wang, H. Lee Men, M. Billinghurst, and H. Kato, "Game-City: a ubiquitous large area multi-interface mixed reality game space for wearable computers," in *Sixth International Symposium on Wearable Computers* (2002), pp. 156-157.

45.   R. Cormack, "Stereoscopic depth perception at far viewing distances.," Perception and Psychophysics **35**, 423-428 (1984).

46.   C. E. Coulman, and G. R. Petrie, "Some Notes on the Designing of Aspherical Magnifiers for Binocular Vision," J. Opt. Soc. Am. **39**, 612-613 (1949).

47.   J. A. Cox, T. A. Fritz, and T. Werner, "Application and demonstration of diffractive optics for head-mounted displays," in *Helmet-mounted displays and symbology design requirements*, SPIE Vol. 2218, R. J. Lewandowski, W. Stephens, and L. A. Haworth, eds. (1994).

48.   J. Cross, C. Baber, and S. I. Woolley, "Layered annotations of digital images for data collection in the field," in *Seventh IEEE International Symposium on Wearable Computers* (2003), pp. 154-159.

49.   C. Curatu, H. Hua, and J. Rolland, "Projection-based head-mounted display with eye tracking capabilities," in *Novel Optical Systems Design and Optimization VIII*, SPIE Vol. 5875, (SPIE, San Diego, CA, USA, 2005), pp. 58750J-58759.

50.   C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," The Journal of Comparative Neurology **292**, 497-523 (1990).

51.   R. Daude, and M. Weck, "Mobile approach support system for future machine tools," in *First International Symposium on Wearable Computers* (1997), pp. 24-30.

52.   T. Davenport,"Creation of a uniform circular illuminance distribution using faceted reflective NURBS surface," Ph.D. thesis, University of Arizona (2002)

53.   P. J. Davis, *Interpolation and Approximation* (Blaisdell Publishing Company, Washington D.C., 1963).

54.   D. C. Dilworth, "Extreme wide angle eyepiece with minimal aberrations," U.S. Patent 4,720,183 (1988)

55.   U. S. DOD.,"Military handbook 141, optical design," Thesis Type, University (1962).

56.   J. Dollond, "An Account of Some Experiments concerning the Different Refrangability of Light.," *Philosophical Transactions (1683-1775)* **50** 733-743 (1757).

57.   S. R. Ellis, and U. J. Bucher, "Distance perception of stereoscopically presented virtual objects optically superimposed on physical objects in a head-mounted see-through displays," in *Proc. human factors and ergonomics society* (Nashville, TN, 1994).

58.   D. Faklis, and M. J. Hoppe, "Effects of diffraction on the performance of diffractive relay optics," in *Helmet- and Head-Mounted Displays and Symbology Design Requirements* SPIE Vol. 2218 (1994), pp. 32-40.

59.   G. E. Fasshauer, *Meshfree approximation methods with MATLAB* (World Scientific, 2007).

60.   S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster, "A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment," in *First International Symposium on Wearable Computers* (1997), pp. 74-81.

61.   F. J. Ferrin, "Update on optical systems for military head-mounted displays," in *Helmet- and Head-Mounted Displays IV*, 3689, (SPIE, Orlando, FL, USA, 1999), pp. 178-185.

62.   J. Figoski, "Aberration characteristics of nonsymmetric systems," in *International Lens Design Conference*, SPIE Vol. 554, (1985).

63.   R. Fisher, "Head-mounted projection display system featuring beam splitter and method of making same," U.S. Patent 5,572,229 (1996)

64.   R. Fletcher, *Practical methods of optimization* (Wiley, Chichester ; New York, 1987).

65.   G. W. Forbes, "Shape specification for axially symmetric optical surfaces," Opt. Express **15**, 5218-5226 (2007).

66.   T. Furness, "Virtual retinal display," U.S. Patent 5,467,104 (1995)

67.   K. Garrard, T. Bruegge, J. Hoffman, T. Dow, and A. Sohn, "Design tools for freeform optics," in *Current Developments in Lens Design and Optical Engineering VI*, SPIE Vol. 5874, (SPIE, San Diego, CA, USA, 2005), pp. 58740A-58711.

68.   J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (Wiley, 1978).

69.   P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization* (Academic Press, London ; New York, 1981).

70.   T. Gold, "Visual disparity tolerances for head-up displays," in *Electro-optical system design conference*(Anaheim, CA, 1971), pp. 399-406.

71.   T. Gold, and A. Hyman,"Visual requirements for head-up displays," Thesis Type, S.-R. Corp., University (1970).

72.   J. W. Goodman, *Introduction to Fourier optics* (McGraw-Hill, San Francisco, 1968).

73.   D. M. Green, and J. A. Swets, *Signal detection theory and psychophysics* (Peninsula Pub., Los Altos, CA, 1988).

74.   L. Greengard, and J. Strain, "The Fast Gauss Transform," SIAM Journal on Scientific and Statistical Computing **12**, 79-94 (1991).

75.   Y. Ha, V. Smirnov, L. Glebov, and J. P. Rolland, "Optical modeling of a holographic single-element head-mounted display," in *Helmet- and Head-Mounted Displays IX: Technologies and Applications*, 5442, (SPIE, Orlando, FL, USA, 2004), pp. 254-260.

76.   J. E. Harvey, and R. V. Shack, "Aberrations of diffracted wave fields," Appl. Opt. **17**, 3003-3009 (1978).

77.   D. M. Hasenauer, and J. M. Kunick, "Full-field mapping and analysis of veiling glare sources for helmet-mounted display systems," in *Current Developments in Optical Design and Optical Engineering VIII*, 3779, (SPIE, Denver, CO, USA, 1999), pp. 382-389.

78.   M. J. Hayford, "Optical system design using holographic optical elements," in *Geometrical optics, SPIE Vol. 0531*(1985).

79.   Hazra, and Chatterjee, "A prophylactic strategy for global synthesis in lens design," Opt. Rev. **12** (2005).

80.   R. T. Hennessy, T. Iida, K. Shiina, and H. W. Leibowitz, "The effect of pupil size on accommodation," Vision Research **16**, 587-589 (1976).

81.   H. T. E. Hertzberg, G. S. Daniels, and E. Churchill,"Antropometry of flying personnel," Thesis Type, W.-P. A. F. B. Wright Air Development Center, University

82.   M. Herzberger, and N. R. McClure, "The design of superachromatic lenses," Appl. Opt. , 553-560 (1963).

83.   R. A. Hicks, "Direct methods for freeform surface design," in *Novel Optical Systems Design and Optimization X*, 6668, (SPIE, San Diego, CA, USA, 2007), pp. 666802-666810.

84.   D. Holmgren, and W. Robinett, "Scanned laser displays for virtual reality: a feasibility study," Presence **2**, 171-184 (1993).

85.   H. H. Hopkins, *The wave theory of aberrations* (Oxford press, 1950).

86.   M. Hoppe, and J. E. Melzer, "Optical tiling for wide-field-of-view head-mounted displays," in *Current Developments in Optical Design and Optical Engineering VIII*, 3779, (SPIE, Denver, CO, USA, 1999), pp. 146-153.

87.   H. Hoshi, N. Taniguchi, H. Morishima, T. Akiyama, S. Yamazaki, and A. Okuyama, "Off-axial HMD optical system consisting of aspherical surfaces without rotational symmetry," in *Stereoscopic Displays and Virtual Reality Systems III*, 2653, (SPIE, San Jose, CA, USA, 1996), pp. 234-242.

88.   H. Hua, Y. Ha, and J. P. Rolland, "Design of an ultralight and compact projection lens," Appl. Opt. **42**, 97-107 (2003).

89.   R. B. Huxford, "Wide-FOV head-mounted display using hybrid optics," in *Optical Design and Engineering*, 5249, (SPIE, St. Etienne, France, 2004), pp. 230-237.

90.   K. Inoguchi, M. Matsunaga, and S. Yamazaki, "The development of a high-resolution HMD with a wide FOV using the shuttle optical system," in *Head- and Helmet-Mounted Displays XIII: Design and Applications*, 6955, (SPIE, Orlando, FL, USA, 2008), pp. 695503-695508.

91.   R. E. Jacobson, "Projection screens," in *Applied optics and optical engineering*, Kingslake, ed. (1965).

92.   T. Jewell,"Aberration field properties of simple non-axially symmetric optical systems," M.S. Thesis, University of Arizona (1984)

93.   C. A. Johnson, "Effects of luminance and stimulus distance on accommodation and visual resolution," J. Opt. Soc. Am. **66**, 138-142 (1976).

94.   E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Essentials of neural science and behavior* (Appleton & Lange, Norwalk, CT, 1995).

95.   T. Kawamura, Y. Kono, and M. Kidode, "Wearable interfaces for a video diary: towards memory retrieval, exchange, and transportation," in *Sixth International Symposium on Wearable Computers* (2002), pp. 31-38.

96.   G. Kelly, M. Shenker, and P. Weissman, "Helmet-mounted area of interest," in *Helmet-Mounted Displays III*, 1695, (SPIE, Orlando, FL, USA, 1992), pp. 58-63.

97.   M. J. Kidger, *Fundamental optical design* (SPIE Press, Bellingham, Wash., 2002).

98.   K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods, "An occlusion capable optical see-through head mount display for supporting co-located collaboration," in *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*(2003), pp. 133-141.

99.   D. Knapp,"Conformal optical design," Ph.D. Thesis, University of Arizona (2002)

100.   W. Knapp, G. Blough, K. Khajurivala, R. Michaels, B. Tatian, and B. Volk, "Optical design comparison of 60° eyepieces: one with a diffractive surface and one with aspherics," Appl. Opt. **36**, 4756-4760 (1997).

101.   D. F. Kocian, "Design Considerations for Virtual Panoramic Display (VPD) Helmet Systems," (Armstrong Aerospace Medical Research Laboratory, Visual Systems Branch, Wright-Patterson Air Force Base, Dayton, Ohio, 1988).

102.   R. Kojima, and T. Ojika, "Transition between virtual environment and workstation environment with projective head-mounted display," in *VRAIS '97*(1997), pp. 130-137.

103.   D. Korsch, *Reflective Optics* (Academic Press, San Diego, CA, 1991).

104.   J. A. LaRussa, "Image forming apparatus," U.S. Patent 3,940,203 (1976)

105.   S. H. Lee, *Diffractive and Miniaturized Optics, SPIE Critical Reviews of Optical Science and Technology* (1993).

106.   S. Lerner,"Optical design using novel aspheric surfaces," Ph.D. Thesis, University of Arizona (2000)

107.   L. Li, A. Y. Yi, C. Huang, D. A. Grewell, A. Benatar, and Y. Chen, "Fabrication of diffractive optics by use of slow tool servo diamond turning process," Optical Engineering **45**, 113401-113409 (2006).

108.   S. Liu, D. Cheng, and H. Hua, "An Optical See-Through Head-Mounted Display With Addressable Focal Planes," in *IEEE and ACM International Symposium on Mixed and Augmented Reality* (Cambridge, UK, 2008).

109.   M. Livingstone, and D. Hubel, "Segregation of form, color, movement, and depth: anatomy, physiology, and perception," Science **240**, 740-749 (1988).

110.   E. Ludvigh, and J. W. Miller, "Study of Visual Acuity during the Ocular Pursuit of Moving Test Objects. I. Introduction," J. Opt. Soc. Am. **48**, 799-802 (1958).

111.   V. N. Mahajan, *Optical imaging and aberrations* (SPIE Optical Engineering Press, Bellingham, Wa., 1998).

112.   J. C. Mairhuber, "On Haar's theorem concerning Chebyshev approximation problems having unique solutions," in *Proc. Am. Math. Soc.* , 7, (1956), pp. 609-615.

113.   W. Mandler, "Design of Double-Gauss lenses," in *SPIE* 237, (1980), p. 222.

114.   S. Mann, "An historical account of the `WearComp' and `WearCam' inventions developed for applications in `personal imaging'," in *Wearable Computers, 1997. Digest of Papers., First International Symposium on*(1997), pp. 66-73.

115.   M. Mansuripur, "Distribution of light at and near the focus of high-numerical-aperture objectives," J. Opt. Soc. Am. A **3**, 2086-2093 (1986).

116.   R. Martins, J. P. Rolland, and Y. Ha, "Head-mounted display by integration of phase-conjugate material," U.S. Patent 6,999,239 (2006)

117.   R. Martins, V. Shaoulov, Y. Ha, and J. Rolland, "A mobile head-worn projection display," Opt. Express **15**, 14530-14538 (2007).

118.   A. Matsunaga, "Overview of "Ketai" (mobile phones) services in Japan," in *Ninth IEEE International Symposium on Wearable Computers* (2005), pp. xviii-xvxix.

119.   D. G. McCauley, C. E. Simpson, and W. J. Murbach, "Holographic Optical Element for Visual Display Applications," Appl. Opt. **12**, 232-242 (1973).

120.   J. E. Melzer, "Design evolution of a wide-field-of-view head-mounted display for aircraft training and simulation," in *Helmet- and Head-Mounted Displays VI*, 4361, (SPIE, Orlando, FL, USA, 2001), pp. 24-29.

121.   J. E. Melzer, "Overcoming the field-of-view/resolution invariant in head-mounted displays," in *Helmet- and Head-Mounted Displays III*, 3362, (SPIE, Orlando, FL, USA, 1998), pp. 284-293.

122.   L. Mertz, "Geometrical design for aspheric reflecting systems," Appl. Opt. **18**, 4182-4186 (1979).

123.   G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," The Psycohological Review **63**, 81-97 (1956).

124.   M. D. Missig, and G. M. Morris, "Diffractive optics applied to eyepiece design," Appl. Opt. **34**, 2452-2461 (1995).

125.   K. Moffit, "Designing HMDs for viewing comfort," in *Head-mounted displays: designing for the user*, J. E. Melzer, and K. Moffit, eds. (McGraw-Hill, New York, 1997), pp. 117-142.

126.   G. M. Morris., "Recent advances in diffractive- and micro-optics technology," in *Optika '98, 5th Congress on Modern optics*, Vol. 3573, (1998), pp. 370-375.

127.   R. J. Mortimer, A. L. Dyer, and J. R. Reynolds, "Electrochromic organic and polymeric materials for display applications," Displays **27**, 2-18 (2006).

128.   P. Mouroulis, T. G. Kim, and G. Zhao, "Transverse color tolerances for visual optical systems," Appl. Opt. **32**, 7089-7094 (1993).

129.   P. Mustillo, "Binocular mechanisms mediating crossed and uncrossed stereopsis," Psychological Bulletin **97**, 187-201 (1985).

130.   T. Nakai, "Research on multi-layer diffractive optical elements and their application to photographic lenses.," in *ODF 2002* (Optical Society of Japan, 2002).

131.   K. N. Ogle, and J. T. Schwartz, "Depth of Focus of the Human Eye," J. Opt. Soc. Am. **49**, 273-279 (1959).

132.   Y. Ohta, and H. Tamura, *Mixed reality: merging real and virtual worlds* (Springer-Verlag, New Jersey, 1999).

133.   D. F. Ormerod, B. Ross, and A. Naluai-Cecchini, "Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events," in *Sixth International Symposium on Wearable Computers* (2002), pp. 131-132.

134.   J. Parsons, and J. P. Rolland, "A non-intrusive display technique for providing real-time data within a surgeons critical area of interest," in *Proceedings of Medicine Meets Virtual Reality* (1998), pp. 246-251.

135.   R. Patterson, "Spatiotemporal properties of stereoacuity," Optometry & Vision Science **67**, 123-128 (1990).

136.   R. Patterson, and R. Fox, "The effect of testing method on stereoanomaly," Vision Research **24**, 403-408 (1984).

137.   R. Patterson, and W. L. Martin, "Human stereopsis," Human Factors: The Journal of the Human Factors and Ergonomics Society **34**, 669-692 (1992).

138.   R. Patterson, L. Moe, and T. Hewitt, "Factors that effect depth perception in stereoscopic displays," Human Factors: The Journal of the Human Factors and Ergonomics Society **34**, 655-667 (1992).

139.   R. Patterson, M. Winterbottom, and B. J. Pierce, "Perceptual Issues with the use of head-mounted displays," Human Factors: The Journal of the Human Factors and Ergonomics Society **48**, 555-573 (2006).

140.   E. Peli, "Visual issues in the use of a head-mounted monocular display," Optical Engineering **29**, 883-892 (1990).

141.   A. P. Pentland, M. Petrazzouli, A. Gerega, and T. Starner, "The digital doctor: an experiment in wearable telemedicine," in *First International Symposium on Wearable Computers* (1997), pp. 173-174.

142.   L. A. Piegl, and W. Tiller, *The NURBS book* (Springer, Berlin ; New York, 1997).

143.   C. S. Pinhanez, "The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces," in *UbiComp '01: Proceedings of the 3rd international conference on Ubiquitous Computing*(Springer-Verlag, Atlanta, Georgia, USA, 2001), pp. 315--331.

144.   W. T. Plummer, "Unusual optics of the Polaroid SX-70 Land camera," Appl. Opt. **21**, 196-208 (1982).

145.   W. T. Plummer, J. G. Baker, and J. Van Tassell, "Photographic Optical Systems With Nonrotational Aspheric Surfaces," Appl. Opt. **38**, 3572-3592 (1999).

146.   W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, 2007).

147.   A. Rakich, and N. Rumsey, "Method for deriving the complete solution set for three-mirror anastigmatic telescopes with two spherical mirrors," J. Opt. Soc. Am. A **19**, 1398-1405 (2002).

148.   B. J. Rhodes, "The wearable remembrance agent: a system for augmented memory," in *First International Symposium on Wearable Computers* (1997), pp. 123-128.

149.   A. K. Rigler, and T. P. Vogl, "Spline Functions: an Alternative Representation of Aspheric Surfaces," Appl. Opt. **10**, 1648-1651 (1971).

150.   S. Rippa, "An algorithm for selecting a good value for the parameter c in radial basis function interpolation," Advances in Computational Mathematics **11**, 193-210 (1999).

151.   W. Robinett, and J. P. Rolland, "A computational model for the stereoscopic optics of a head-mounted display," Presence: Teleoper. Virtual Environ. **1**, 45-62 (1992).

152.   J. M. Rodgers,"Nonstandard representations of aspheric surfaces in optical design," Ph.D. thesis, University of Arizona (1984)

153.   J. M. Rodgers, and K. P. Thompson, "Benefits of freeform mirror surfaces in optical design," in *American society for precision engineering* (2004).

154.   J. R. Rogers, "Techniques and tools for obtaining symmetrical performance from tilted-component systems," Optical Engineering **39**, 1776-1787 (2000).

155.   P. J. Rogers, "Biocular magnifiers - a review," in *International Lens Design Conference, SPIE Vol. 0554*
T. a. Moore, ed. (1985).

156.   P. J. Rogers, "The optics/vision interface," in *Critical review vol. CR41, lens design*(1992).

157.   J. Rolland, "Fundamentals of Optics for Head-Mounted Displays," in *Virtual Reality Annual International Symposium Tutorials* (1995).

158.   J. Rolland, and H. Hua, "Displays: head-mounted," in *Encyclopedia of optical engineering* (Marcel-dekker, 2005).

159.   J. P. Rolland, "Towards quantifying depth and size perception in virtual environments," Presence **4**, 24-49 (1995).

160.   J. P. Rolland, F. Biocca, F. Hamza-Lup, Y. Ha, and R. Martins, "Development of Head-Mounted Projection Displays for Distributed, Collaborative Augmented Reality Applications," Presence **14**, 528-549 (2005).

161.   J. P. Rolland, M. W. Krueger, and A. Goon, "Multifocal Planes Head-Mounted Displays," Appl. Opt. **39**, 3209-3215 (2000).

162.   J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci, "Method of Adjustments versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Head-Mounted Displays," Presence: Teleoperators & Virtual Environments **11**, 610-625 (2002).

163.   J. P. Rolland, A. Yoshida, L. D. Davis, and J. H. Reif, "High-Resolution Inset Head-Mounted Display," Appl. Opt. **37**, 4183-4193 (1998).

164.   M. Rosete-Aguilar, and J. L. Rayces, "Eye Rotation and Vignetting in Visual Instruments," Appl. Opt. **41**, 6593-6602 (2002).

165.   S. D. Russell, V. W. Michael, R. Bijan, M. Ken, S. Peter, A. S. Gregory, S. Venkataraman, M. Warren, P. Javier, and F. O. Tobibio, "Electrochromic variable transmission optical combiner," 5801, G. H. Darrel, W. F. Eric, C. M. David, E. B. Charles, and J. G. Henry, eds. (SPIE, 2005), pp. 268-277.

166.   B. Schmidt, Central-Zeitung f. Optik u. Mekanik **52** (1931).

167.   B. T. Schowengerdt, and E. J. Seibel, "True Three-Dimensional Displays that Allow Viewers to Dynamically Shift Accommodation, Bringing Objects Displayed at Different Viewing Distances Into and Out of Focus," CyberPsychology & Behavior **7**, 610-620 (2004).

168.   Schulz, "Aspheric surfaces," in *Progress in Optics XXV* (1988).

169.   K. Schwarzschild, Astr. Mitt. Konigl. Sternwarte Gottingen (1905).

170.   W. H. Scidmore, and M. D. Flanagan, "Wide angle eyepiece with large eye relief," U.S. Patent 3,384,434 (1968)

171.   H. C. Self,"Optical tolerances for alignment and image differences for binocular helmet-mounted displays," Thesis Type, A. A. M. R. Laboratory, University (1986).

172.   D. Shafer, "Design with two-axis aspheric surfaces," in *Computer-aided optical design, SPIE Vol. 147* (1978), p. 171.

173.   R. R. Shannon, *The art and science of optical design* (Cambridge University Press, Cambridge, U.K. ; New York, NY, USA, 1997).

174.   J. Sheedy, and N. Bergstrom, "Performance and comfort on near-eye computer displays," Optometry & Vision Science **79**, 306-312 (2002).

175.   F. P. Shvartsman, "Replication of diffractive optics," in *Diffractive and miniaturized optics, SPIE critical reviews of optical science and technology*, CR49, S. H. Lee, ed. (1993).

176.   W. P. Siegmund, S. E. Antos, R. M. Robinson, and M. L. Thomas, "Fiber optic development for use on the fiber optic helmet-mounted display," Optical Engineering **29**, 855-862 (1990).

177.   A. Sisodia, A. Riser, and J. R. Rogers, "Design of an advanced helmet mounted display (AHMD)," in *Cockpit and Future Displays for Defense and Security*, 5801, (SPIE, Orlando, FL, USA, 2005), pp. 304-315.

178.   W. J. Smith, *Modern Lens Design* (McGraw-Hill, 2005).

179.   G. H. Spencer, and M. V. R. K. Murty, "General Ray-Tracing Procedure," J. Opt. Soc. Am. **52**, 672-676 (1962).

180.   J. E. Stacy,"Design of reflective optical systems with general surfaces," Ph.D. thesis, University of Arizona (1983)

181.   T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R. Picard, and A. Pentland, "Augmented reality through wearable computing," Presence **6** (1997).

182.   A. State, K. P. Keller, and H. Fuchs, "Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display," in *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*(2005), pp. 28-31.

183.   T. Stone, and N. George, "Hybrid diffractive-refractive lenses and achromats," Appl. Opt. **27**, 2960-2971 (1988).

184.   G. Strang, *Computational Science and Engineering* (Wellesley-Cambridge Press, 2006).

185.   G. Strang, *Introduction to Linear Algebra* (Wellesley-Cambridge, 2003).

186.   E. H. Stupp, and M. S. Brennesholtz, *Projection displays* (Wiley, Chichester [England] ; New York, 1999).

187.   I. E. Sutherland, "A head-mounted three-dimensional display," in *Fall Joint Computer Conference, AFIPS Conference Proceedings, Vol. 33* (1968), pp. 757-764.

188.   G. Swanson,"Binary optics technology: theory and design of multi-level diffractive optical elements," Thesis Type, Massachusetts Institute of Technology, University (1989).

189.   W. C. Sweatt, "Describing holographic optical elements as lenses," J. Opt. Soc. Am. **67**, 803-808 (1977).

190.   E. W. Tatham, "Technical opinion: getting the best of both real and virtual worlds," Commun. ACM **42**, 96-98 (1999).

191.   B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski, "ARQuake: an outdoor/indoor augmented reality first person application," in *The Fourth International Symposium on Wearable Computers* (2000), pp. 139-146.

192.   C. Thompson, J. J. Ockerman, L. J. Najjar, and E. Rogers, "Factory automation support technology (FAST): a new paradigm of continuous learning and support using a wearable," in *First International Symposium on Wearable Computers, 1997* (1997), pp. 31-38.

193.   K. Thompson, "Description of the third-order optical aberrations of near-circular pupil optical systems without symmetry," J. Opt. Soc. Am. A **22**, 1389-1401 (2005).

194.   K. P. Thompson,"Aberration fields in tilted and decentered optical systems," Ph.D. Thesis, University of Arizona (1980)

195.   K. P. Thompson, "Practical methods for the optical design of systems without symmetry," in *Design and Engineering of Optical Systems*, 2774, (SPIE, Glasgow, United Kingdom, 1996), pp. 2-12.

196.   L. N. Trefethen, *Spectral Methods in MATLAB* (SIAM: Society for Industrial and Applied Mathematics, 2001).

197.   P. J. Twardowski, and P. Meyrueis, "Design of some achromatic imaging hybrid diffractive-refractive lenses," in *Holographic Optics III: Principles and Applications*, 1507, (SPIE, The Hague, Netherlands, 1991), pp. 55-65.

198.   T. Uchida, K. Sato, and S. Inokuchi, "An optical see-through MR display wih digital micro-mirrow device," Transactions of the virtual reality society of Japan **7** (2002).

199.   H. Urey, "Retinal scanning displays," in *Encyclopedia of optical engineering*, R. Driggers, ed. (Marcel dekker, 2003), pp. 2445-2457.

200.   H. Urey, "Spot Size, Depth-of-Focus, and Diffraction Ring Intensity Formulas for Truncated Gaussian Beams," Appl. Opt. **43**, 620-625 (2004).

201.   H. Urey, C. Kan, and W. O. Davis, "Vibration mode frequency formulae for micromechanical scanners," Journal of Micromechanics and Microengineering **15**, 1713-1721 (2005).

202.   L. Vaissie, and J. P. Rolland, "Head mounted display with eyetracking capability," U.S. Patent 6,433,760 (2002)

203.   E. M. Vaskas, "Note on the Wasserman-Wolf Method for Designing Aspheric Surfaces," J. Opt. Soc. Am. **47**, 669-670 (1957).

204.   M. Velger, *Helmet-mounted displays and sights* (Artech House, Boston, 1998).

205.   V. Vlahakis, J. Karigiannis, M. Tsotros, N. Ioannidis, and D. Stricker, "Personalized augmented reality touring of archaeological sites with wearable and mobile computers," in *Sixth International Symposium on Wearable Computers* (2002), pp. 15-22.

206.   J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural problems for stereoscopic depth perception in virtual environments," Vision Research **35**, 2731-2736 (1995).

207.   G. D. Wasserman, and E. Wolf, Proc. Phys. Soc. B **62** (1931).

208.   M. Watanabe, H. Takayama, N. Asai, R. Matsuda, and S. Yamada, "A retinal scanning display with a wavefront curvature modulator," Journal of the Society for Information Display **11**, 511-515 (2003).

209.   R. H. Webb, G. W. Hughes, and O. Pomerantzeff, "Flying spot TV ophthalmoscope," Appl. Opt. **19**, 2991-2997 (1980).

210.   I. Weingartner, "Real and achromatic imaging with two planar holographic optical elements," Opt. Comm. **58**, 385-388 (1986).

211.   W. H. Welch, M. R. Feldman, and R. D. TeKolste, "Diffractive optics for head-mounted displays," in *Stereoscopic Displays and Virtual Reality Systems II*, 2409, (SPIE, San Jose, CA, USA, 1995), pp. 209-210.

212.   W. T. Welford, *Aberrations of optical systems* (A. Hilger, Bristol, Eng. ; Philadelphia, 1991).

213.   H. Wendland, *Scattered Data Approximation* (Cambridge University Press, 2004).

214.   W. B. Wetherell, "Afocal lenses," in *Applied optics and optical engineering*, Wyant, and Shannon, eds. (1987).

215.   D. R. Williams, and N. J. Coletta, "Cone spacing and the visual resolution limit," J. Opt. Soc. Am. A **4**, 1514-1523 (1987).

216.   D. Williamson, "The eye in optical systems," in *Geometrical optics, SPIE Vol. 0531* (1985), pp. 136-147.

217.   G. C. d. Wit, "Contrast of displays on the retina," Journal of the Society for Information Display **13**, 178-179 (2005).

218.   G. C. d. Wit, "A retinal scanning display for virtual reality.," T. U. Delft (1997)

219.	G. C. d. Wit, "Retinal scanning displays: light sources moving over the retina," Science progress **82**, 135-149 (1999).

220.	A. P. Wood, "Design of infrared hybrid refractive-diffractive lenses," Appl. Opt. **31**, 2253-2258 (1992).

221.	S.-T. Wu, and D.-K. Yang, *Reflective liquid crystal displays* (Wiley, Chichester [Eng.] ; New York, 2001).

222.	A. D. Yalcinkaya, H. Urey, D. Brown, T. Montague, and R. Sprague, "Two-axis electromagnetic microscanner for high resolution displays," J. Microelectromechanical Systems **15**, 786-794 (2006).

223.	S. Yamazaki, K. Inoguchi, Y. Saito, H. Morishima, and N. Taniguchi, "Thin wide-field-of-view HMD with free-form-surface prism and applications," in *Stereoscopic Displays and Virtual Reality Systems VI*, 3639, (SPIE, San Jose, CA, USA, 1999), pp. 453-462.

224.	S. Yamazaki, K. Inoguchi, Y. Saito, H. Morishima, and N. Taniguchi, "A thin, wide field-of-view HMD with free-form-surface prism and applications," in *IS&T/SPIE Conference on the engineering reality of virtual reality, SPIE Vol 3639* (San Jose, California, 1999).

225.	W. Y. Yang, W. Cao, T.-S. Chung, and J. Morris, *Applied Numerical Methods using MATLAB* (Wiley-Interscience, 2005).

226.	Y.-Y. Yeh, and L. D. Silverstein, "Limits of Fusion and Depth Judgment in Stereoscopic Color Displays," Human Factors: The Journal of the Human Factors and Ergonomics Society **32**, 45-60 (1990).