Optical Design for a Head-Mounted Display

by

Jiantao Ma

B.S., Peking University, China (1987)

SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF
MASTER OF ENGINEERING IN BIOMEDICAL ENGINEERING

DEPARTMENT OF BIOMEDICAL ENGINEERING
MCGILL UNIVERSITY, MONTREAL

November, 1992

Optical Design for a Head-Mounted Display

by

Jiantao Ma

Submitted to the Faculty of Graduate Studies and Research
on November, 1992 in partial fulfillment of the
requirements for the Degree of
Master of Engineering in Biomedical Engineering

ABSTRACT

This thesis reports on the design of an optical relay for use in a color, stereo Head-Mounted Display (HMD) system.

Based on reviews and discussions of the requirements of the human visual system, major factors affecting the visual acuity, the aberration tolerances of the human eye, and optical design limitations, we derive the design criteria for the optical relay. A survey of alternate approaches to the three components of HMDs is presented.

A brief review of first order optics, aberration theory, general design principles, and computer aided lens design is also given.

Two multi-spherical lens systems, a straight structure and a folded layout, are presented. Their aberrations (distortion, coma, lateral color, field curvature and astigmatism) have been well corrected. Each of them has a 20 mm eye relief and an instantaneous field-of-view greater than 60°.

RÉSUMÉ

Cette thèse décrie conceptuellement un système optique pouvant être utilisé dans un "Head-Mounted Display" couleur et stéréo.

Les critères de conception d'un système optique sont déterminés à partir de revues et de discussions des exigences du système visuel humain, des facteures majeurs affectant l'acuité visuel, des tolérances de l'oeil humain aux abérrations et des limites du système optique. Un inventaire des approches alternatives au HMD à trois composantes est présenté.

Un court résumé d'optique du premier ordre, de la théorie des abérrations, des principes généraux de conception et de la conception de lentille assistée par ordinateur est aussi présenté.

Deux systèmes à lentilles sphériques multiples sont présentés. Le premier a une structure linéaire, le second, une structure pliée. Leurs abérrations (distortion, coma, couleur latérale, courbure de champ et astigmatisme) sont bien corrigées et analysées. Les deux systèmes sont placés à 20mm des yeux et possèdent des champs de vision instantanés de plus de 60°.

Acknowledgments

I would like to thank my thesis supervisor, Professor John M. Hollerbach, for all the supervision, encouragement and patience provided throughout the course of this thesis. Without his effort, this dissertation would have been impossible. It has been a rewarding experience working under his guidance.

Next I thank Professor Ian W. Hunter for his valuable advice at the beginning of the project and Dr. Yangming Xu for many discussions we had. Professor Ian W. Hunter initiated the idea of building the Head-Mounted Display.

My thanks also go to my fellow students in our lab for the help and support and for having made our lab a pleasant environment in which to work. I also thank Mr. Steve Kacani for his help in writing this thesis.

I am indebted to my wife Hongyan and my son Joshua. I thank my wife for her continued understanding, encouragement, and help.

Last, but not least, I would like to thank my parents, brother and sister for having been very supportive during the course of all my studies.

This research was supported by a Hydro-Quebec Fellowship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis reports on the progress in developing a color, stereo Head-Mounted Display (HMD) system, especially the design of an optical relay between displays and operator's eyes, at the Biorobotics Laboratory of McGill University.

A remote vision system is required for teleoperation in inaccessible environments. The particular tele-microbotics research at the Biorobotics Laboratory indicates a need for a vision system which magnifies images sent by a robot operating under a closed environmental shield. Similarly, a visual display is also necessary for microsurgical robots to perform human-guided retinal surgery and for endoscopic surgery. At a macrorobotic level, a visual display would be useful for teleoperation in space, underseas, in nuclear power plants, and for remote power line maintenance. A head-mounted display system was designed for use in these tasks.

A HMD is a device attached to an operator's head or helmet that produces a virtual image display for the operator. The device typically consists of three subsystems, a display or an image source, a relay optics, and a combiner. Often a HMD is used in conjunction with another device, called a head-mounted sight (HMS) or head-sight system (HSS), which is capable of determining the line of sight of the operator's head and controlling sensors' orientations in the remote system so that they point in the same direction as the operator's head. Thus the sensors are coupled to the orientation of the head, and the image produced by the sensors are displayed for the operator on the HMD, forming a closed loop system. Figure 1.1 [4] shows pictorially the HMD concepts.

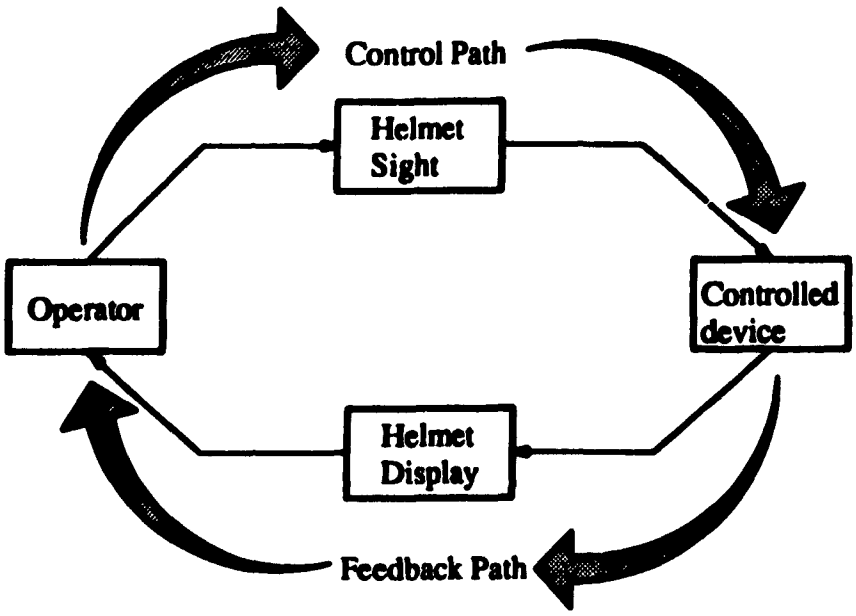


Figure 1.1: Helmet-Mounted Display system concepts. From [4].

1.1 Statement of the Problems

HMDs are often associated with the military, but recently HMDs have been used in teleorobotics [16]. Developing a HMD system for telemanipulation is motivated by the following facts.

- HMDs provide operators a quick access to important information from various sensors in the system, and the remote scene as well. This helps the operator guide the robot well.
- HMDs use extremely small displays to supply the human operators with panoramic views covering a very wide viewing angle of the remote scene, which is hard to achieve with conventional fixed (desk-top or panel-mounted) displays. This gives the human operator telepresence, the feeling of being physically present at the remote site during the performance of telemanipulation.
- A HMD in conjunction with a graphics processor is an essential component in achieving virtual reality, such as in a flight simulator.

Although considerable effort has been expended in developing commercial HMDs, the products fall into one of two categories: very good but very expensive, or inexpensive but not very good. We hope to develop an advanced HMD by a judicious mix between commercial components and our own developments.

The eventual goal of this HMD project is to design and construct a stereo, color vision HMD system (Biorobotics Laboratory HMD or BLHMD system) with a wide instantaneous field of view (IFOV), and with compensation for the operator's head and eye motion. The first key component that has to be designed is a wide IFOV optical relay.

The wide IFOV is vital for HMDs to achieve a realistic sensation of telepresence, and to enhance safety and efficiency. Because the optic relay determines the quality of images presented to operators, the aberrations of the optical relay should be well corrected. Moreover, because HMDs have been investigated as a possible means of man-machine interface, the design criteria of the optical relay comes mainly from human factors requirements. Hence careful consideration of human factors must be made in specifying the design criteria to ensure the operational utility of HMDs for a particular application. The main objectives of this thesis are:

1. to derive optical design criteria by considering human factors and optical constraints; and

2. to design an optical relay for use in the BLHMD system, which has a IFOV greater than 60°, well-corrected aberrations, while satisfying human factor constraints and optical constraints.

The following three sections review the typical components of both see-through and non-see-through HMDs, in order to give a clearer view of each component's functions, relative advantages and disadvantages of various approaches, and the choice of components for the BLHMD.

In Section 1.5 several existing HMD systems are introduced to give a full picture of applications and of alternate approaches towards HMDs, including our BLHMD. The design criteria of the optical relay and human factors criteria are discussed in Section 1.6. The contributions and the organization of the thesis are presented in Section 1.7 and Section 1.8 respectively.

1.2 Display or Image Source

Video displays convert a camera (or similar device) generated representation of a scene into a two dimensional image that can be viewed by human eyes. The roles of displays in see-through and non-see-through HMDs are different. A display of a non-see-through HMD presents an image which is an external world scene superimposed with another image providing some useful information from various sensors (called the information image). But a display in a see-through HMD only serves as an image source which shows the information

from various sensors and whose image is superimposed with the real world scene by a combiner. The ideal display for use in HMDs should have the following attributes:

- small display size but high resolution,
- small volume and light weight,
- full color, adequate luminance and high contrast,
- fast response,
- low operating voltage for safety consideration, low power consumption, and
- low cost.

1.2.1 Cathode-Ray Tubes

Of all devices that have been tried as HMD displays or images sources, the cathode-ray tube (CRT) is by far the most versatile and the most capable of producing a high quality image.

CRTs have high resolution (spot size can be smaller than 0.5 square mm), fast response (the actual speed of a CRT is limited by the electronics that drive it [22, page 99]), the widest viewing angle of any two dimensional electronic display devices, and very wide operating temperature range (from $-50^{o}C$ to $100^{o}C$). It can provide high brightness at the cost of reduced resolution. Usually CRTs are massive, bulky, have high power consumption, and require driving voltages as high as 20 kV. The resolution of color CRTs is lower than that of monochrome CRTs, since one color pixel consists of three primary color pixels (red, blue, and green).

Another important reason why most CRTs in see-through HMDs are monochrome is that combiners need monochrome displays whose primary wavelength falls into a narrow wavelength region in which combiners have a high reflection coefficient. By this wavelength match, combiners can have a high transmittance of an external world scene in all wavelengths except this very narrow region while still maintaining a high reflection of a monochrome information image.

1.2.2 Liquid Crystal Displays

Liquid crystal displays (LCDs) are light modifiers (modifying either external light or backlight), not light producers. The reflectance or transmittance of a liquid crystal changes when an electric field is applied.

As flat-panel displays, LCDs are superior to CRTs in

1. volume (LCDs are very thin),

2. weight,

3. power consumption (several tens of $\mu W/cm^2$), and

4. operation voltage (only 2 to 5 V).

LCD resolution is increasing at a very rapid rate, from 528 pixels per square cm in the early 1980's to 4039 pixels per square cm three years ago. These latter LCDs had resolutions four times greater than CRTs in television service today. Now, a commercial Thin-Film Transister (TFT) active matrix LCD (2.7 inch diagonal Sony Watchman) has the same pixel size, 86,400 pixels in an area of 22.57 square cm. In fact, resolution is no longer a problem for LCDs. The best LCDs can rival or surpass good-quality CRTs when the viewing angle is zero. The drawbacks are slow response and smaller viewing angles. The high cost of LCDs has previously limited the application of LCDs, but the price is rapidly decreasing.

The small viewing angle of LCDs is not a problem for use in HMDs, since the display is always positioned at the zero viewing angle in HMDs. The future holds the possibility of LCDs with higher resolution, and faster response.

1.2.3 Other Displays

Light-emitting diode displays (LEDs) are usually used as symbol-only image sources or as status indicators in HMDs due to their inherent low resolution and poor full color performance. Advantages of LEDs are fast response, low power consumption, low operation voltage, light weight and small volume.

Miniaturization of the display and its support equipment has made a vacuum fluorescent display (VFD) attractive for less demanding applications. Its low resolution and lack of full color capability limit its application in HMDs, in spite of its many advantages, such as high brightness, wide operation temperature, and low power consumption.

1.2.4 The BLHMD Display

The BLHMD design is based on commercial color LCDs. The length of CRTs would create difficulties in the arrangement of optical relays and CRTs, i.e., CRTs are usually put on the side of the opertor's head, not in front of the eyes. The high operation voltage of CRTs raises safety concerns. The weight of CRT is also a major problem. LCDs not only can produce almost the same quality image as CRTs within a small viewing angle associated with HMDs, but also have the beneficial characteristics of being thin and lightweight which

facilitate optical design. We believe that the slow response of current LCDs will soon be overcome, and LCDs have the potential of much higher resolution than CRTs. In fact, the current color LCDs already have smaller pixel sizes than color CRTs, but the total available pixel number is not enough. Meanwhile the cost of LCDs has already come down in the these two years. For example, a Sony 2.7 inch Watchman cost $599 at the end of 1990, and $278 six months later. Based on these considerations, LCDs will soon be ideal for HMD use.

1.3 Combiners

Both see-through and non-see-through HMDs require combiners to superimpose an external world scene or image with information images from sensors. The design of combiners for see-through HMDs is relative simple compared to that for non-see-through HMDs. The ideal combiner should:

- not only provide high quality external world image, but also a clear information color image properly overlapped on the external world image,
- be light weight, small and easy to install,
- combine the images in real time, and
- not limit the IFOV of HMDs.

The optical combiner for non-see-through HMDs usually consists of simple beamsplitters which inject the information image into the primary external display image. This is especially true in a simulator or location where size and weight are of little importance. Beamsplitter combiners are simple but relative heavy if used in binocular HMDs, and don't have much flexibility to control two overlapped images. Also this combiner requires a complicated and heavy optical relay, since two are needed, one for the external world image and one for the information image.

Another novel approach, which is only possible when the digital image processing can be realized in real time, is a "computer combiner". The advantages of this computer combiner approach are:

1. better control of image quality in various application situations, such as daylight viewing or night viewing. It is impossible for see-through HMDs to be used in both daylight and night viewing.
2. image size and centering adjustment are controlled;

3. image combining can be carried out "off head", thereby reducing the weight of the HMD;

4. more functions other than image combining can be performed, i.e., choosing different information from various sensors to display, changing working mode, and zooming in and out;

5. it provides a possible means of correcting optics aberrations introduced by optical relays.

The disadvantage of this approach is the requirement of powerful video processors or computers to implement the image processing task in real time, and non-see-through capability. Today, computer technologies are growing so fast that computers and video processors are becoming more and more powerful and affordable.

The design of a combiner with see-through capability is where nearly all the effort lies. The combiner reflection and transmittance coefficients, which are probably the most critical see-through HMD characteristic to be specified, fundamentally determine the ratio in luminance between the information image and the external world scene. Basically two kinds of technologies, dichromatic coating and diffractive optics or holography, have been applied to the design of see-through combiners to control the reflectivity or transmissivity of combiners. To improve the transmittance coefficient of the combiner while still maintaining a high reflection coefficient, both technologies "tune" the combiner to have a high reflection coefficient in only a narrow range of wavelengths encompassing the primary wavelength of the information image display. It is for this reason that see-through HMDs use only monochromatic CRTs.

Holographic combiners are superior to coating combiners because a greater number of variables are available for correction of optical aberrations. The major drawback of holographic combiners is that multiple solutions of the diffraction equation may cause secondary images. Both coating and holographic combiners are angularly sensitive.

In order to obtain a large IFOV while retaining an acceptable eye relief (the distance from the last surface of the optical system to the entrance pupil of the human eye), several attempts have previously been made:

- optical power was added to a flat reflective combiner to reduce the size of the objective lens for the same IFOV, and also extend eye relief [11]. The addition of power to the combiner introduces several optical aberrations (coma, astigmatism, and distortion) which have to be corrected by a complex lens system.

- double- or multi-element combiners [34], and

- holographic optics [34, 10].

Another point to consider is the particular HMD application: daylight viewing or night viewing, since these two application situations dictate the design of the combiner. It is not possible to optimize one optical combiner for both day and night use. This is a major drawback of all optical combiners.

Although optical combiners have several drawbacks as indicated above, they are the only choice for see-through HMDs. The BLHMD is intended for use in telerobotics, in which case see-through ability is of little importance, and an optical combiner is not necessary. Since computer combiners have many advantages over optical combiners, we employ a computer combiner in the BLHMD system. The design of relay optics now is simplified to the design of an eyepiece.

A combiner element may be thought of as a part of the optical relay which bring images from the display to the viewer's eyes. Hence, the design of both the see-through and non-see-through combiners must be considered with the design of optical relays.

1.4 Optical Relays

An optical relay is usually called a "collimating lens" for see-through HMDs or an "eyepiece" for non-see-through HMDs. These two kinds of relay lenses have almost the same function: to make the image from a display appear as a virtual image at infinity with respect to the viewer. But the arrangement of these optical relays is different. In see-through HMDs, the optical relay is put between the display and combiner such that the virtual image formed by relay lenses is superimposed by a combiner with an external world scene at infinity. In contrast, non-see-through HMDs usually have optical relays between the display and the viewer's eyes such that the virtual image of a superimposed display image is formed at infinity.

An ideal optics relay should

- provide as large an IFOV as possible,
- have little or well-corrected aberrations, or if necessary, deliberately introduce aberrations to negate aberrations introduced by other optics components such as the combiner,
- satisfy all human constraints,
- be light weight, and

- safe.

Most design problems of the optical relay arise from the requirement of a large IFOV, because a large IFOV is formed by very oblique light rays which cause serious aberrations. The IFOV problem contains constraints such as the display is usually placed close to the optical relay, and the exit pupil of the optical relay is fixed at the location of the viewer's eye. When a display is located near the optical relay, the objective points on the margin of the display are extra-off-axis points and will be poorly imaged. The aperture of an optical system controls the rays that pass through the system to form a image. The optical relays of HMDs with a fixed aperture location lose a very important means of controlling aberrations.

1.4.1 Eyepiece or Magnifier

The most basic approach to the HMD optical relay is a simple magnifier, or eyepiece. The structure of the eyepiece is simple if the required IFOV is small, and is very complex if the IFOV is intended to be large while satisfying all other design criteria. An eyepiece differs fundamentally from a photographic objective, which is also a positive lens, in that the entrance and exit pupils are outside the system. The diameters of lenses are determined far more by the angular field to be covered than by the relative aperture set by the display or by the objective lens. The larger the IFOV to be covered, the larger the diameter of lenses required.

The various approaches to this class of optical relays include:

- multi-element spherical lenses providing more design freedoms for aberration correction, and relatively large exit pupil and long eye relief [11, 30];
- aspherical lenses [11] or gradient index lenses for aberration correction;
- removal of some power from lenses to combiner to increase the IFOV, to simplify the lens system, and to make the HMD more compact [11];
- prisms as relay elements to fold the whole optical relay to save space, and to correct chromatic aberrations [24].

The eyepiece or magnifier approach is generally simple and relatively light, and permits reasonably efficient transmission of the image from the display to the eye. There are many existing designs to choose as a starting point.

However, this simple magnifier approaches has a relatively small eye relief, a small exit pupil, and a limitation on the maximum achievable apparent IFOV. Moreover, there is a limitation on the size of display for binocular vision.

1.4.2 Objective-Eyepiece Structure

More complex optical design approaches have been constructed [15]. These are characterized by intermediate image planes, which then are re-imaged by optical relays. Usually this kind of optical relay consists of two folding mirrors and two relay lenses (an objective and an eyepiece).

This objective-eyepiece structure approach has a number of benefits:

- the display can be mounted further from the eye, so the limitation on the size of display is eliminated,
- the center of gravity of HMD may be easily set,
- the eye relief distance can be made larger, and
- more design freedom is available such that aberrations can be well controlled.

Disadvantages include the addition of more weight as a consequence of an additional lens system (the objective), and no see-through capability because no combiner is used. Obviously, this optical relay is more complex, and is more difficult to optimize.

1.4.3 Flexible Fiber Optics Bundle

Another approach to optical relay design is the use of a flexible fiber optics bundle (FFOB) for transmitting images on a display to the operator [31]. This approach removes the display weight by mounting the display off the operator's head. However, the size and flexibility of the FFOB are adversely affected as the resolution (determined by number of fibers) of the bundle is increased. For this reason, the optimum use of this approach lies with applications requiring intermediate resolution levels (40,000 to 150,000 elements). The cost of FFOBs is also a disadvantage. Another drawback is that two lenses are needed to couple images in and out of the FFOB.

1.4.4 The BLHMD Optical Relay

Our first approach to the optical relay is an eyepiece consisting of multi-spherical lenses. This system has a wide instantaneous FOV, well-corrected aberrations, and meets all the design criteria. Secondly, a folded design, consisting of a field lens system (objective), an eye lens system (eyepiece), and two folding mirrors, is considered as an alternative to the first approach.

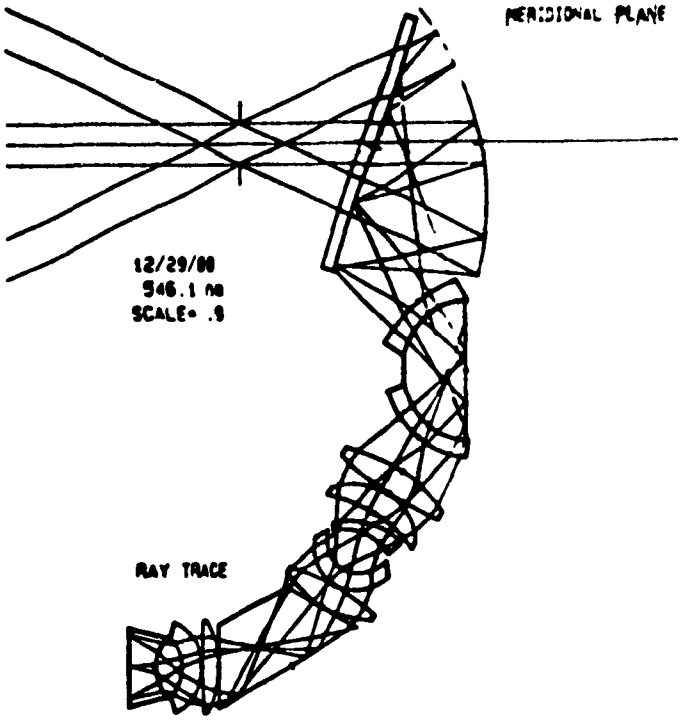


Figure 1.2: $50^o \times 60^o$ tilted cat display optics in the Integrated Helmet and Display Sighting System (IHADSS). From [10].

1.5 HMD Systems

In order to provide a full picture of HMD systems, several existing HMD systems are introduced in this section. The review of this section focuses on optical design, display technology, and other unique characteristics of these HMD systems.

1.5.1 Tilted Cat HMD

Figure 1.2 shows the Integrated Helmet and Display Sighting System (IHADSS) developed by Honeywell Systems and Research Center [10].

The display optics consistes of four major areas. Proceeding from the eye to the CRT, they are (1) combiner with objective collimating optics, (2) relay lens system with turn prism, (3) tilted field lens, and (4) second relay lens system with turn prism. The combiner consists of the beamsplitter with its angle-sensitive coating and a spherically curved, partially reflective combiner. Two mildly aspheric surface lenses are used in the second relay lens system. The system achieves satisfactory results but has very complicated display optics in which two mild aspheric lenses are used.

The combiner is a multielement combiner with 12^o tilt angle. The curved combiner coatings were designed to reflect only the CRT's P43 spectral output (from 510 nm to 670 nm wavelength). The display of this system is a monochrome CRT.

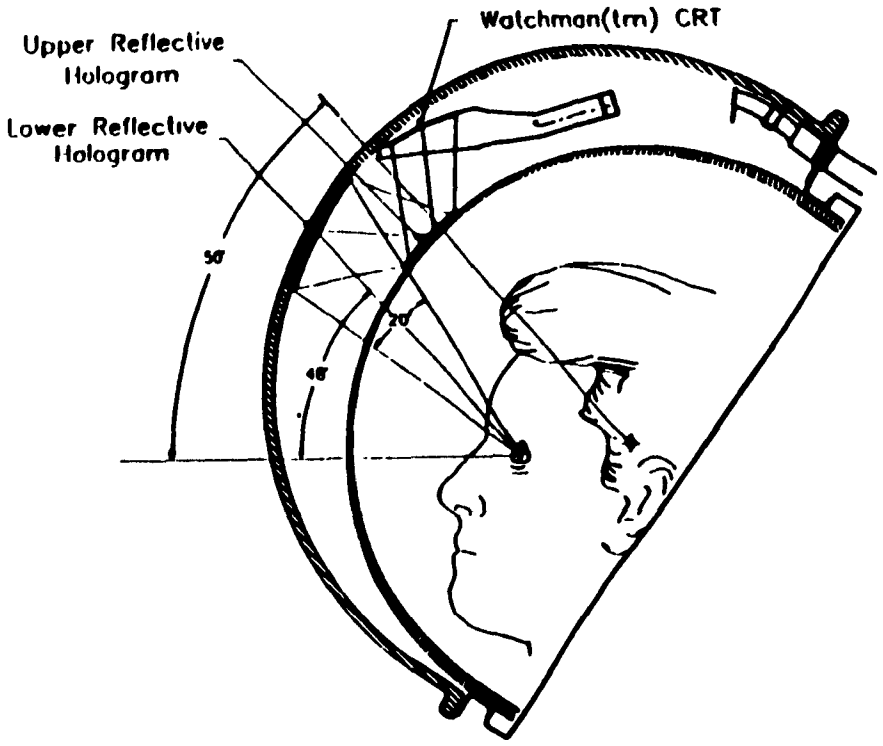


Figure 1.3: NASA Holographic HMD designed for the extra-vechicular mobility unit (EMU) on the Space Station Freedom, with head in viewing position. From [34].

The system has a 10 mm diameter exit pupil, a rectangular field of view of $50° \times 60°$, 30 mm on-axis eye relief, and 18 mm eye clearance when measured from the plane of the exit pupil to the lower edge of the beamsplitter. This gives an optimum eye position clearance of 13 to 25 mm.

1.5.2 NASA EMU Holographic HMD

Figure 1.3 shows a holographic HMD developed by the Technology Innovation Group (TIG)/Lockheed Engineering and Sciences Company (LESC) for the Crew and Thermol Systems Division and NASA-Johnson Space Center [34]. This HMD is unique in its use of holographic optical elements (HOEs) on the helmet and protective visor surfaces to relay an image from a CRT directly to the exit pupil. This HMD provides the user with a biocular virtual image in a 25 degrees diagonal FOV. Since it is "off axis", it usually has greater field aberration than symmetrical optical designs. A modified CRT from a 2.7 in. diagonal, 525 line Sony Watchman was used as an image source. The design was optimized to reduce the aberrations, particularly astigmatism and distortion, as much as possible, but a detailed aberration analysis was not presented [34].

Figure 1.4: Photo of the CAE helmet and fiber optics. From [31].

1.5.3 CAE Fiber Optic HMD

One approach to HMD design is to use a flexible fiber optics bundle (FFOB) for transmitting the source image to the operator. Figure 1.4 shows a fiber optic HMD (FOHMD) developed by CAE Electronics Ltd. for the U.S. Air Force Human Research Laboratory [31]. With a 38° stereoscopic overlap, the Phase V generation of the FOHMD has an IFOV of 160° horizontally and 80° vertically.

The advantages of this approach are that 1) the image source weight can be removed by mounting the source off the helmet; 2) the HMD optics can be folded more tightly along the helmet to improve the form factor of the display; and 3) the FFOB can be used to magnify, and therefore more effectively use the light from image sources. In low resolution applications the fiber structure is not objectionable, but in high resolution, wide angle HMDs, the problem is significant. It is hard to get large coherent fiber optic bundles without broken fibers, which blemish the image.

1.5.4 GEC NVG HMD

Some HMDs were designed to be used with night vision goggle (NVG) systems. Figure 1.5 showes the GEC Ferranti NITE-OP/NIGHTBIRD aviator's NVG developed by GEC Ferranti, UK [36]. Figure 1.6 illustrates an imaging telescope formed by objective and eyepiece lenses and a Image Intensifier (I^2) Tube. This system has a full 40 degrees circular FOV, 10 mm diameter exit pupil, and eye relief of 25 mm. Distortion across the full IFOV is 4 percent.

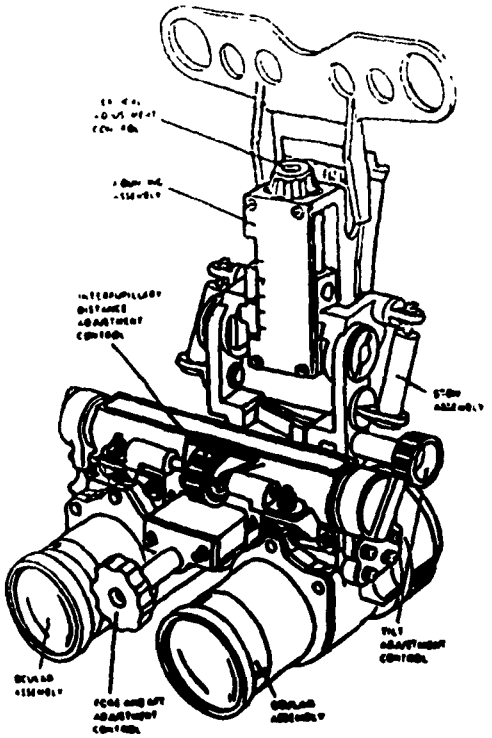


Figure 1.5: GEC Helmet-Mounted NVG system. The system includes a helemt interface and adjustment unit and a monocular unit. From [36].

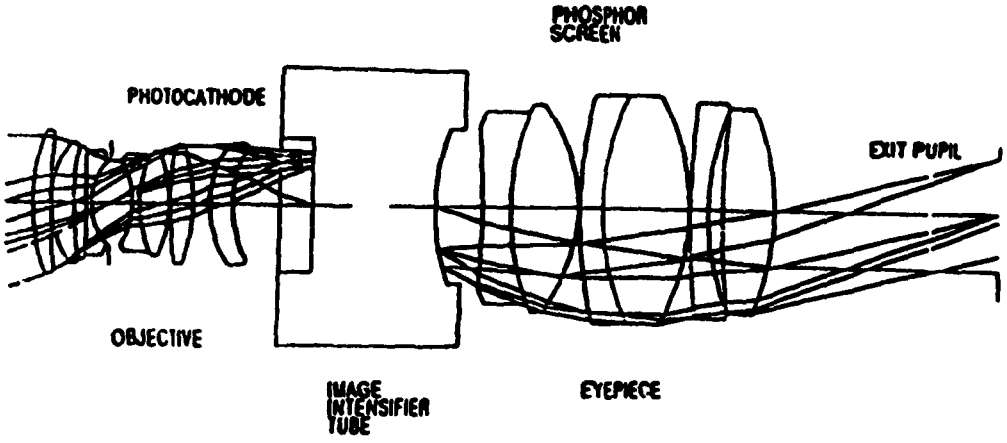


Figure 1.6: Optical Design of GEC's HelmtMounted NVG System. From [36]

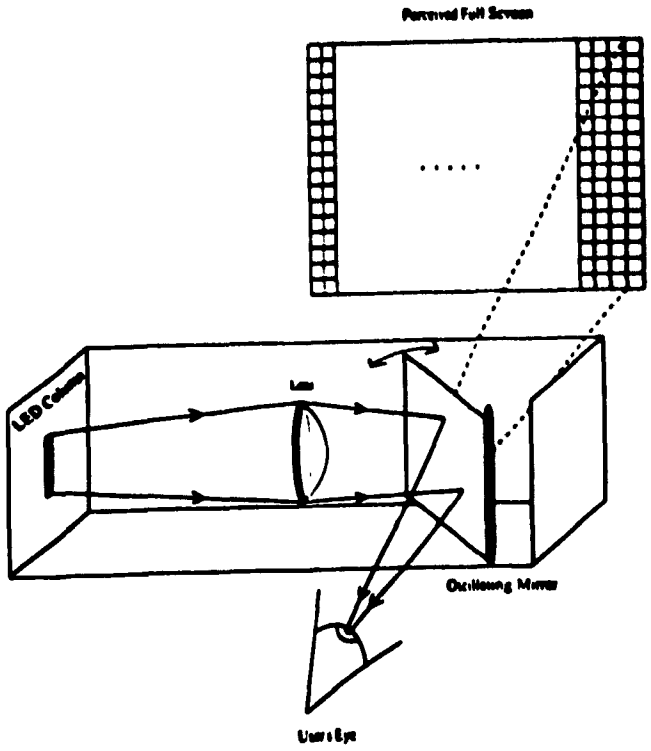


Figure 1.7: Schematic of the Private Eye's design. A linear array of LEDs is driven with one column of imaged data at a time. The horizontally oscillating mirror (50 Hz) scans the column across the observer's retina. The focusing lens serves to create a virtual image of the display about 2 ft in front of the observer and to correct for the user's spherical refractive error. From [21]

1.5.5 Private Eye HMD

A very interesting HMD design, called Private Eye (Figure 1.7), was built by Reflection Technology Inc., Waltham, Mass. [21]. The Private Eye display differs from most HMDs in its use of an LED array, a lens, and a scanning mirror to obtain a full alphanumeric and graphic display. Previously used LED HMDs were limited to a small number of symbols presented with a few elements.

This display creates a virtual image of a 12 in. monochrome monitor in a package of $1.1 \times 1.2 \times 3.2$ in., weighing about 2 oz. It is designed to be used as head mounted in front of one eye, with the other eye's view of the environment uninterrupted. The display provides $720(H) \times 280(V)$ pixels, and a $21^{o} \times 14^{o}$ field of view. The brightness is 2 fL nominal and is refreshed at 50 frames/s (non-interlaced).

1.5.6 Summary and General Description of BLHMD System

In their development history of more than two decades, HMDs were restricted to various military applications. Because of the requirements for image sources, optical designs of HMD pushed the related technologies forward; in return, HMDs have been improved significantly by the development of new technologies. There are many alternate design approaches to the components of HMDs, yet none is superior. Which approach is chosen depends upon the HMD's application and technologies available. Today there are many successful HMDs designed and tested for various applications.

The BLHMD system, which will have stereo color vision and very wide instantaneous FOV, is intended to be used in teleoperation. Figure 1.8 schematically shows the overall BLHMD system.

The remote scene is viewed by two charge-coupled device (CCD) color cameras. Two slightly different images from these two cameras are digitized by two frame grabber boards in an IBM RISC/6000 computer. The digitized real images are superimposed with two computer generated stereo images which presents useful information from various sensors. The two frame grabber boards send and display the superimposed stereo images on two color LCD screens which are mounted with two optical relays on a helmet. Images on the LCDs are imaged by the optical relays to form a pair of magnified virtual images at optical infinity. Stereo color vision is achieved by the disparity of these two virtual images viewed by both eyes.

In the meantime the camera mount is driven by signals from the head tracking sensor and eye position measurement system such that the cameras point to the same direction as the head and eyes.

The IBM RISC/6000 computer manipulates the images by two frame grabber boards, and also serves the head tracking sensor and eye position measurement system.

1.6 Human Factors and Design Criteria

Because HMDs are intended to be used as man-machine interface and the display, optical relay and head orientation sensor are usually mounted on the operator's head, human visual characteristics and mounting constraints dictate if a HMD is operational and safe. Moreover, since it is impossible to design and construct an ideal HMD, trade-offs must be made to enhance some visual parameters and compromise others. A good understanding of human factors is the only way to make good trade-offs.

There are four categories of human factors to be considered:

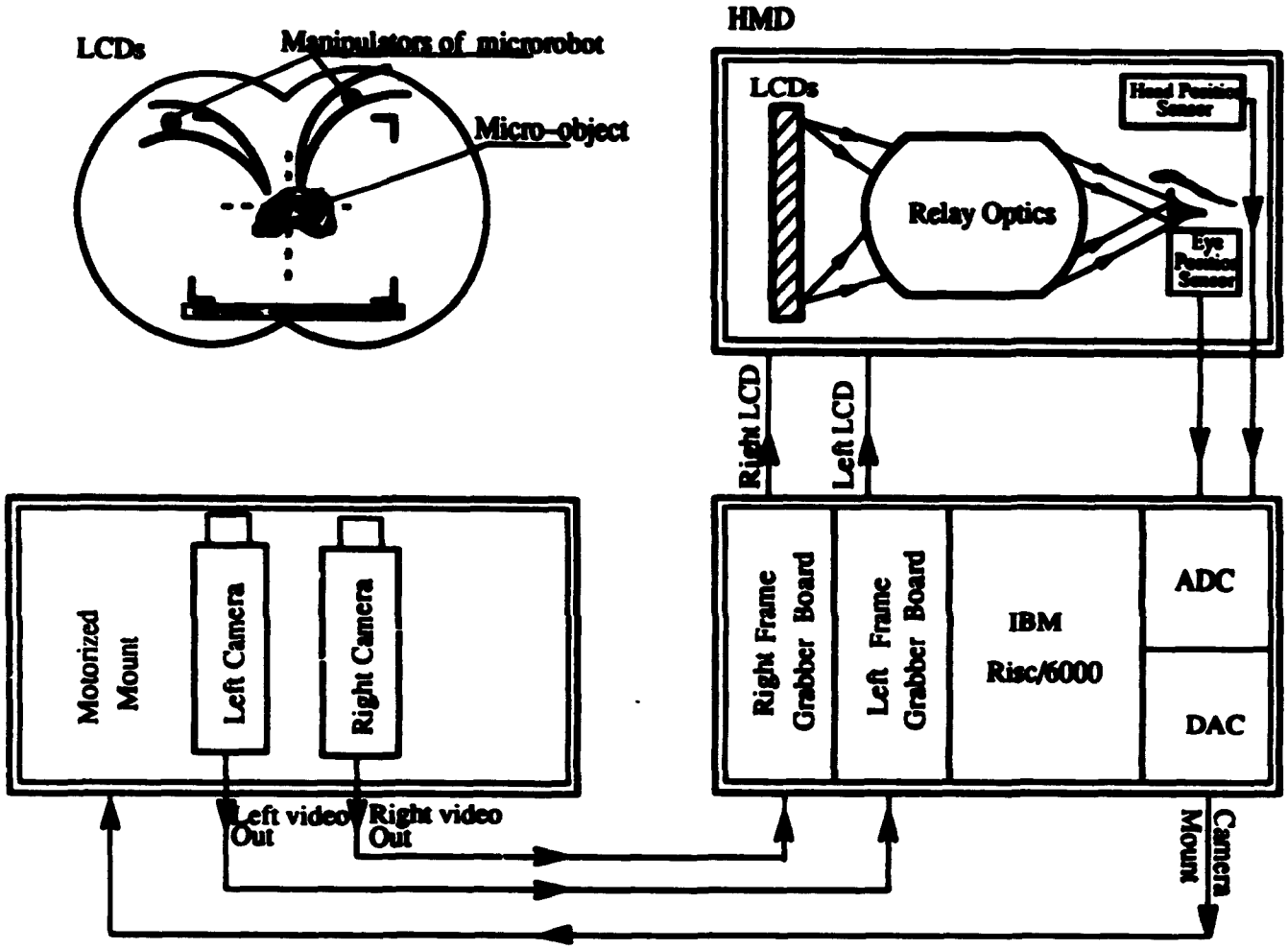


Fig 1.8 The BLHMD system.

1. factors affecting detection and recognition ability of human eyes, including luminance level, contrast, retinal location of targets, viewing distance, field-of-view, binocular, biocular or monocular vision, and color;

2. geometry factors, such as interpupillary distance, pupil size, other geometric parameters of the human head, and eye relief;

3. safety and endurance factors, including the maximum weight a human head can support during a long period of practical operation, and shifts in the center of gravity; and

4. tolerance of aberrations such as lateral color, distortion, field curvature, astigmatism, and coma.

Considerations of the first class of factors indicate that an advanced HMD should have color and binocular vision, an IFOV greater than 60^{o}, adequate luminance ($40-1000cd/m^2$), and high contrast, and that a virtual image should be located at a reasonable distance range set by aberration corrections.

The interpupillary distance, which is around 63 mm, and the space needed to install lenses set the limitation of the diameters of binocular lenses to no larger than 55 mm. The exit pupil of the optical relay, determined by the rotation of the eyeball and the pupil size which is in the range of 3 to 8 mm diameter, should have a diameter larger than 10 mm. The shape and size of a human head dictate the arrangement of the combiner and elements of the optical relay.

The important optical aberrations in the optical design of HMDs include distortion, lateral color, field curvature, astigmatism, and coma. The tolerance of aberrations depends upon applications and requirements. A wide IFOV optical relay usually has large aberrations, hence low acuity. The compromise is to ensure small aberrations and high acuity in the central field of view, and large aberrations on the margin of a wide IFOV.

The HMD design criteria, especially the optical design criteria, can be derived by considerations of human factors and optical constraints. Details of this issue are given in Chapter 3.

For cost and manufacturing reasons, only spherical lenses are used in the optical relay of the BLHMD system.

1.7 Contributions of the Thesis

The contributions of this thesis are the following:

1. Derive optical design criteria by considering human factors, which include 1) factors affecting visual acuity, 2) geometry factors, 3) aberration tolerances of the human eye, and 4) safety and endurance factors.

2. Design an optical relay for use in the BLHMD with the following characteristics:

   - IFOV: larger than 60 degrees,
   - Distortion on full field: less than 2.5 percent,
   - Lateral color: less than 3.5 minutes,
   - Field curvature: the tangential field curvature lies within 0.8 diopters of the central image plane, and the sagittal field curvature is less than 3 diopters within a $\pm 20^{o}$ field-of-view, and no part of the field is beyond infinity,
   - Coma: OSC less than 0.001,
   - Diameter of exit pupil: equal to 10 mm,
   - eye relief: equal to 20 mm,
   - Total length (of the straight optical relay): less than 100 mm,
   - All spherical lenses, and
   - Diameter of lenses: less than 54 mm.

1.8 Organization of the Thesis

This thesis mainly reports on the optical design for the BLHMD system.

Chapter 2 addresses optical design theories, including first order optics, aberration theory, general principles to correct aberrations, and computer aided lens design.

Chapter 3 shows the HMD system optical design constraints and targets by considering human factors and optical limits.

Chapter 4 discusses design problems of a wide IFOV optic relay. Basic considerations and design procedures of the optic relay are presented, as well as two resulting designs and corresponding analyses.

Chapter 5 contains the conclusions based on the work performed.

Chapter 2

Lens Design Theory

2.1 Introduction

In order to provide a background in lens design, this chapter briefly reviews first order optics, aberration theory, some general design principles, and computer aided lens design. For detailed discussions see [18, 29, 32].

All lens design procedures are based on the principles of geometrical optics, which assumes that light travels along rays that are straight in a homogeneous medium. Light rays are refracted or reflected at a lens or mirror, whence they proceed to form an image. An optical image system always has an object and an image. The space containing the entering rays at a surface under consideration is known as the *object space* for that surface; the space containing the rays emerging from a surface is called the *image space* for the surface. Because of the existence of virtual objects and virtual images, we must regard the object and image spaces as overlapping to infinity in both directions.

The basic and important characteristics of an optical system can be obtained by first order optics. Due to the inherent properties of refracting and reflecting surfaces and the dispersion of refracting media, the image of a point is seldom perfect but is generally afflicted with aberrations. To classify, identify, calculate, analyze, and correct aberrations are tasks of aberration theory. Because of the high non-linearity and complexity of most optical systems, aberration corrections require delicate designs and huge calculations. It is impossible in all cases to obtain analytic relations between aberrations and structural parameters of optical systems. Some empirical principles are very useful tools in the design of optical systems.

2.2 First order Optics

2.2.1 Ideal Optical Systems

First-order (or Gaussian) optics is often referred to as the optics of ideal optical systems. An ideal optical system is an abstract, structure-independent model which can be equivalent to any particular optical system. An ideal optical system should have these following properties [32]:

1. A point in the object space of this system corresponds to a unique point in the image space of this system. This pair of points is called the conjugate points.
2. A line in the object space of this system corresponds to a unique line in the image space of this system. This pair of lines is called the conjugate lines.
3. If a point is on a line in the object space of this system the conjugate point must be on the conjugate line in the image space of this system.

Based on the above definition, the theory of optical systems was derived by Gauss in 1841, and later named Gaussian optics.

2.2.2 Cardinal Points of an Optical System

A perfect, or well corrected, optical system can be treated as a "black box" whose characteristics are defined by its cardinal points, which are its first and second *focal points*, its first and second *principal points*, and its first and second *nodal points*.

Figure 2.1 illustrates the locations of the focal points and principal points of a generalized optical system. The object space is on the left of the system, the image space on the right. The *focal points* are those points at which light rays (from an infinitely distant axial object point) parallel to the optical axis are brought to a common focus on the axis. The optical axis is a line through the centers of curvature of the surfaces which make up the optical system. If the rays entering the system and those emerging from the system are extended until they intersect, the points of intersection will define a surface called the *principal plane*. In a well corrected optical system, the principal surfaces are spheres, centered on the object and image. In the paraxial region where the distances from the axis are infinitesimal, the surfaces can be treated as if they were planes. The intersection of this surface with the axis is the *principal point*. The "second" focal point and the "second" principal point are those defined by rays approaching the system from the left (object space). The "first" points are those defined by rays from the right (image space).

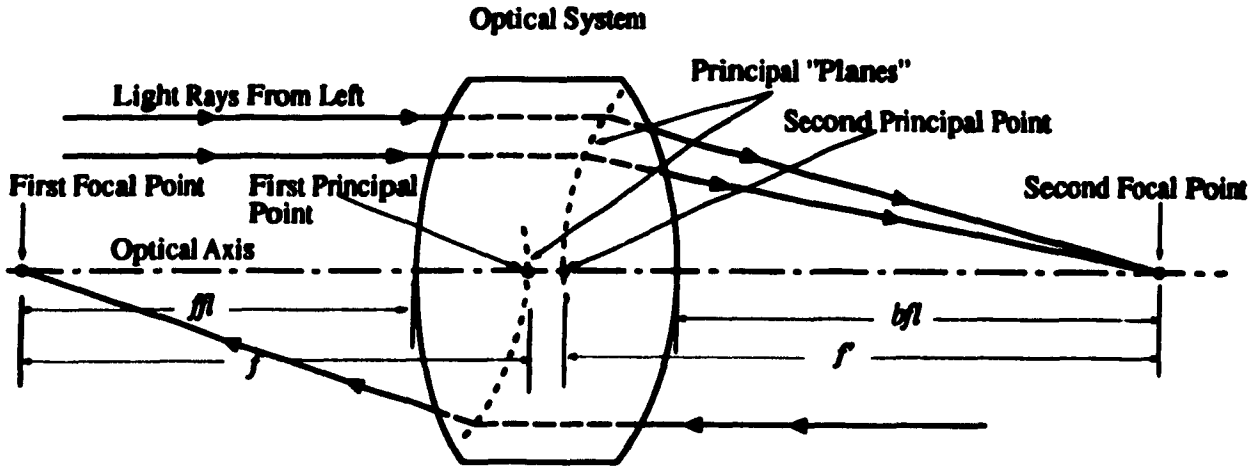


Figure 2.1: Illustrated the location of the focal points and principal points of a generalized optical system.

The object space *effective focal length* f of the system is the distance from the first principal point to the first focal point (Figure 2.1). The image space *effective focal length* f' of the system is the distance from the second principal point to the second focal point. If the refractive index of the object space medium differs from that of the image space, $f \neq f'$. The *back focal length* (bfl) is the distance from the vertex of the last surface of the system to the second focal point. The *front focal length* (ffl) is the distance from the vertex of the front surface to the first focal point.

Figure 2.2 shows the definition of nodal points. The *nodal points* are two axial points such that a ray directed toward the first nodal point appears (after passing through the system) to emerge from the second nodal point parallel to its original direction. When an optical system is bounded on both sides by air the nodal points coincide with the principal points.

2.2.3 A Simple Lens System

Figure 2.3 shows a simple lens system, which has radii r_1 and r_2, a center thickness t_c, and a refractive index n''. It is immersed in an object space medium with homogeneous refractive index n, while the image space medium has a homogeneous refractive index n'. The effective focal lengths f and f' can be found by equation 2.1:

$$f = \frac{n}{k} \text{ and } f' = \frac{n'}{k} \tag{2.1}$$

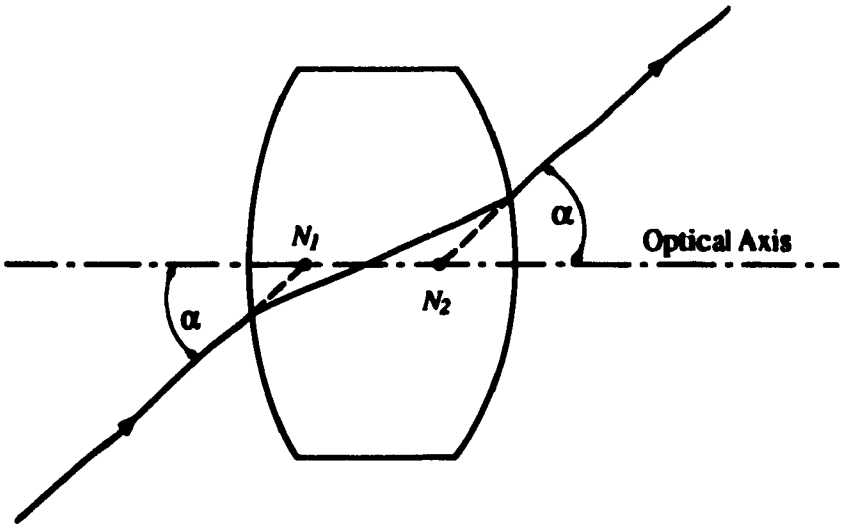


Figure 2.2: A ray directed toward the first nodal point N_1 of an optical system emerges from the system without angular deviation and appears to come from the second nodal point N_2.

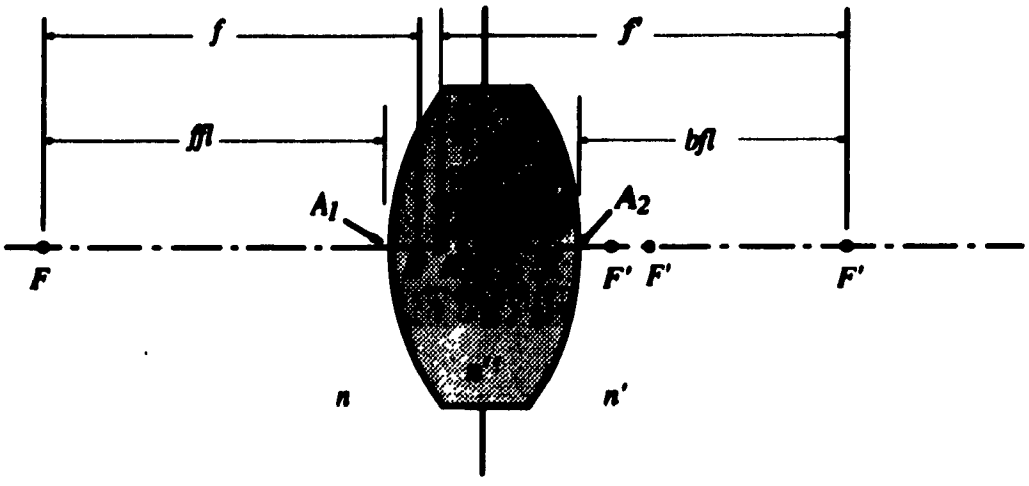


Figure 2.3: A simple lens with disparate object and image space indices.

where

$$k = \frac{n'' - n}{r_1} + \frac{n' - n''}{r_2} - \frac{t_c(n'' - n)(n' - n'')}{n'' r_1 r_2} \tag{2.2}$$

The locations of principal points H and H' with respect to the two vertices of the lens A_1 and A_2 are given by equations 2.3 and 2.4:

$$A_1H = \frac{nt_c}{k}\left(\frac{n' - n}{n'' r_2}\right) \tag{2.3}$$

$$A_2H' = -\frac{n't_c}{k}\left(\frac{n'' - n}{n'' r_1}\right) \tag{2.4}$$

Please note the following sign conventions.

- r_1 or r_2 of a surface is positive for the center of curvature to the right of the surface, and negative for the center of curvature to the left of the surface.
- A_1H is positive for the first principal point H to the right of vertex A_1 (interior to the lens), and negative to the left of vertex A_1 (exterior to the lens).
- A_2H' is positive for the secondary principal point H' to right of vertex A_2 (exterior to the lens), and negative to the left of vertex A_2 (interior to the lens).

Nodal point locations are:

$$A_1N = A_1H + HN \tag{2.5}$$

$$A_2N' = A_2H' + H'N' \tag{2.6}$$

where

$$HN = H'N' = (n' - n)/k, \tag{2.7}$$

which is positive for N to the right of H and N' to the right of H'.

2.2.4 Combination of Ideal Optical Systems

Many complex imaging tasks cannot be performed with a single lens. Any number of lenses can be dealt with by considering the coaxial combination of two lenses and then combining this "combination" with the next lens, and so on, until all the lenses in the system are accounted for.

Figure 2.4 shows the coaxial combination of two ideal optical systems. Given the focal lengths, f_1, f_1', f_2, f_2', the locations of cardinal points, H_1, H_1', H_2, H_2', F_1, F_1', F_2, F_2',

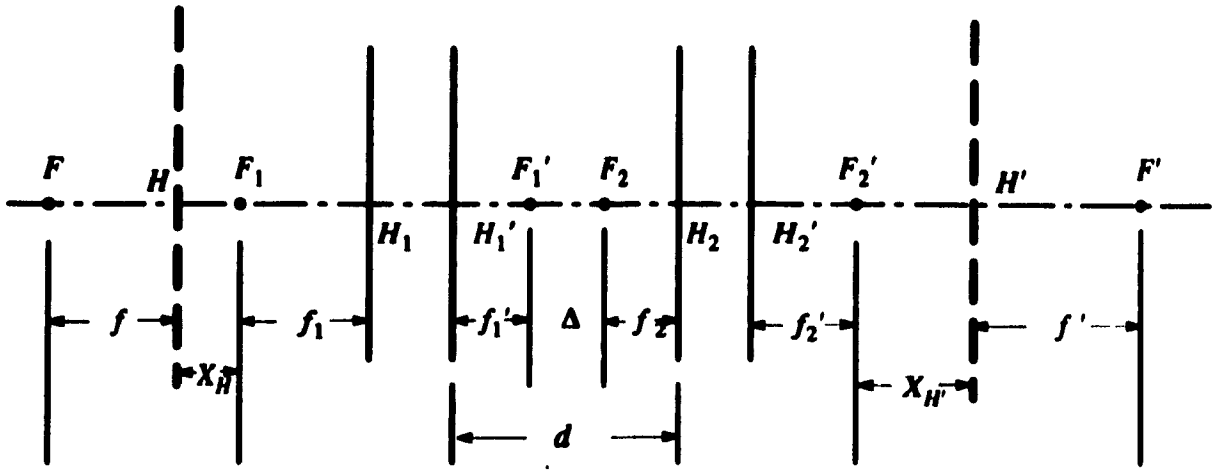


Figure 2.4: Coaxial combination of two ideal optical systems.

of these two systems, and the separation of the two system represented by d (the distance between H_1' and H_2) or Δ (the distance between F_1' and F_2), we can find the cardinal points and focal lengths of the combined system: H, H', F, F', f, f'. The focal lengths, f and f', are given by equations 2.8 and 2.9:

$$f = -\frac{f_1 f_2}{\Delta} \tag{2.8}$$

$$f' = -\frac{f_1' f_2'}{\Delta} \tag{2.9}$$

The locations of the principal points which are measured by X_H, the distance from H to F_1, and $X_{H'}$, the distance between H' and F_2', may be found by equations 2.10 and 2.11:

$$X_H = f_1 \frac{\Delta + f_1' + f_2}{\Delta} = f_1 \frac{d}{\Delta} \tag{2.10}$$

$$X_{H'} = f_2' \frac{\Delta + f_1' + f_2}{\Delta} = f_2' \frac{d}{\Delta} \tag{2.11}$$

Note the sign conventions. Assume that incident rays of light come from the right.

1. Δ is positive if F_1' is to the left of F_2, negative otherwise.

2. d is positive if H_1' is to the left of H_2, negative otherwise.

3. X_H is positive if H is to the left of F_1, negative otherwise.

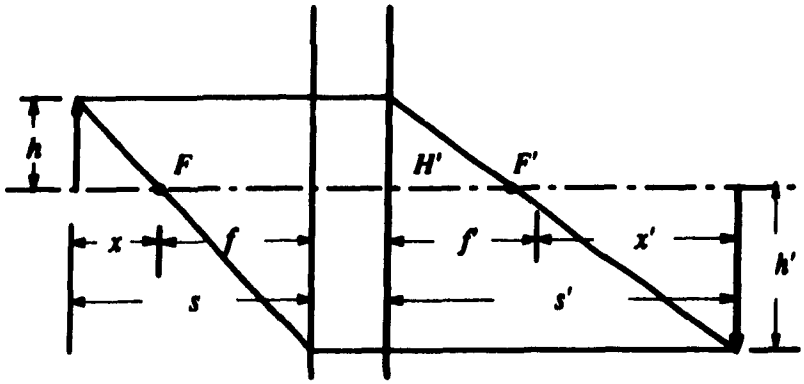


Figure 2.5: Image position and size of an ideal optical system.

4. $X_{H'}$ is positive if H' is to the right of F_2', negative otherwise.

2.2.5 Image Position and Size

When the cardinal points of an optical system are known, the location and size of the image formed by the optical system can be readily determined (see Figure 2.5). The image position x' (the distance from the second focal point to the image plane) or s' (the distance from the second principal point to the image plane) can be obtained by the Newtonian image equation 2.12 or the Gaussian image equation 2.13 separately. The Newtonian image equation is:

$$ff' = -xx' \qquad \Longrightarrow \qquad x' = -\frac{ff'}{x} \tag{2.12}$$

The Gaussian image equation is:

$$\frac{1}{s'} = \frac{1}{f} + \frac{1}{s} \qquad \Longrightarrow \qquad s' = \frac{sf}{(s+f)} \tag{2.13}$$

Note the sign conventions.

- Distances measured from the left of a reference point are negative, to the right, positive.
- The focal length of a converging lens is positive and the focal length of a diverging lens is negative.

There are two parameters to describe the size of an image of an optical system (Figure 2.5), the *lateral* (or *transverse*) *magnification* m, which is given by the ratio of image size to object size, and *longitudinal magnification* $\overline{m}$, which is the magnification of the

longitudinal thickness of the object. m and $\overline{m}$ can be found by equations 2.14 and 2.15.

$$m = \frac{h'}{h} = \frac{s'}{s} = -\frac{f}{f'}\frac{s'}{s} \tag{2.14}$$

$$\overline{m} = \frac{s_2' - s_1'}{s_2 - s_1} = \frac{s_1'}{s_1} \cdot \frac{s_2'}{s_2} \approx m^2 \qquad \text{(for small thickness)} \tag{2.15}$$

This indicates that longitudinal magnification is ordinarily positive. Note that heights above (below) the optical axis are positive (negative).

2.2.6 Limitations of First Order Optics

The first-order optics theory is only completely accurate for an infinitesimal threadlike region about the optical axis, known as the *paraxial region*. The value of first-order expressions lies in the fact that a well-corrected optical system will follow the first-order expressions almost exactly and also that the first-order image position and sizes provide a convenient reference from which to measure departures from perfection. In addition, the paraxial expressions are much easier to use than trigonometrical equations.

When the behavior of lenses with *finite* aperture and field of view is considered, the amount of aberrations of the lenses must be determined and hence the aberration theory of lenses is needed.

2.3 Aberration Theory

Aberrations of optical systems are measured by the amount by which rays miss the paraxial image point. This work of determining the aberrations cannot be done easily until the various types of image faults are classified and the behavior of each type is well understood. To evaluate each aberration, only a few rays need be traced. Thus the problem assumes more manageable proportions.

There are two classes of aberrations: monochromatic aberrations and chromatic aberrations. If the image of an optical system is formed only by monochromatic light, the optical system has five kinds of aberrations: spherical aberration, coma, astigmatism, Petzval curvature, and distortion. Of these aberrations, spherical aberration is only a function of aperture (becoming significant only when the aperture increases). Petzval curvature and astigmatism are only functions of field of view. The others are functions of both aperture and field of view.

But most optical systems form images in white light or multi-chromatic light. Since the

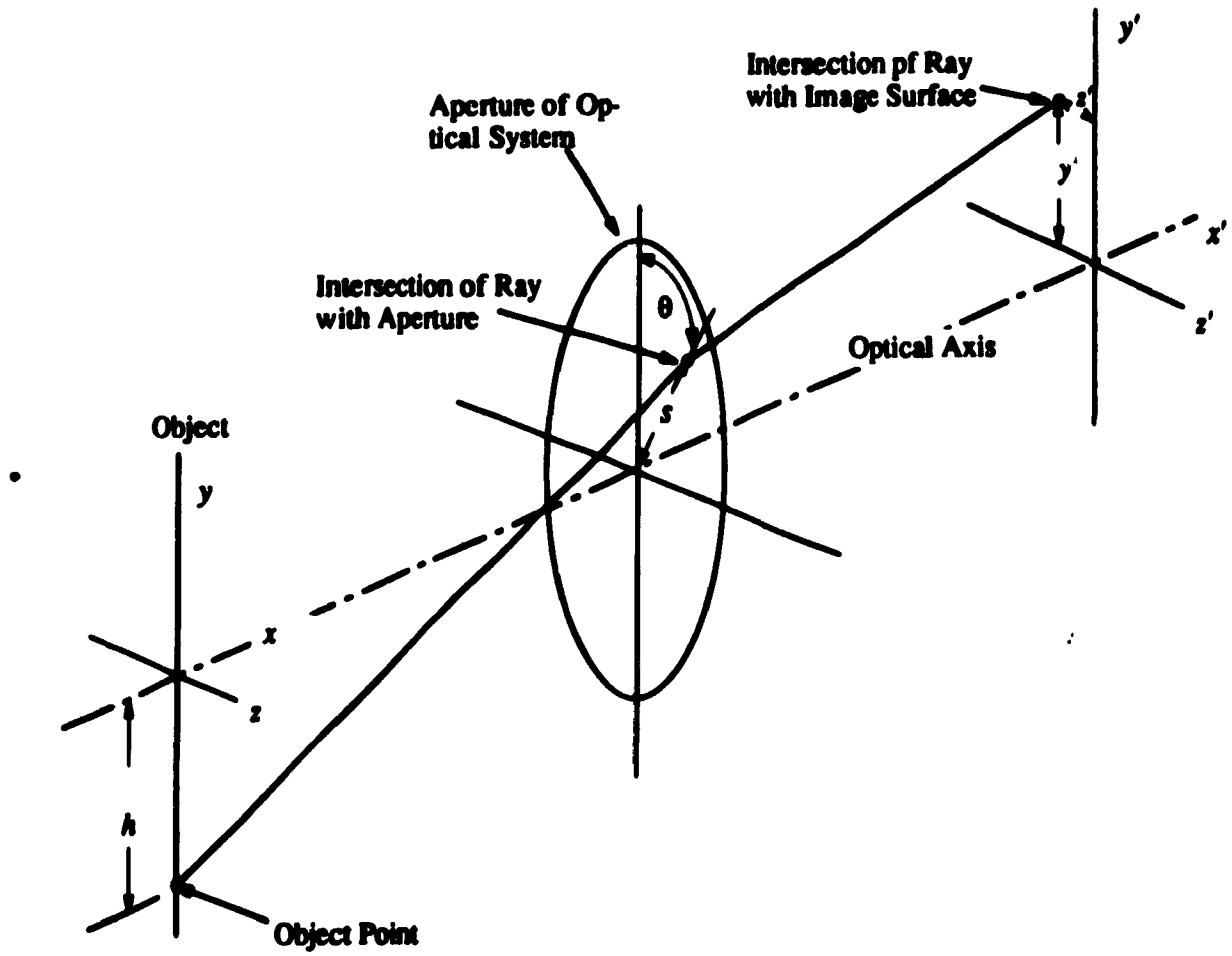


Figure 2.6: A ray from the point $y = h$, $(z = 0)$ in the object passes through the optical system at a point defined by its polar coordinates, (s, θ), and intersects the image surface at y', z'.

index of refraction varies as a function of the wavelength of light, chromatic aberrations are another concern.

2.3.1 Seidel Aberrations

Seidel investigated and codified the primary aberrations and derived analytical expressions for their determination. For this reason, the primary aberrations are usually referred to as the *Seidel aberrations.*

Considering an optical system with symmetry about the optical axis, and one coordinate system in the object space and one in the image space (Figure 2.6), we define a ray starting from an object point ($y = h$, $z = 0$) and passing through the optical system at a point defined by its polar coordinates, (s, θ). The ray intersects the image surface at (y', z').

Considering the axial symmetry of the system, the general form of the expressions of y' and z' can be derived as follows [29]:

$$\begin{aligned} y' &= A_1 s\cos\theta + A_2 h \\ &\quad + B_1 s^3\cos\theta + B_2 s^2 h(2+\cos 2\theta) + (3B_3 + B_4) s h^2 \cos\theta + B_5 h^3 \\ &\quad + C_1 s^5\cos\theta + (C_2 + C_3) s^4 h + (C_4 + C_6\cos^2\theta) s^3 h^2 \cos\theta \\ &\quad + (C_7 + C_8\cos 2\theta) s^2 h^3 + C_{10} s h^4 \cos\theta + C_{12} h^5 \\ &\quad + D_1 s^7 cos\theta + \cdots \end{aligned} \tag{2.16}$$

$$\begin{aligned} z' &= A_1 s\sin\theta \\ &\quad + B_1 s^3\sin\theta + B_2 s^2 h\sin 2\theta + (B_3 + B_4) s h^2 \sin\theta \\ &\quad + C_1 s^5\sin\theta + C_3 s^4 h\sin 22\theta + (C_5 + C_6\sin^2\theta) s^3 h^2 \sin\theta \\ &\quad + C_9 s^2 h^3\sin 2\theta + C_{11} s h^4 \sin\theta \\ &\quad + D_1 s^7\sin\theta + \cdots \end{aligned} \tag{2.17}$$

where A_i, B_i, etc., are constants and h, s, and θ have been defined above and in Figure 2.6.

In the above equations, the A_i terms are first-order terms relating to the paraxial imagery discussed in Section 2.2. All the other terms are called transverse aberrations. They represent the distance by which the ray misses the ideal image point as described by the paraxial imaging equations. The B_i terms are called the third-order, Seidel, or primary aberrations. B_1 is the spherical aberration, B_2 is the coma, B_3 is the astigmatism, B_4 is the Petzval curvature, and B_5 is the distortion. Similarly, the C_i terms are called the fifth-order or secondary aberrations. C_1 is the fifth-order spherical aberration; C_2 and C_3 are the linear coma; C_4, C_5 and C_6 are the oblique spherical aberration; C_7, C_8 and C_9 are the elliptical coma; C_{10} and C_{11} are the Petzval curvature and astigmatism; and C_{12} is the distortion. The 14 D_i terms are the seventh-order or tertiary aberrations.

In an axially symmetrical system there are no even-order aberrations; only odd-order terms may exist.

This section will present the definitions and representations of the Seidel aberrations (the B_i terms), and will discuss each aberration's characteristics and factors affecting it.

Spherical Aberration (the B_1 Term)

Spherical aberration can be defined as the variation of focus with aperture. Figure 2.7 is a sketch of a simple lens forming an *image* of an axial object point a great distance away. The spherical aberration can be measured in the vertical direction (called the transverse

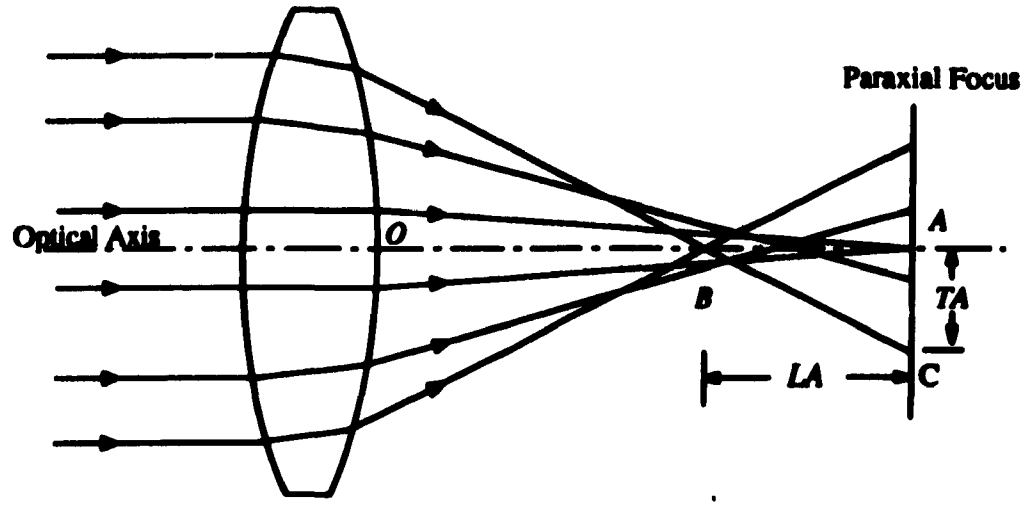


Figure 2.7: A simple converging lens with undercorrected spherical aberration. *TA* is the transverse spherical aberration; *LA* is the longitudinal spherical aberration. The rays farther from the axis are brought to a focus nearer the lens.

spherical aberration), or in the axial direction (called the longitudinal spherical aberration).

For a given aperture and focal length, the amount of spherical aberration in a simple lens is a function of object position and the shape, or bending, of the lens.

The image of a point formed by a lens with spherical aberration is usually a bright dot surrounded by a halo of light. The effect of spherical aberration on an extended images is to soften the contrast of the image and to blur its details.

Coma (the B_2 Term)

Coma is an off-axis aberration as shown in Figure 2.8. At non-zero field angles, coma appears as a difference in magnification for different parts of the lens surface. The appearance of a point image formed by a comatic lens is a comet-shaped flare as indicated in Figure 2.9. This causes blurring in the image plane of off-axis object points. Coma may be expressed by OSC - Offence against the Sine Condition.

Coma varies with the shape of the lens elements and also with the position of any aperture or diaphragms which limit the bundle of rays forming the image. In an axially symmetrical system there is no coma on the optical axis. The size of the coma patch varies directly with its distance from the axis.

In a particular design it may be controlled either by blending of the elements or by appropriate matching with associated objectives and transfer lenses.

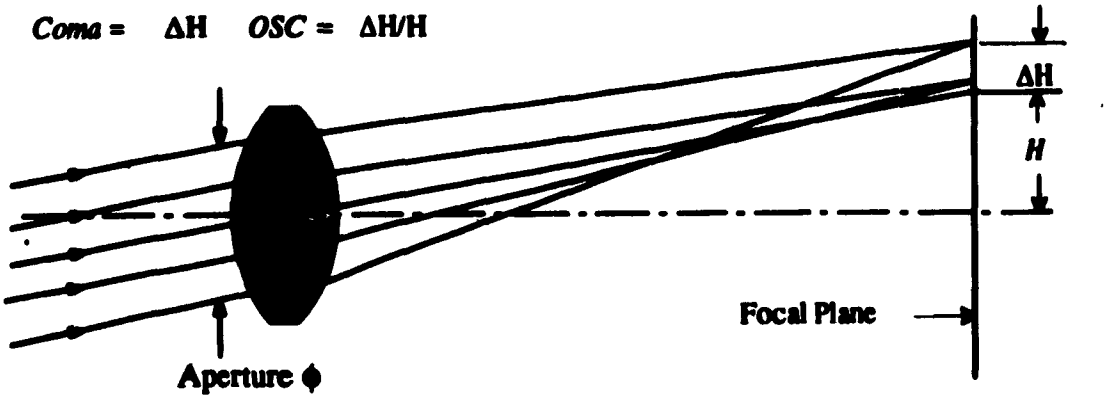


Figure 2.8: Coma appears as a difference in magnification for different parts of the lens surface. Coma can be described by OSC (Offense against the Sine Condition).

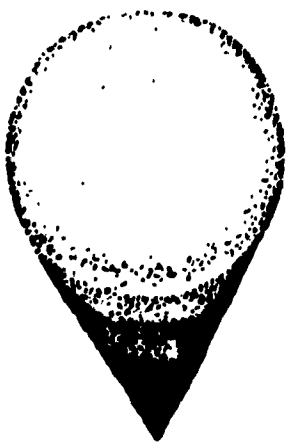


Figure 2.9: The coma patch. The image of a point source is spread out into a comet-shaped flare.

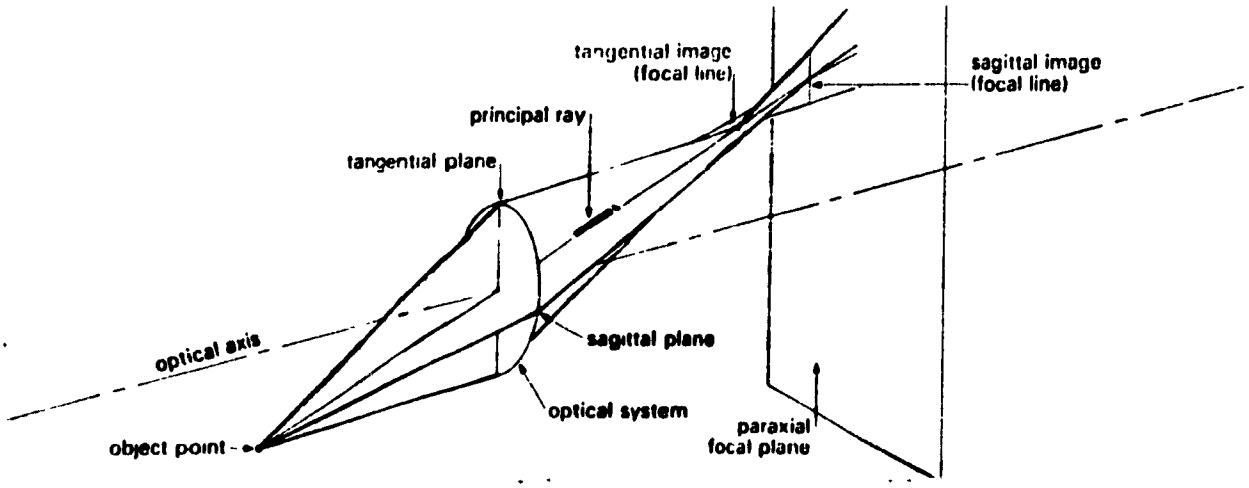


Figure 2.10: Astigmatism can be presented by these sectional views. From [13].

Astigmatism and Filed Curvature (the B_3 and B_4 Terms)

A schematic view of an optical system imaging an off-axis point is shown in Figure 2.10 [13]. A pair of focal lines, tangential and sagittal, can be found. In between these lines, the image is either an elliptical or circular blur. The separation of these two line is astigmatism. Figure 2.11 shows the definitions of field curvature and astigmatism.

Field curvature refers to the effect that optical systems image better on curved surfaces rather than on flat planes. In the presence of astigmatism, there are two separate astigmatic focal surfaces corresponding to the tangential and sagittal fields, and the tangential image surface lies three times as far from the Petzval surface as the sagittal surface. In the absence of astigmatism, tangential and sagittal focal planes fall into the same curved surface called the Petzval surface. Usually *field curvature* means the longitudinal departure of the tangential or sagittal image surfaces from the ideal flat image surface. We express the field curvature in diopters, which is the reciprocal of the longitudinal distance in meters from the tangential or sagittal image point to the location of the human eye - the exit pupil of the optical relay.

The amount of astigmatism in a lens is a function of the shape of the lens and its distance from the aperture which limits the size of the bundle of rays passing through the lens. The Petzval curvature is a function of the index of refraction of the lens elements and their surface curvature (this will be discussed later this chapter).

Owing to the high concentration of positive power of eyepieces, only moderate control exists over the Petzval sum, so that in the absence of astigmatism the field is highly curved

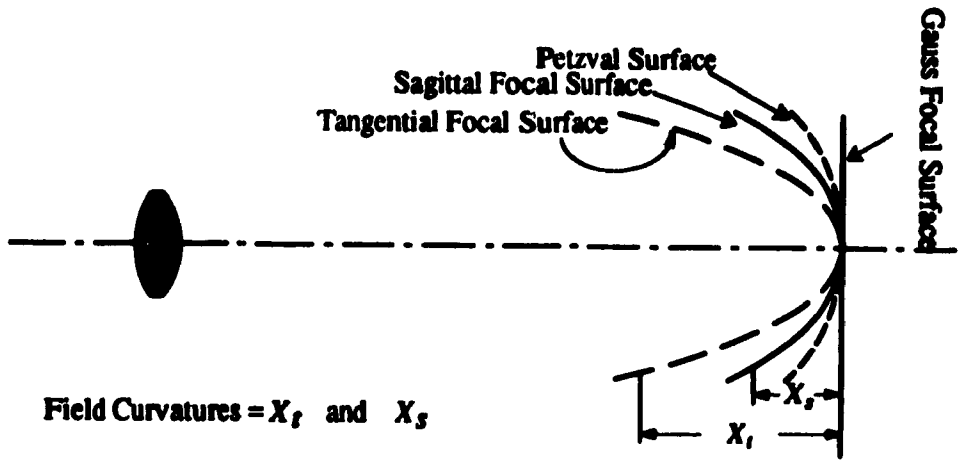


Figure 2.11: Field curvature is the tendency to image better on curved surfaces than the flat planes. The astigmatism is the separation between the tangential surface and the sagittal surface.

convex toward the eye. It may be neutralized by introduction of over corrected astigmatism, but too much of this will badly blur the outer image. The compromise is up to the designer, who must consider the end use of the instrument. If the tangential and sagittal field curves all lie within 1 diopter of the central focus, the image will be reasonably well defined over the field.

A flat tangential field combined with a 3 diopters curved sagittal field will just about correspond to the largest astigmatism that can be tolerated; in this case the outer field is useful only for identifying the presence of a possible target.

In the case of wide-angle eyepieces, the presence of higher-order astigmatism should be watched for, since violent changes of the tangential field can occur and make the design unusable.

Distortion (the B_5 Term)

Distortion is the separation of the actual image point from the paraxially predicted location on the image plane, and can be expressed either as an absolute value or a percentage of the paraxial image height (Figure 2.12). Distortion means that even if a perfect off-axis point is formed, its location on the image plane is not right. Hence distortion does not lower system resolution.

The amount of distortion ordinarily increases as the image size increases. There are two kinds of distortions: overcorrected, or pincushion, distortion and negative, or barrel, distortion.

Distortion is a function of aperture position and magnification. When a thin lens coin-

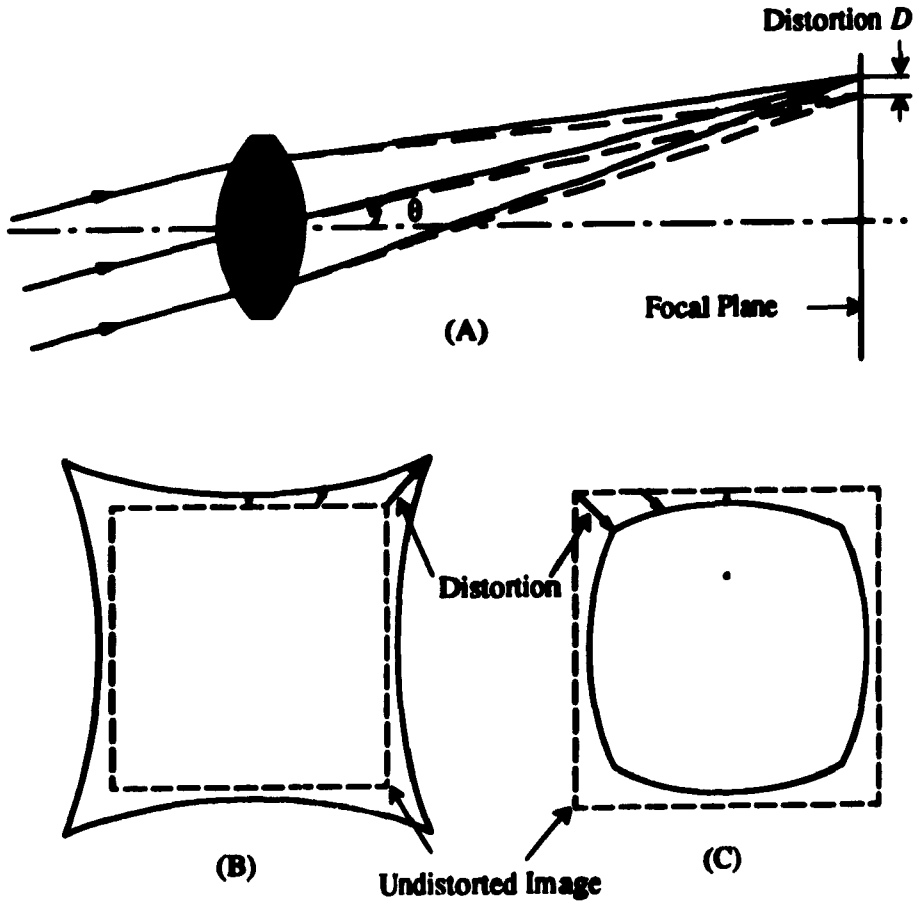


Figure 2.12: (A) Distortion is the separation of the actual image point from the predicted location on the image plane. (B) Positive, or pincusion, distortion. (C) Negative, or barrel, distortion. The sides of the image are curved because the amount of distortion varies as the cube of the distance from the axis.

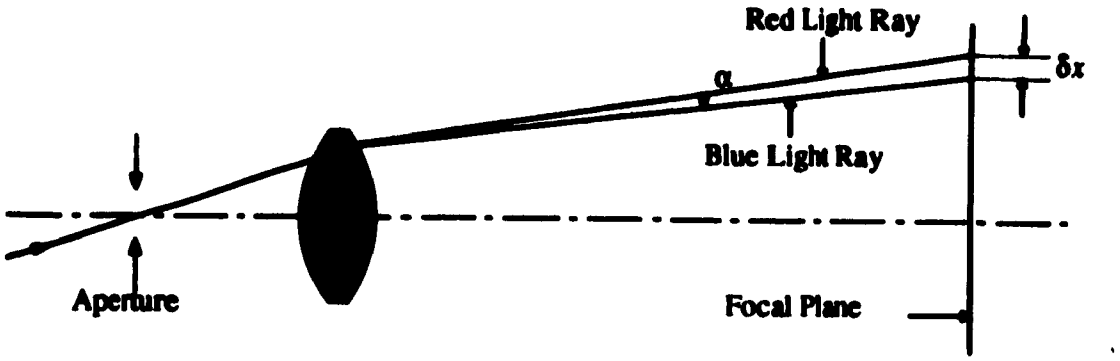


Figure 2.13: Lateral color, or chromatic difference of magnification, of a simple lens. The lateral color, which is defined as the difference δx in the image height, or the angle α between red and blue rays of light. results in different-sized images for different wavelengths.

cides with an aperture, there is no distortion. Also, perfectly symmetrical optical systems at 1:1 magnification have no distortion. It should be apparent that a lens or lens system has opposite types of distortion depending on whether it is used forwards or backwards.

2.3.2 Chromatic Aberrations

Chromatic aberrations are caused by the fact that the index of refraction varies as a function of the wavelength of light. In general, the index of refraction of optics materials is higher for short wavelengths than for long wavelengths; this causes the short wavelengths to be more strongly refracted at each surface of a lens, therefore they have different magnifications. When a lens system forms images of different sizes for different wavelengths, or spreads the image of an off-axis point into a rainbow, the difference between the image heights for different colors is called lateral color, or chromatic difference of magnification. Figure 2.13 shows the definition of lateral color.

Another chromatic aberration is the longitudinal chromatic aberration, which is the distance along axis between two focus points formed by blue light rays and red light rays.

The image of an axial point in the presence of chromatic aberration is a central bright color dot surrounded by a color halo; the dot color and halo color are dependent on the type of chromatic aberration (undercorrected or overcorrected) and the position of the screen on which the image is formed.

Chromatic aberrations vary with the optical materials, the shape of the lens elements, and also with the position of the aperture of the system. As in the case of spherical aberration, positive and negative elements have opposite signs of chromatic aberration.

Aberration	Aperture	Image Height
Lateral Spherical	y^3	-
Longitudinal Spherical	y^2	-
Coma	y^2	h
Astigmatism	-	h^2
Field Curvature	-	h^2
Distortion	-	h^3
Chromatic	-	-

Table 2.1: Variation of aberrations with aperture and image height.

Hence, chromatic aberrations can be compensated by grouping of positive elements of low dispersion and negative elements of high dispersion, or by using high dispersive prisms. For a simple lens lateral color is zero when the stop is in contact with the lens.

The chromatic variation of index also produces a variation of the monochromatic aberrations discussed in Section 2.3.1. In general the effects are of practical importance when the basic aberrations are well corrected.

2.3.3 Summary

By classifying the various types of image faults and by understanding the behavior of each type, the work of determining the aberrations of a lens system can be simplified greatly. There are five classes of monochromatic aberrations and two kinds of chromatic aberrations.

The above discussions of each aberration showed the strong dependency of aberrations on aperture and field size or angle. Table 2.1 summaries the relationships between the primary aberrations and the semi-aperture height y and image height h. Aberrations are also functions of shapes of lenses, optical media, location of the aperture, and structure of lens systems.

Equations 2.16 and 2.17 shows that the fifth or higher order aberrations are much more complicated than the third order aberrations. For many optical systems, the third order term is all that may be needed to quantify the aberrations. In highly corrected systems or those having large apertures or a large angular field-of-view, third order theory is inadequate. In these cases, exact trigonometric ray tracing is essential.

2.4 Correction of Aberrations

To correct aberrations of an optical system to a level required by a particular application is where almost all design effort lies. The correction of aberrations is a highly non-linear, multi-

variable optimization process through the adjustment of available structural parameters.

The design freedoms of an optical system usually include the radii of the surface, the thicknesses and air spaces, the position of the aperture, and the refractive index and dispersive powers of the glass from which the lenses are made. The best we can do to correct aberrations is to use our knowledge of aberrations and available design freedoms to improve the resulting image.

2.4.1 "Blending Technology"

The lens focal length equation 2.1 reveals that for a given index and thickness, there is an infinite number of combinations of r_1 and r_2 which will produce a given focal length. The aberrations of the lens are changed markedly as the shape or "blending" is changed. This effect is the basic tool of optical design.

For a single lens, applied under certain conditions, it is not difficult to find the curves of aberrations as functions of lens shape [29, 18]. For example, it was shown in [29] that with a stop at a lens and at $f/10$ covering a $\pm 17^o$ field, the lens has a zero value of coma and minimum spherical aberration when the lens shape is near convex-plano form. In a complex optical system, there are no such simple and general relations between lens shapes and aberrations of the system. In this case, experience plays an important role.

For reasons of manufacture and cost, most lenses have spherical shapes. But aspherical shape lenses have more advantages in controlling aberrations than spherical lenses. A Schmidt aspheric corrector plate, for example, is an aspherical lens used with eyepieces to flatten the field-of-view.

2.4.2 Petzval Sum and its Correction

Consider a surface radius of r in a lens system which has k surfaces. If the indexes of reflection of mediums before and after the surface are n and n' separately, the radii of curvature of the object and the image of this surface are represented by ρ and ρ', and all aberrations are somehow eliminated except for the field curvature, then for this surface, we have [18]

$$\frac{1}{n'\rho'} - \frac{1}{n\rho} = \frac{n' - n}{nn'r} \tag{2.18}$$

Write this expression for every surface in the lens (k surfaces) and add up, we find that

$$\frac{1}{n'_k\rho'_k} - \frac{1}{n_1\rho_1} = \sum_{i=1}^{k} \frac{n' - n}{nn'r} \tag{2.19}$$

It should be noted that a positive value of ρ corresponds to a negative sag, or an inward-curving surface. Hence the sag of the curved image of a plane object with $\rho_1 = \infty$, in the absence of astigmatism, is given by:

$$\begin{aligned} X'_{ptz} &= -\frac{1}{2}h'^2_k\frac{1}{\rho'_k} \\ &= -\frac{1}{2}h'^2_k n'_k \sum_{i=1}^{k}\frac{n'-n}{nn'r} \end{aligned} \tag{2.20}$$

This is the famous Petzval theorem, and the Petzval sum is defined as

$$\text{Petzval sum} \equiv \frac{1}{\rho} = n'_k \sum_{i=1}^{k}\frac{n'-n}{nn'r} \tag{2.21}$$

The Petzval sum provides an excellent indication of the likely amount of field curvature that will be found in the central part of the field of a lens after the astigmatism has been removed, or, alternatively, an indication of the residual astigmatism that may be expected after the meridional field locus has been flattened.

To secure a flat field free from astigmatism it is necessary to reduce the Petzval sum; this can be done either by widely separating the positive and negative components of the lens, by the use of a very thick meniscus form, or by the use of a high-index crown glass combined with a low-index flint glass.

2.4.3 Location of Aperture

Every optical system has apertures or stops. The function of apertures is to limit the bundle of rays forming images. It is obvious that, depending on its position in a lens system, an aperture selects some rays from an oblique pencil and rejects others. If the stop is moved along the axis, some of the former useful rays will be excluded while other previously rejected rays are now included in the image-forming beam. Consequently, unless the lens happens to be perfect, a longitudinal stop shift changes all the oblique aberrations in a lens. But similar to the case of lens shape, the relations between position of aperture and aberrations of a lens system depend strongly on the structure of the lens system. The equations of third-order aberrations of thin lenses as the function of aperture shift can be obtained if the thickness of elements in the system can be negligible [29].

If the aperture is located in the middle of a completely symmetrical optical system, i.e., the elements behind the aperture are mirror images of those ahead of the aperture and the system operates at unit magnification, the coma, distortion, and lateral color of the system are identically zero. This is the **symmetrical principle**, a principle of great utility, not

only for systems working at unit power, but even for systems working at infinite conjugates. This is due to the fact that, although coma, distortion, and lateral color are not completely eliminated under these conditions, they tend to be drastically reduced when the elements of any system are made symmetrical, or even approximately so. For this reason many lenses which cover appreciable field with low distortion and low coma tend to be generally symmetrical in construction.

2.4.4 Optical Materials

The most common lens material is optical glass, but crystals and plastic are frequently used.

The optical glasses are classified roughly as crowns and flints. Crowns have a V-value of 55 or more if the index is below 1.60, and 50 or more for an index above 1.60. The flint glasses are characterized by V-value less than these limits. A high refractive index leads to weaker surfaces and therefore smaller aberration residuals, but high-index glasses are usually expensive, more dispersive, and dense. Various glass types are often combined to control chromatic aberrations. An achromatic doublet, for example, consists of a low-relative-dispersion element of the same sign power as the doublet and a high-relative-dispersion element of the opposite sign. Several factors other than aberrations, such as cost, weight, and application conditions, have to be weighed while choosing glass types.

Plastic optical materials are rarely used for precision optical elements, because of their softness, high thermal expansion (eight times that of glass), high temperature coefficient of refractive index (120 times that of glass), and other disadvantages.

Gradient index media, whose refractive index varies from point to point within the media, are attractive because they may result in cost saving in some cases, perhaps, for example, where aspheric surfaces can be replaced by gradient elements. But it is not easy to make gradient index lenses to have a desired index profile.

2.5 Computer Aided Lens Design Optimization

There are many approaches to the computer aided lens design (CALD); almost all of them are characterized by the use of a "merit function". The merit function is a single numerical value which indicates to the computer whether any given change has improved the lens or not. Usually the merit function is a collection of aberrations and departures from desired conditions. The departures from desired conditions are counted as aberrations in the construction of the merit function in the following section. The most popular method in CALD is the *Levenberg-Marquardt algorithm*. The merit function of this algorithm is equal to the weighted sum of the square of all aberrations concerned plus the sum of weighted squares

of all the parameter changes (damping term). The adding of the damping term penalizes any large parameter changes and tends to stabilize the process.

Assume an optical system has k configuration parameters, $\mathbf{x} = (x_1, , x_2, \ldots, x_k)^T$. Its starting point is $\mathbf{x^0} = (x_1, , x_2, \ldots, x_k)^T$. For a given location of an objective and a given aperture or a FOV, the aberrations (including the departures from desired conditions) of the system are totally determined, and may be written as functions of the k variables as follows (assume a total m kinds of aberrations, f_1', , f_2', $\cdots$, f_m', are concerned):

$$f_i' = f_i'(x_1, , x_2, \ldots, x_k) \qquad i = 1, 2, \ldots, m \tag{2.22}$$

The merit function is

$$\begin{aligned} \Psi &= \sum_{i=1}^{m} \mu_i^2 f_i'^2 + \lambda \sum_{j=1}^{k} (\Delta x_j)^2 \\ &= \sum_{i=1}^{m} f_i^2 + \lambda \sum_{j=1}^{k} (\Delta x_j)^2 \end{aligned} \tag{2.23}$$

where μ_i, $i = 1, 2, \ldots, m$, are weight factors, λ is the damping factor, and $f_i = \mu_i f_i'$ is the ith weighted aberration. Aberrations are non-linear functions of the configuration variables, and can be linearized by expanding them in Taylor series at the starting point $\mathbf{x^0}$, and only taking the linear terms:

$$\mathbf{f} = \mathbf{f_0} + \mathbf{A}\Delta\mathbf{x} \tag{2.24}$$

where $\mathbf{f}$ and $\mathbf{f_0}$ are the aberrations in matrix form after and before the configuration parameters change, $\Delta\mathbf{x}$ is the changes of the configuration parameters, and $\mathbf{A}$ is a $m \times k$ matrix. They are expressed as:

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}, \qquad \mathbf{f_0} = \begin{bmatrix} f_1^0 \\ f_2^0 \\ \vdots \\ f_m^0 \end{bmatrix}, \qquad \Delta\mathbf{x} = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_k \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_k} \end{bmatrix}$$

The element $\frac{\partial f_i}{\partial x_j}$ in the matrix **A** can be approximated as follows:

$$\frac{\partial f_i}{\partial x_j} = \frac{f_i(x_1^0, x_2^0, \ldots, x_{j-1}^0, x_j^0 + \Delta x_j, \ldots, x_k^0) - f_i(x_1^0, x_2^0, \ldots, x_k^0)}{\Delta x_j} \tag{2.25}$$

The conditions that Ψ is a minimum are:

$$\frac{\partial \Psi}{\partial x_1} = 0, \frac{\partial \Psi}{\partial x_2} = 0, \cdots, \frac{\partial \Psi}{\partial x_k} = 0,$$

or the following simultaneous equations:

$$\left.\begin{aligned}
\tfrac{1}{2}\tfrac{\partial \Psi}{\partial x_1} &= f_1\tfrac{\partial f_1}{\partial x_1} + f_2\tfrac{\partial f_2}{\partial x_1} + \cdots + f_m\tfrac{\partial f_m}{\partial x_1} + \lambda\Delta x_1 = 0 \\
\tfrac{1}{2}\tfrac{\partial \Psi}{\partial x_2} &= f_1\tfrac{\partial f_1}{\partial x_2} + f_2\tfrac{\partial f_2}{\partial x_2} + \cdots + f_m\tfrac{\partial f_m}{\partial x_2} + \lambda\Delta x_2 = 0 \\
&\vdots \\
\tfrac{1}{2}\tfrac{\partial \Psi}{\partial x_k} &= f_1\tfrac{\partial f_1}{\partial x_k} + f_2\tfrac{\partial f_2}{\partial x_k} + \cdots + f_m\tfrac{\partial f_m}{\partial x_k} + \lambda\Delta x_k = 0
\end{aligned}\right\} \tag{2.26}$$

In matrix form, equation 2.26 is

$$\mathbf{A^T f} + \lambda\Delta\mathbf{x} = 0 \tag{2.27}$$

From equation 2.27 and 2.24, we get

$$(\mathbf{A}^T\mathbf{A} + \lambda I)\Delta\mathbf{x} = -\mathbf{A}^T\mathbf{f}_0 \tag{2.28}$$

where **I** is a $k \times k$ unit matrix. The solution to equation 2.28 is

$$\Delta\mathbf{x} = -(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{f}_0 \tag{2.29}$$

The new configuration parameters are

$$\mathbf{x} = \mathbf{x}^0 + \Delta\mathbf{x} \tag{2.30}$$

Choose this **x** as another starting point, repeate the same solution process until Ψ < a desired value. During the optimization process, the configuration variables must be inside certain ranges (boundary conditions).

This algorithm will seek out the nearest local minimum of the merit function, and the selection of the starting point for the process is vitally important. The knowledge of successful design types and features can direct the computer to good starting points. The other important issues are the selections of the weight factors, the damping factor, and the

configuration variables. The general design principles and experiences can help for these selections.

2.6 Conclusions

Optical design is always guided by the design theories. The important properties of almost all optical systems can be described by first-order optics. But rays of light do not follow first-order optics strictly to pass through a curved refractive surface in an optical system; this causes various aberrations.

Aberrations typically are classified as spherical aberration, coma, astigmatism, field curvature, distortion, and chromatic aberrations. But in any given lens all aberrations appear mixed together and correcting one aberration will improving the resulting image only to the extent of the elimination of that particular aberration from the over-all mixture. Some aberrations can be easily varied by merely changing the shape of one or more of the lens elements; others require a drastic alteration of the entire lens structure.

The design freedoms of an optical system usually include the radii of the surfaces, the thicknesses of lenses, the air spaces, the position of the aperture, and the refractive index and dispersive powers of the glasses of which the lenses are made. Another valuable means for aberration control is symmetry. In almost all cases the designer is restricted to the use of spherical refracting or reflecting surface. The attempts to use aspheric surfaces lead to extremely difficult manufacturing problems, and consequently such surfaces are used only when no other solutions to the problem can be found. Gradient index lenses can simplify the design and can replace aspherical lenses, but it is still difficult to make gradient lenses with designed index profiles.

Computers provide a powerful means for the lens design. The most popular algorithm is the Levenberg-Marquardt algorithm whose merit function is equal to the weighted sum of the squares of all aberrations concerned plus the sum of the weighted squares of all the parameter changes. This method is capable of driving a rough preliminary design form to the nearest local minimum of the merit function. To find another minimum, the designer could 1) select another starting point, 2) change the weight factors, or 3) change the damping factor. The design experiences and knowledge of general design principles are very important in the course of CALD.

Chapter 3

Optical Design Criteria and Human Factors

Investigated as a possible means of man-machine interface, HMDs should meet human requirements as much as possible. The understanding of human visual characteristics will help to set design criteria of the optical relays of HMDs.

This chapter focuses on the derivation of the optical design criteria by considering human visual characteristics. Section 3.1 discusses the human eye structure, factors affecting visual acuity, comparison of stereo and monocular vision, effect of field of view, and color vision. Section 3.2 addresses the geometrical constraints on optical relays of HMDs. Section 3.3 discusses human visual tolerances. The last section summarizes the design criteria obtained from the first three sections.

3.1 Human Eye Characteristics

3.1.1 General

Figure 3.1 shows the cross section of the right human eye [5]. The anatomical structure of the eye is roughly analogous to the optical imaging apparatus of a camera. The eye focuses rays of light from objects in the visual field so that a reasonably accurate, integrated image forms at the back of the eye, on the retina. Figure 3.2 shows in simplified form the eye forming an image of a distant point object [13]. The cornea and the crystalline lens together provide the refractive or focusing power of about 60 diopters (16.7 mm focal length) when focused on distant objects and 69.4 diopters (14.4 mm focal length) when focused on extremely close objects.

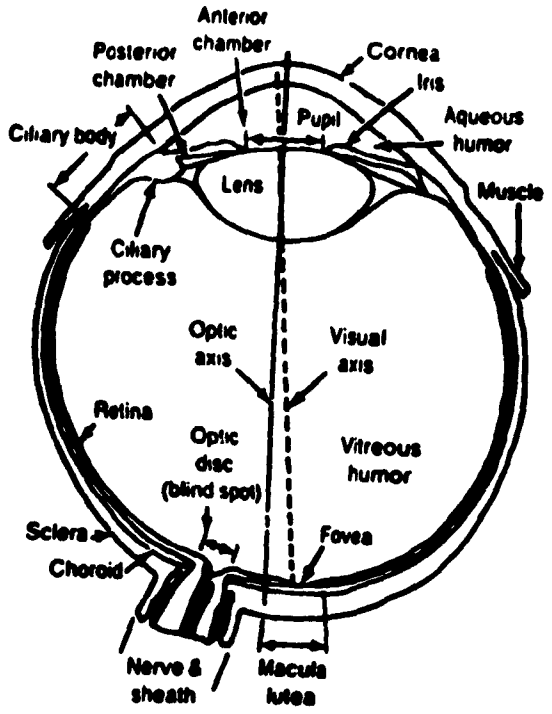


Figure 3.1: Horizontal section of the right human eye. From[5].

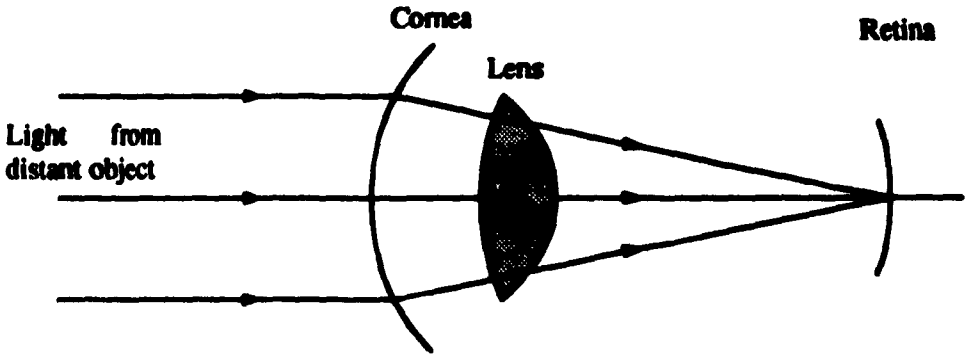


Figure 3.2: Schematic illustration of the formation on the retina of the image of a distant point source of light. From [12].

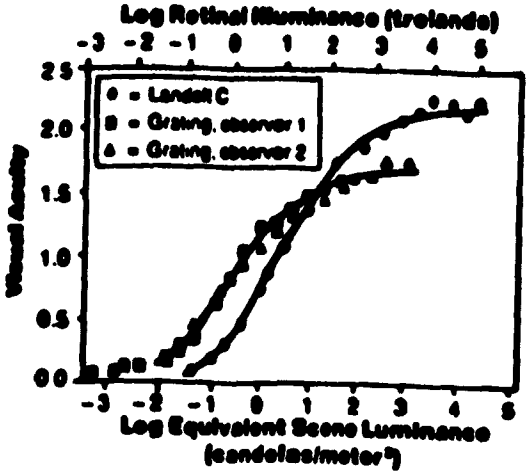


Figure 3.3: Visual acuity as a function of background luminance measured with a Landolt-C target (1 observer) and square-wave grating (2 observers). Data are averages across one observation at each of four orientations; observers dark-adapted for 20 min. From [28].

The aperture of the optical system of the eye is controlled by an iris that forms a pupil of 7 to 8 mm diameter in dark conditions to a minimum of about 2 mm in a high luminance level. The diffractive effect due to the small size of the pupils sets the highest acuity of the human eye, which is based on the Rayleigh criterion. Considering other factors, such as luminance level, target contrast, location of the target on retina, etc., the human visual acuity is lower than that set by the Rayleigh criterion.

3.1.2 Visual Acuity

This section discusses factors affecting human visual acuity. Visual acuity is the ability to resolve spatial detail. Decimal acuity is expressed as the reciprocal of the smallest pattern or pattern detail (in minutes of arc of visual angle) that can be detected or identified at the given viewing distance. The factors affecting the visual acuity mainly include luminance level and contrast, target location on visual field, and viewing distance.

Luminance Level and Target Contrast

Figure 3.3 [28] shows that 1) visual acuity improves with increasing luminance and 2) acuity approaches an asymptote at a luminance of $40 - 1000\ Cd/m^2$, depending on the type of target used. Usually most monochrome CRTs are capable of area brightness in excess of $300\ Cd/m^2$. The brightness of an LCD depends on the light source it modified. A color 2.7 in. LCD Sony Watchman has a luminance from $150\ Cd/m^2$ to $220\ Cd/m^2$. Usually current

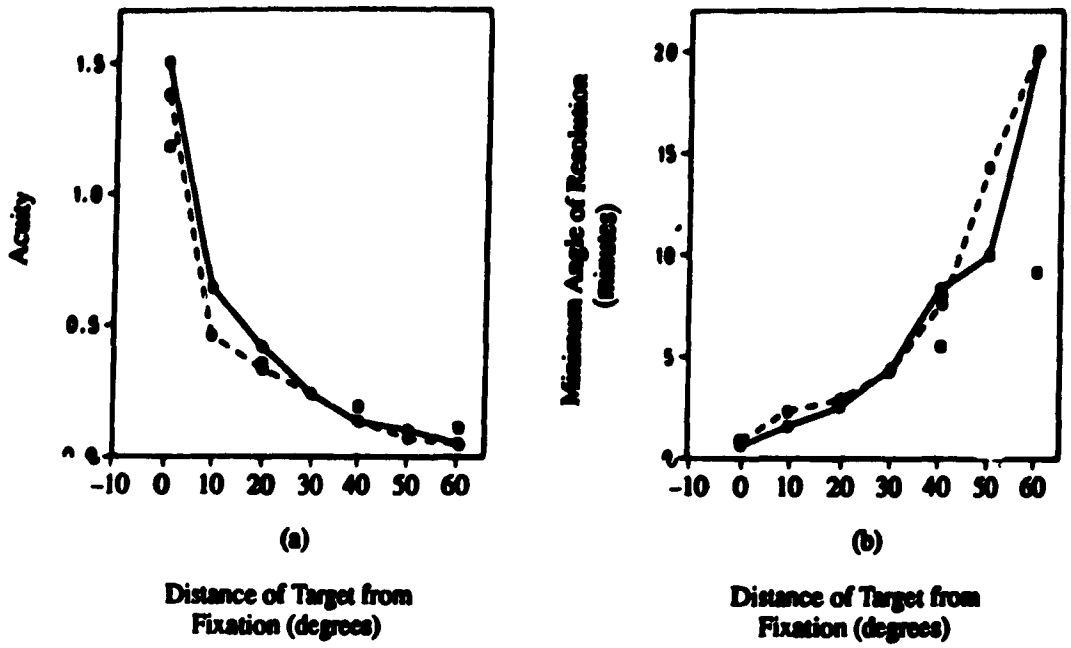


Figure 3.4: Visual acuity as a function of target distance from fixation. (a) Decimal acuity. Three different targets were used: Landolt-C rings at a background luminance of 2.45 cd/m^2 (open circles); Landolt-C rings at a background luminance of 245 cd/m^2 (close circles) and sine-wave gratings (squares). For the Landolt-C target, acuity is the reciprocal of the width (in minutes) of the smallest gap that can be localized; for the grating target, acuity is the reciprocal of the width (in minutes) of one bar of the finest grating that is resolvable into bars. (b) Minimum angle of resolution (reciprocal of decimal acuity) for the three targets. From [19].

displays can provide enough luminance, and are not a factor affecting visual acuity.

Studies also show that, at a given luminance level, acuity increases as target contrast or target exposure duration increases [2].

Target Location in Visual Field

Figure 3.4 shows that for photopic (daylight) levels of illumination ($> 0.03\ Cd/m^2$), the acuity of the human eye decreases as the target distance from the fixation point (retinal eccentricity) increases [19]. Two useful conclusions can be drawn from Figure 3.4.

- The minimum angle of resolution at a field angle of 30 degrees is about 5 minutes. Considering other factors that may decrease the visual acuity, a lateral color criterion of less than 7 minutes is reasonable.

- The fact that visual acuity at the center of the visual field is much higher than that on the periphery suggests that it is acceptable for an optical device to provide the human

eye with high resolution in the central visual field and low resolution in the periphery. This is useful since it is very hard to correct aberrations in the visual margin field.

Viewing Distance or Dioptric Setting

Figure 3.5 is a compilation of the results about ideal viewing distance from a number of studies [13]; the most notable thing about these is their lack of agreement. The simple conclusion is that there is not an optimum dioptric setting. The best hint is to choose a dioptric setting that suits the specific aberration residuals, as long as the axial setting is within a range of 0.75 and 2.5 diopters and that no part of the field of view is beyond infinity [26].

3.1.3 Binocular vs. Monocular Vision

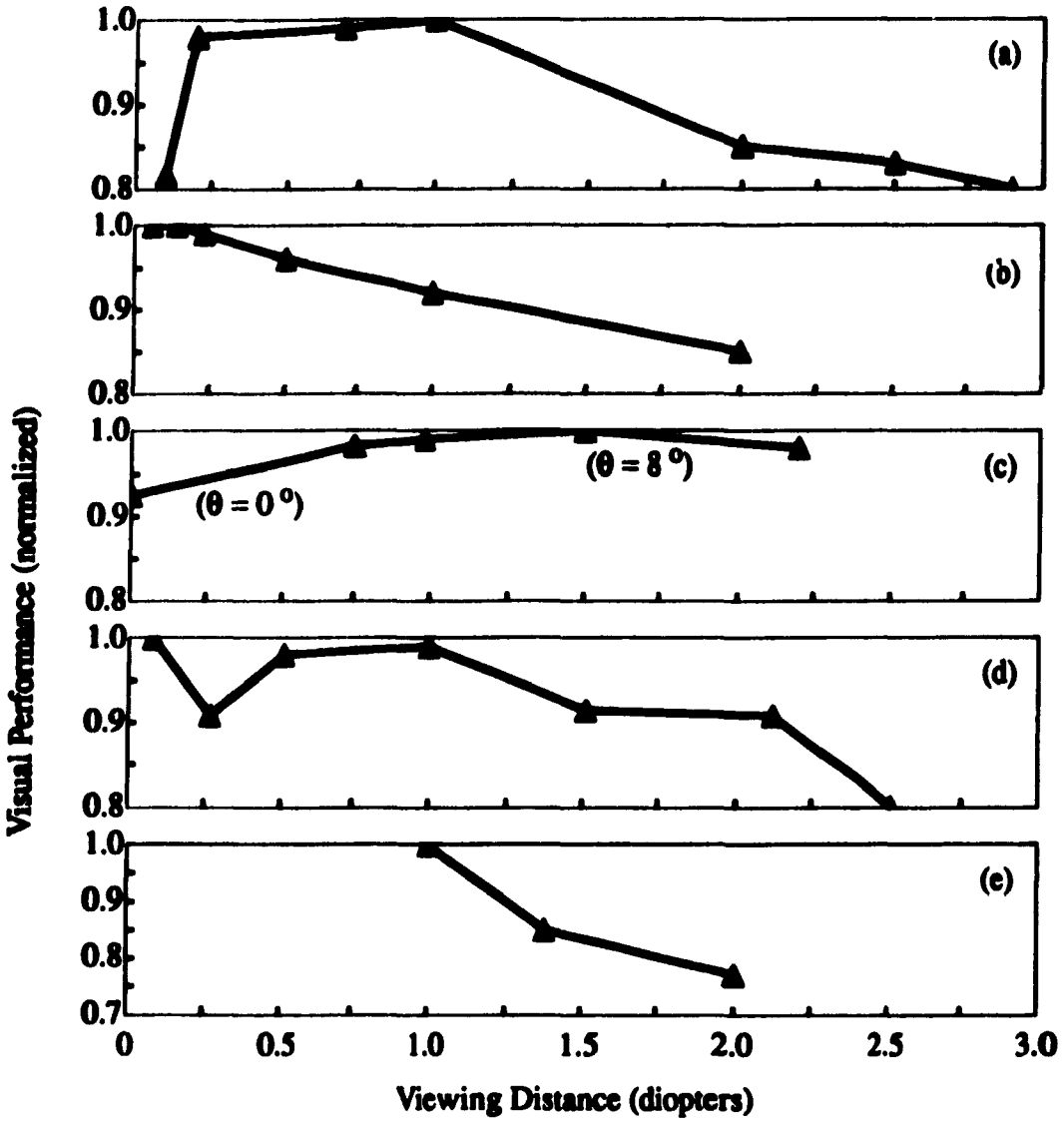
Binocular vision not only provides the advantage of stereopsis, but also produces superior characteristics to monocular vision:

- Visual detection at threshold, including absolute light detection and contrast sensitivity.
- Visual acuity and a high spatial frequency range of the contrast sensitivity threshold function [7] (Figure 3.6). The contrast sensitivity threshold of the eye is defined as the contrast level that can just be detected in a cyclical target of a given angular spatial frequency.
- Form recognition, especially in simple displays.
- Reaction time to onset of light flashes and bar patterns.

There are two possible sources of this binocular advantage. One is probability summation between the eyes, which is due to each eye's independent chance of detecting a stimulus and engenders better overall performance than if only one eye is tested. The other is neural summation which refers to the actual convergence of monocular neural pathways to produce a physiological "sum".

When the perspective points on the two images, due to imperfect setting of IPD and/or centering the eyes on the eyepiece exit pupils, and small errors in magnification of relay optics, are not the same for both eyes, binocular suppression or retinal rivalry will occur.

Hence, binocular vision is better than monocular if retinal rivalry can be avoided and the weight of optical systems is of little importance.



Visual Function Measured	(a) Visual Acuity	(b) Visual Acuity	(c) Stereo Acuity	(d) Stereo Acuity	(e) Stereo Acuity
Type of Test	Checker Board	Cobb 2–Bat Test	CL Stereo Test	2 vertical Bars	2 vertical Bars
Subject	400 Young Adults	7 Adults	32 Selected Adults	3	3

Figure 3.5: Ideal viewing distance (Dioptric setting. Results from five different studies.) From [12].

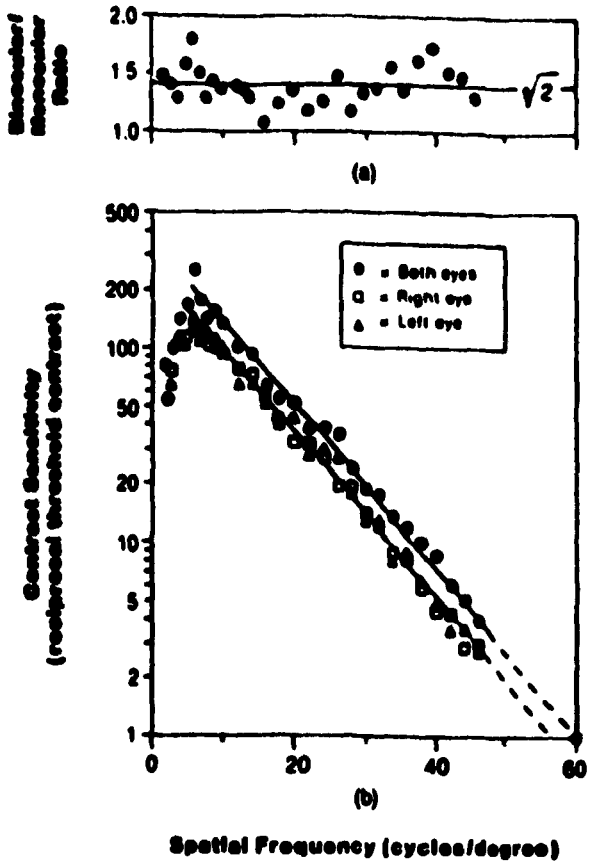


Figure 3.6: (a) Ratio of binocular to mean monocular sensitivity and (b) binocular versus monocular contrast sensitivity. Note logarithmic axis. From [3].

3.1.4 Wide Instantaneous Field of View

The normal, achromatic field-of-view size for a human observer is approximately $\pm 100^o$ horizontally and $\pm 60^o$ vertically. A large IFOV of HMDs allows operators to use both head and eye movements to search large portions of the operation area. Eye movements are faster than head movements, and coordination of the head and eyes to acquire visual targets is a much-practised and well-coordinated activity. In addition, a large IFOV affords operators the use of their peripheral vision. In contrast, with small IFOV, the head and eyes have to point to the same direction for operators to extract the same information. Therefore, the synergistic effect of the eye movement is lost. Moreover, peripheral vision provides less information if the IFOV is smaller.

A HMD with a wide IFOV not only gives operators a realistic sensation of telepresence, but also has the benefit of enhancing safety and increasing efficiency. A study conducted by M.J. Wells and M. Venturino [35] showed that

1. reducing the size of IFOV of a HMD reduced performance at a search and shoot task, as measured by the number of threats hit and the time for which the operator was threatened, and

2. the minimum IFOV required to achieve optimum performance increased with increasing complexity of the task.

The conclusion is that an advanced HMD should have a wide IFOV.

3.1.5 Color Vision

Color would enable the viewer to better discriminate different types of objects, giving, in addition to the normal morphological information of a monochrome display, the cues of color. But human color vision cannot make an objective determination of true color. A uniformly colored background, after awhile, will look like white-grey to the human eye. Thus if color were used to code an event, its accurate identification could be difficult.

3.1.6 Summary

This section discussed the human eye characteristics relative to the design of advanced HMDs. The discussions suggest that an advanced HMD system should have binocular (stereo) vision, color vision, a wide IFOV, adequate contrast ratio and luminance level, reasonable image location setting, and small aberrations.

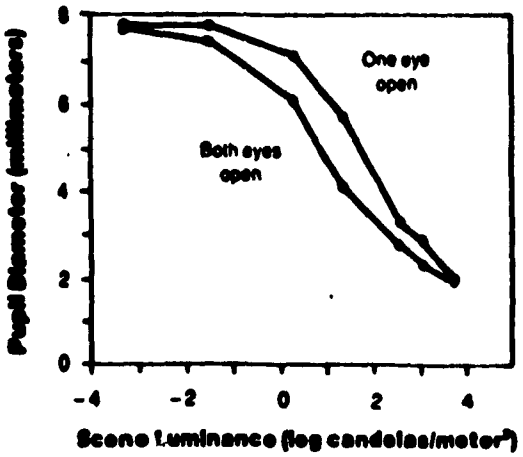


Figure 3.7: Average monocular and binocular pupil diameter as a function of luminance level. From [23].

3.2 Geometry Factors

3.2.1 Pupil Size

The diameter of the pupil of the eye varies from 8 mm to 2 mm with the level of illumination and is smaller when both eyes are illuminated than when only one is illuminated [23] (Figure 3.7). In the daytime, the diameter of the eye is around 4 mm.

The eyeball pivots in its socket about a rotation center which is 10 mm behind the entrance pupil (Figure 3.8). For normal, unaided viewing, the comfort limit for eye rotation from straight ahead is about 40 degrees [12]; however, the observer may avoid approaching this limit by rotation of the head. The positions of the optical relay of a binocular HMD system with respect to the helmet remain unchanged when the pupil positions vary. This places a demand on display optics to have exit pupils much larger than the nominal pupil size (see Figure 3.9). Assume that the eye rotates by the angle α, the radius of eye pupil is r and the radius of rotation is R, then the diameter of the exit pupil for an unvignetting margin field is

$$D = 2 \cdot (R \sin\alpha + r \cos\alpha) \tag{3.1}$$

Assume that the eye usually rotates within $\pm 20^{\circ}$, and the pupil diameter of the eye is 4 mm (Figure 3.9), then the diameter of the exit pupil is

$$D = 2 \times (10 \times \sin 20^{\circ} + 2 \times \cos 20^{\circ}) = 10.60\, mm \tag{3.2}$$

The optical relay of the BLHMD has an exit pupil of 10 mm. Please note that optical quality will be measured by the quality within the eye pupil, not the display pupil as a

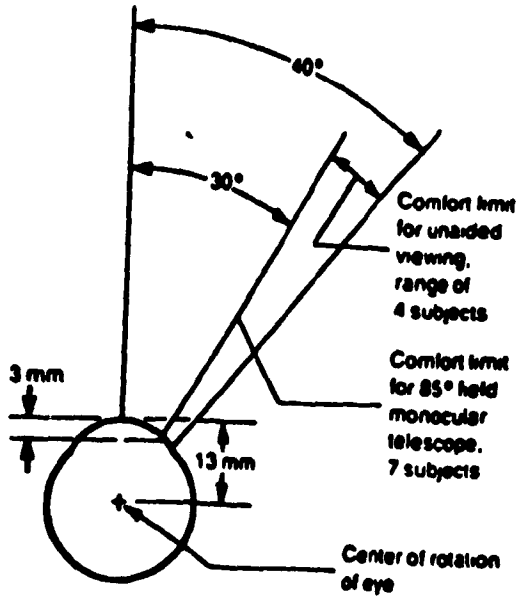


Figure 3.8: Comfort limits of eye rotation for unaided and aided viewing. From [3].

whole.

3.2.2 Interpupillary Distance (IPD)

Interpupillary distance data is important in designing a dual-eye display. The IPD does not have a fixed value either between individuals or even for a single individual. There is a reduction of IPD as the eyes pivot around their centers of rotation to converge on an object at a finite distance. The range of IPD is suggested in Table 3.1 (from [3]). Take the mean value of 63 mm as a reasonable value for IPD.

3.2.3 Eyerelief

Eyerelief is defined as the distance between the last element of the optical display and eye entrance pupil, not cornea, which is 3 mm ahead of the eye pupil.

Between 30 to 50% of the population, aged 20 to 45, will be using spectacles [12], and about half those people prefer to keep wearing spectacles when looking into optical equipment, rather than take them off and refocus. Therefore, viewing the full field of view while wearing eyeglasses requires a large eyerelief, at least 20 mm. A larger eyerelief is critical for a wide FOV visual system to achieve a designed IFOV. But for the situation without glasses, the eyerelief can be as small as 12 mm. The BLHMD has a designed eyerelief of 20 mm.

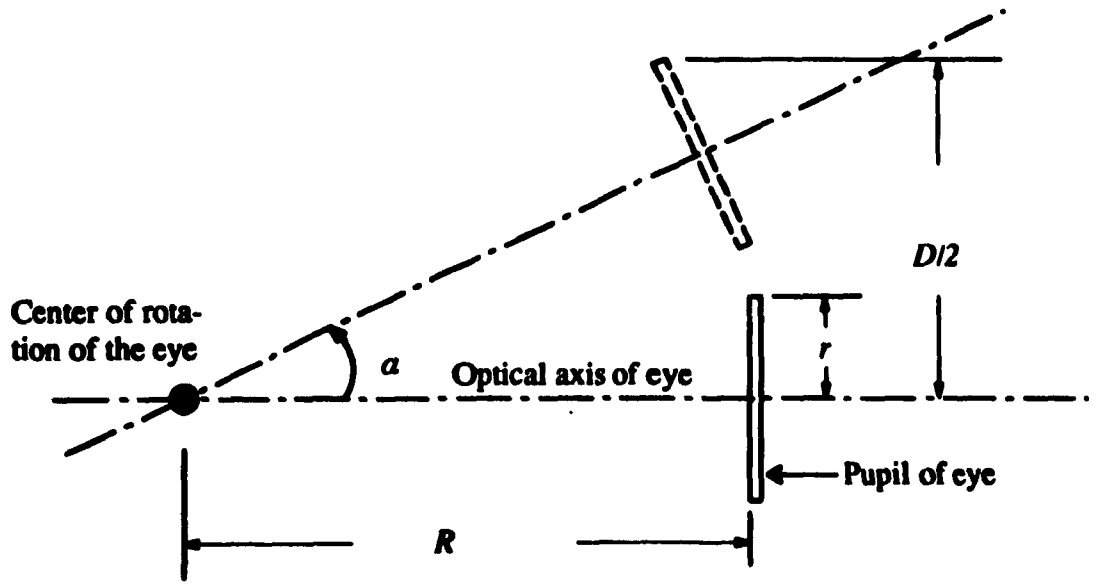


Figure 3.9: The diameter D of the exist pupil of a HMD optical relay, considering the rotation of eye, is larger than the eye pupil. α is the rotation angle of the eye, R is the radius of rotation.

3.3 Aberration Tolerances

It is impossible that a visual optical system can be fully corrected for all Seidel aberrations, unless extremely complex optics are employed. Considering human eye acuity limitations and accommodation ability, it is not necessary to fully correct all aberrations in one visual system. But it is possible and necessary to know visual aberrations and aberration tolerances of the eye before we begin the design of a visual optical system.

Our straight optical relay, a special eyepiece, has a short focal length, a relatively small aperture, and a large IFOV. Hence the oblique aberrations, including lateral color, astigmatism, coma, distortion and field curvature, are the main concerns. These aberrations fall into two categories: image-degrading aberrations (lateral color, astigmatism, coma) and non-image-degrading aberrations (distortion and field curvature). In this section, we will find out how much aberrations are acceptable in our BLHMD.

3.3.1 Aberration Tolerances of Human Eye

How much aberrations are acceptable is a very important and complex question which may not be regarded as solved. This is because the aberration tolerances of an optical system are dependant on image evaluation methods, the conditions and requirements of the application

Group	Number in Sample	Percentile Rank 5	50	95	SD	Ref
Aviators	4057	57.7	53.2	69.9	3.6	[16]
Army drivers (white)	431	54.1	58.9	64.0	3.0	[9]
Army drivers (black)	79	57.9	62.0	71.7	3.8	[9]
Army & Air Force (females)	3205	51.0	–	65.0	–	[1]

SD = standard deviation of measurement within the group

Table 3.1: Interpupillary distance for different groups, based on observations taken from military personnel. From [3].

of the system, and the characteristics of the receiver (such as human eyes) following the system.

Please note that the use of the word "tolerance" in this connection does not carry the same go, no-go connotation that it does in matters mechanical where parts may suddenly cease to fit or function when tolerances are exceeded. *Any* amount of aberration degrades the images; a larger amount simply degrades it more.

Rayleigh Criterion

Although being related to each other, directly or indirectly, various image evaluation approaches, such as the Rayleigh criterion, modulation transfer function (MTF), and image energy distribution, are from different viewpoints, and their applications are all limited to particular optical systems. For example, the MTF is widely used to describe the performance of a lens system, but it has been shown that the MTF is not a useful quality test for visual optical systems [6]. Of these methods, the Rayleigh criterion is practical and convenient. The Rayleigh criterion employs optical path difference (OPD) to evaluate images, and OPD has direct and simple relations with Seidel aberrations. Hence, aberration tolerances can be obtained from the Rayleigh criterion.

An image point may be thought to be formed by a spherical wave. When the actual image point departs from the ideal or reference point, the maximum departure of the actual spherical wave front from an ideal sphere is often referred to as peak-to-peak of peak-to-valley (P-V) OPD. If the wave front is irregular, the root-mean-square (RMS) OPD is a

better measure of the effect of the wave-front deformation. RMS OPD is the square root of the mean of the squares of all the OPD values over the full aperture of an optical system. For the detailed derivation of the relation between aberrations and OPD see [29, pages 327 - 339].

The Rayleigh criterion allows no more than one-quarter wavelength of OPD over the wave front with respect to a reference sphere about a selected image point in order that the image may be *sensibly* perfect.

The following tabulation [29] indicates the amount of image-degrading aberration corresponding to one-quarter wavelength of OPD when the reference point is chosen to minimize the P-V OPD.

$$OSC = \frac{0.5\lambda}{H' NA} \tag{3.3}$$

$$Lateral\ color = \frac{\lambda}{NA} \tag{3.4}$$

$$Astigmatism = \frac{\lambda}{n \sin^2 U_m} \tag{3.5}$$

$$\tag{3.6}$$

where λ is the wavelength of the light, n is the index of the medium in which the image is formed, U_m is the slope angle of the marginal axial ray at the axial image, $NA = n \sin U_m$ is the numerical aperture, and H' is the image height. In the BLHMD system, $\lambda = 5500\ \AA = 5.5 \times 10^{-4}\ cm$, $n = 1$, $U_m = 30^o$, $NA = 0.5$, and $H' = 25\ cm$. From the equations above, we obtain

$$OSC = 2.2 \times 10^{-5} \tag{3.7}$$

$$Lateral\ Color = 2\lambda \tag{3.8}$$

$$Astigmatism = 4\lambda \tag{3.9}$$

Experimental studies of the effects of Seidel aberrations (primary defocus, spherical aberration, astigmatism, and coma) on visual target discrimination by Burton and Haig [6] determined the threshold values of the human visual system to different levels and combinations of the aberration types by a forced-choice discrimination technique. The resulting threshold levels, expressed in units of wavelength, specify the changes in wave-front aberration that can be detected with some defined probability and represent *just-noticeable differences* in image quality. The results agreed well with the Rayleigh criterion. Table 3.2 shows aberration values that could be discriminated in high contrast targets at 75% prob-

Aberration	OPD Level	Level Allowing for Refocusing
Spherical Aberration	0.21 λ	0.67 λ
Coma	0.46 λ	0.46 λ
Astigmatism	0.30 λ	0.39 λ
Defocus	0.22 λ	–

Table 3.2: Visual aberration values that could be discriminated in high contrast targets at 75% probability levels. From [6].

ability levels.

The image-degrading aberration tolerances shown above, whether derived from the Rayleigh criterion or from experiments, only are the image-degrading aberration values in wave-front "that can be detected with some defined probability and represent just-noticeable differences in image quality" [6]. These values are only suitable for optical systems with small aberrations such as objectives of telescopes and microscopes. It is extremely difficult to correct a system to this level of quality over an appreciable field. Most optical systems exceed this allowance many times over.

For optical systems with large aberrations exceeding the Rayleigh criterion many times, geometrical image energy distribution may be used to predict the appearance of a point image with a fair degree of accuracy. It is difficult to attempt to derive the aberration tolerances of the human eye based on theory because of the coherent coupling between the optics and the eye, and the non-linear behavior of the latter.

As for non-image degrading aberrations, field curvature and distortion, only emprical aberration tolerances can be found [26, 33, 19].

In short, the aberration tolerances for the optical relay can be summarized [26, 33, 19]:

- Distortion: less than $5 \sim 10\%$.
- Lateral color: less than $5 \sim 7$ minutes.
- Field curvature: the tangential field curvature lies within 0.9 diopters of the central image plane, the sagittal field within 3 diopters, and no part of the fields is beyond infinity.
- Astigmatism $<$ 2 diopters.
- Coma: OSC is less than 0.0025.

3.4 Optical Design Criteria

In this chapter, we discussed the characteristics of the human eye, which has a separate optical system and retina to receive and form an image of a scene. It has been shown that use of two eyes allows a better perception of depth and an improvement in visual acuity. Color images present more information than monochromic. Hence color stereo vision will be more acceptable for human eyes than monocular and monochrome vision. Based on these discussions, we believe that an advanced HMD should have binocular (stereo) and color vision.

Considering factors affecting visual acuity of an eye, such as luminance level, contrast, retinal location, and viewing distance, the ideal conditions that a visual system provides to the eye are adequate luminance ($40 - 1000\ Cd/m^2$), high contrast, high resolution on the central visual field, and reasonable dioptric setting.

The following design criteria of the optical relay summarize the discussions of Section 3.2 and Section 3.3.1 on geometry factors and aberration tolerances of the human visual system.

Optical Design Criteria

Human Factor Constraints:

- IFOV > 60^o
- Diameter of lenses < 55 mm
- Eyerelief > 18 - 20 mm
- Diameter of the exit pupil > 10 mm
- Total length of the straight optical relay < 100 mm
- as light as possible

Aberration Tolerances:

- Distortion on full field $< 10\%$
- Lateral color $< 5 \sim 7$ minutes
- Field curvature: the tangential field curvature lies within 0.9 diopters of the central image plane, the sagittal field within 3 diopters, and no part of the fields is beyond infinity.
- Astigmatism < 2 diopters
- Coma: $OSC < 0.0025$

Other Constraints:

- All spherical lenses
- Display size ($41.1mm \times 54.4mm$)

Chapter 4

Optical Design of HMD

4.1 Introduction

In Chapter 3, we discussed the desired characteristics of the BLHMD optical relay. This chapter determines the structural parameters, based on the design theories discussed in Chapter 2.

Optical design is a process in which the desired characteristics of an optical system are given and the constructional parameters are to be determined. A large part of optical design is concerned with analysis of the system. Figure 4.1 is a simple chart of the process.

For computer aided lens design, there are several steps to follow. The first and very important step is to choose a likely starting system. There is no sure procedure that will lead (without foreknowledge) from a set of characteristics to a suitable starting design. Usually a likely starting approach to the desired lens can be obtained by either a mental guess which may work well for an experienced designer, or a search through previously designed lenses in patent files. The second step is to construct a merit function which describes all desired characteristics, which are suitably weighted according to their importance. The last step is optimization of the lens according to the merit function and initial starting point. Almost all of the design work in this thesis was carried out on ACCOS V, a lens design software by Optikos Corporation, New York.

This chapter first reviews the design history of the BLHMD optical relay, then focuses on two types of approaches to the BLHMD optical relay, straight and folded designs. Basic design considerations and design procedures are discussed. The resultant designs and their analyses are also presented.

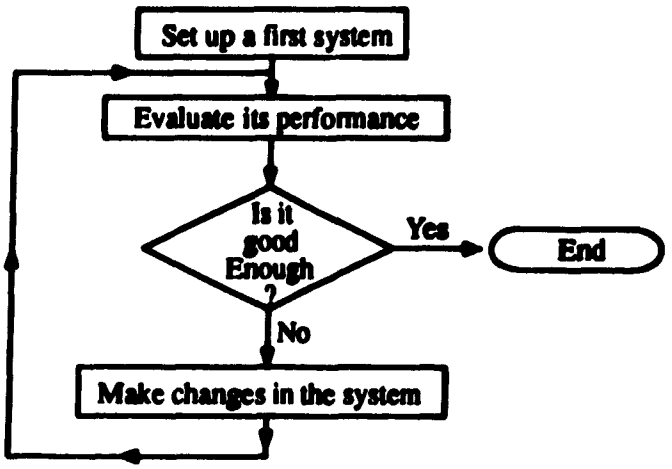


Figure 4.1: Lens design flow chart. From [17].

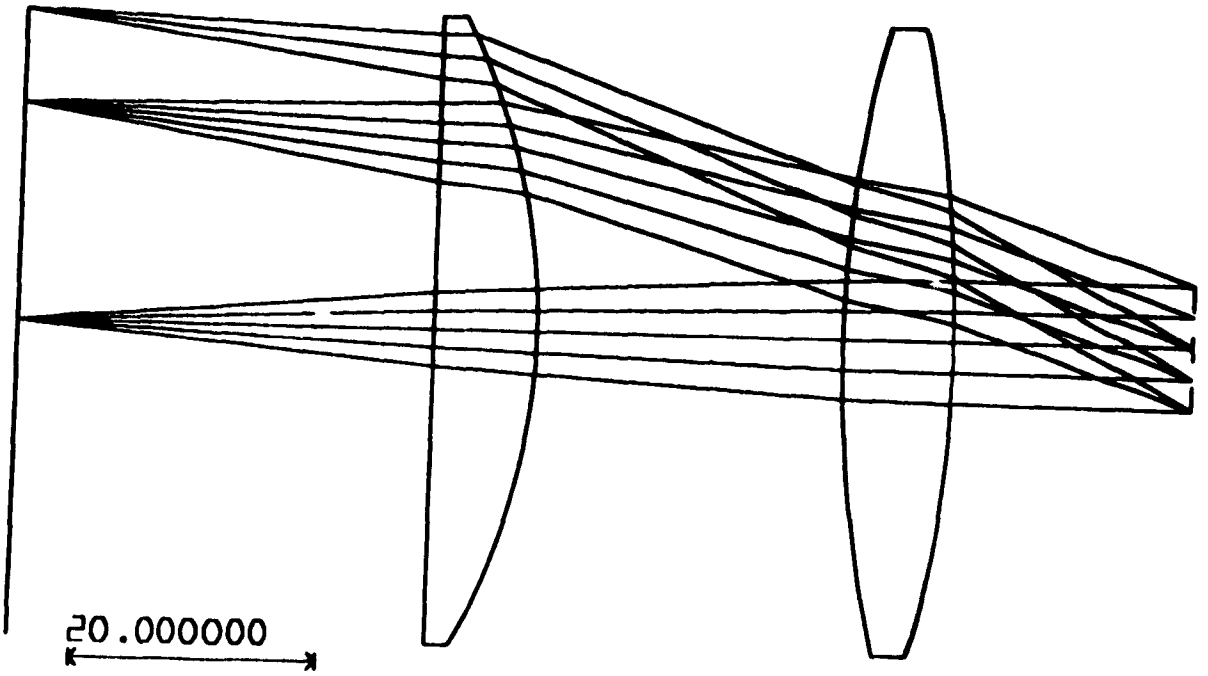


Figure 4.2: The first design of the BLHMD optical relay.

Surface	Radius (mm)	Thickness (mm)	Medium	Index	Clap (mm)
obj	0.000	37.00	AIR	1.000000	
1	52.00	8.22	BK7	1.516799	25.00
2	0.00	25.00	AIR	1.000000	25.00
3	102.500	8.700	BK7	1.516799	25.00
4	-102.500	20.00	AIR	1.000000	25.00
5	0.000	-1076.122	AIR	1.000000	5.00
6	0.000	0.000	AIR	1.000000	

Table 4.1: Structural parameters of the first BLHMD optical relay.

4.2 History of BLHMD Optical Relay Design

The first BLHMD optical relay (Figure 4.2) has a structure similar to a Ramsden eyepiece. It consists of a plano-convex lens and a biconvex lens, and the structural parameters are summarized in Table 4.1.

Aberration analyses in Figure 4.3 using ACCOS V show some major drawbacks in this design:

- the non-vignetted IFOV is as small as 42^o, and
- the aberrations are large: the distortion is 16% and the lateral color is 18 minutes. They make the useful IFOV even smaller. The coma (OSC) is surprisingly small. The field curvature is not good enough, and a part of the FOV goes beyond infinity.

Certain improvements in aberrations can be achieved by reversing the field lens (the one adjacent to the object), i.e., letting the convex side of the field lens face the object (Figure 4.4). The aberrations of this improved design are illustrated in Figure 4.5. Still, the IFOV is smaller than the 60^o required, and the aberrations do not meet our design criteria.

The design experience of the first optical relay showed that the biggest problem in the optical relay design is the conflict among the requirement of a wide IFOV, the limit on the size of lenses, and the aberration tolerances of the human eye. Our next approach to the BLHMD optical relay is an eyepiece with a complex structure. Eyepieces having a wide file of view ($> 60^o$) always cause problems that have to do with both aberrations and logistics. The logistics problem arises from the fact that the eye pupil is not at the center of rotation

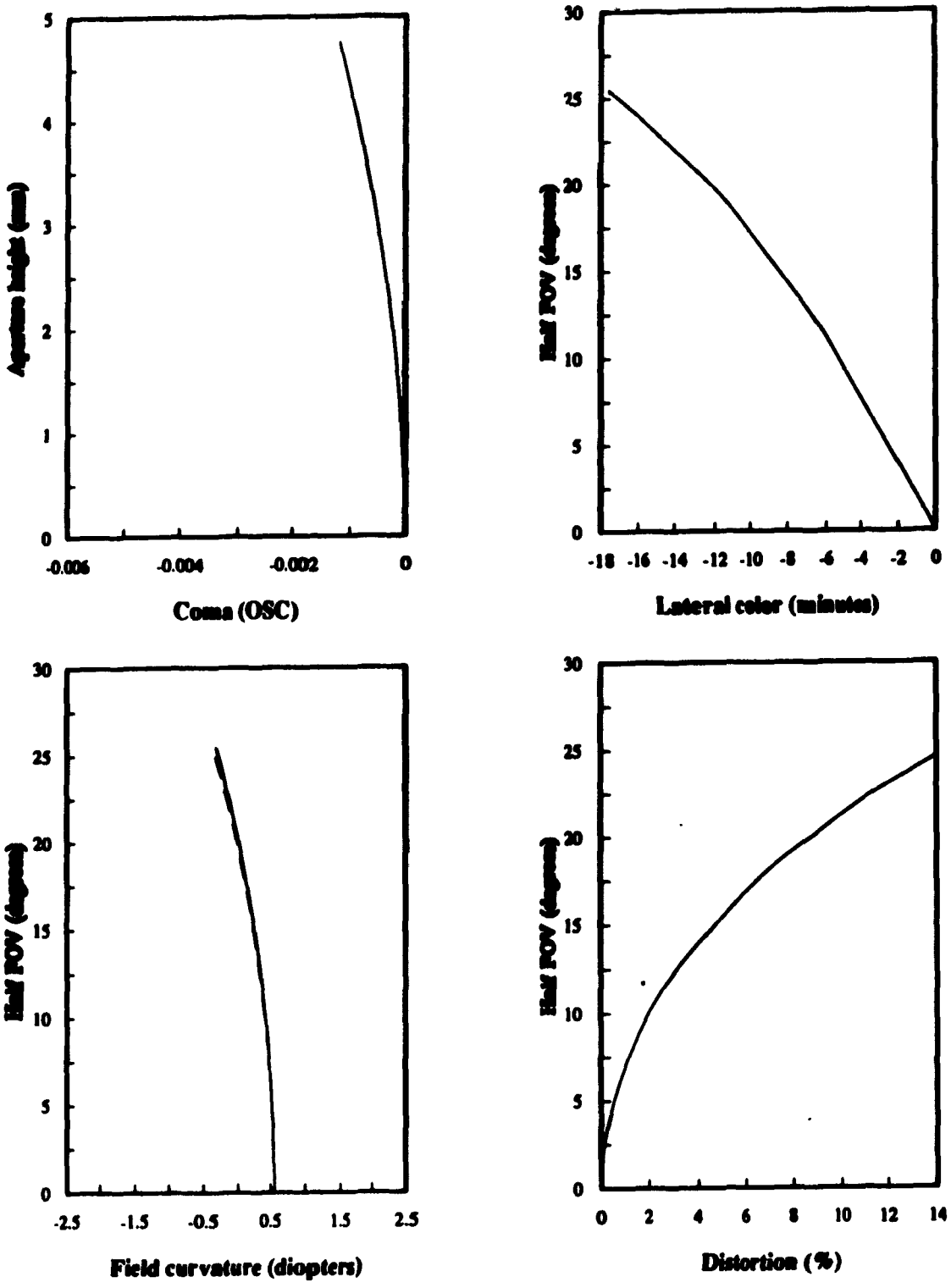


Figure 4.3: Aberration analyses of the first optical relay.

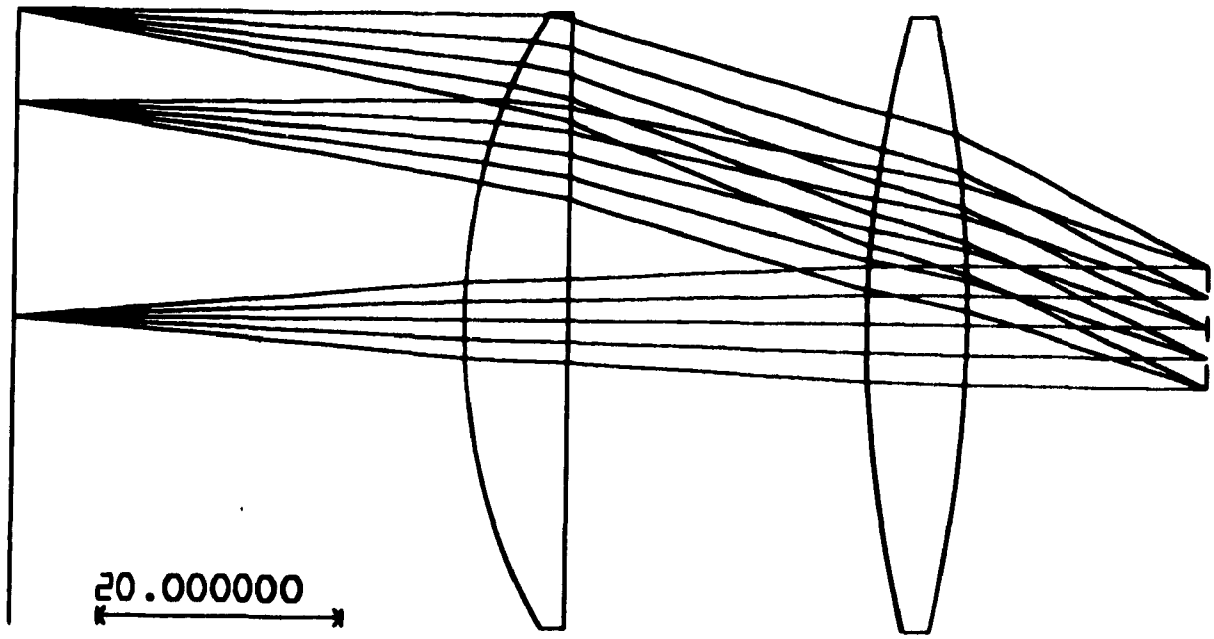


Figure 4.4: Optical relay improved from the first design.

of the eye. Rotation of the eye in order to view the edge of the field of view can result in significant or total vignetting. Figure 4.6 shows this effect.

Besides the requirement of a wide IFOV, other conditions which complicate the design task of an eyepiece are (1) a relatively large eye relief, and (2) a large exit pupil.

In order to have a wide IFOV, a large eye relief and a large exit pupil, and to meet other design criteria obtained in Chapter 3, the relay lens system should have a more elaborate structure and more available parameters than the first design.

4.3 Straight Design

This section discusses the construction of a starting system for the straight optical relay and the optimization procedure.

4.3.1 Basic Structure of the Optical Relay

The first optical relay is a simple eyepiece. Based on this structure, we construct the basic structure of the straight optical relay through the following steps.

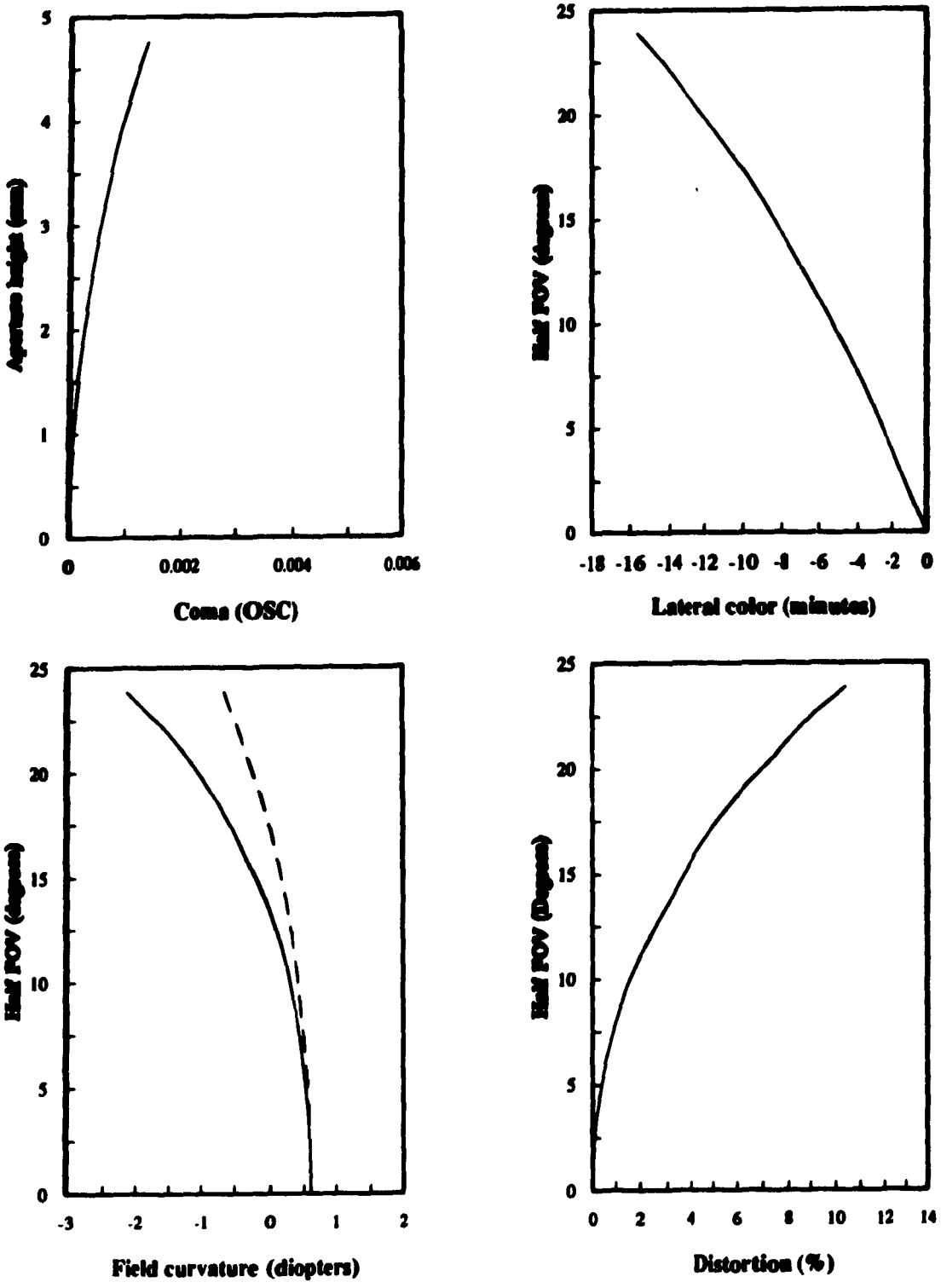


Figure 4.5: Aberration analyses of the improved first optical relay design.

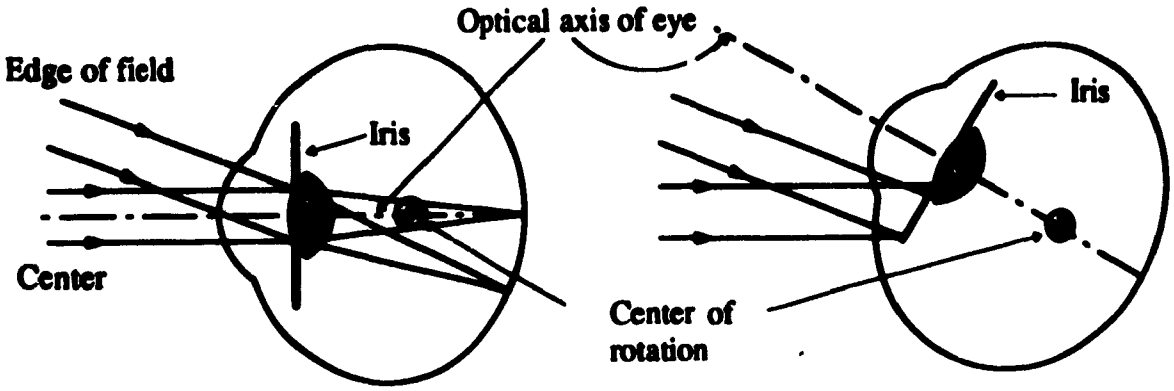


Figure 4.6: The logistics problem with wide IFOV eyepieces. Rotation of the eye to view the edge of FOV may lose all of the picture.

Step 1 Replace the two single lenses with two doublets to control the lateral color. This type of eyepiece, consisting of two doublets, is called the military eyepiece [17].

Step 2 Insert a biconvex element between the two doublets as a relay such that marginal rays of the wide field still can pass through the doublet near the eye (called the eye lens). Thus the system can achieve a 60^o IFOV. Meanwhile, since the biconvex lens shares part of the power of the eyepiece, it should weaken the inner convex surfaces of the two doublets to keep the same focal length. This reduces aberrations as well. This type of eyepiece is the Erfle eyepiece patented in 1921 by H. Erfle [11].

Step 3 Add a positive meniscus lens right after the eye lens to increase the eye relief. The reason for this is explained in the next section.

The basic structure of the starting system thus consists of four elements: two doublets, one biconvex lens and one meniscus lens. The thickness and the optical material of each lens, and the power distribution among the lenses, remain to be determined.

4.3.2 Further Considerations

For an eyepiece of a given focal length, it is desirable to make the eye relief as large as possible. This can be realized by concentrating the convergent power in the element or elements nearest the eye, and reducing the power (or even making it negative) in those adjacent to the object [27]. This actually pushes the principal points toward the eye and the exit pupil plane right along them. That is why a positive meniscus lens is added right after the doublet near the eye. Meanwhile, the power of the doublet adjacent to the object

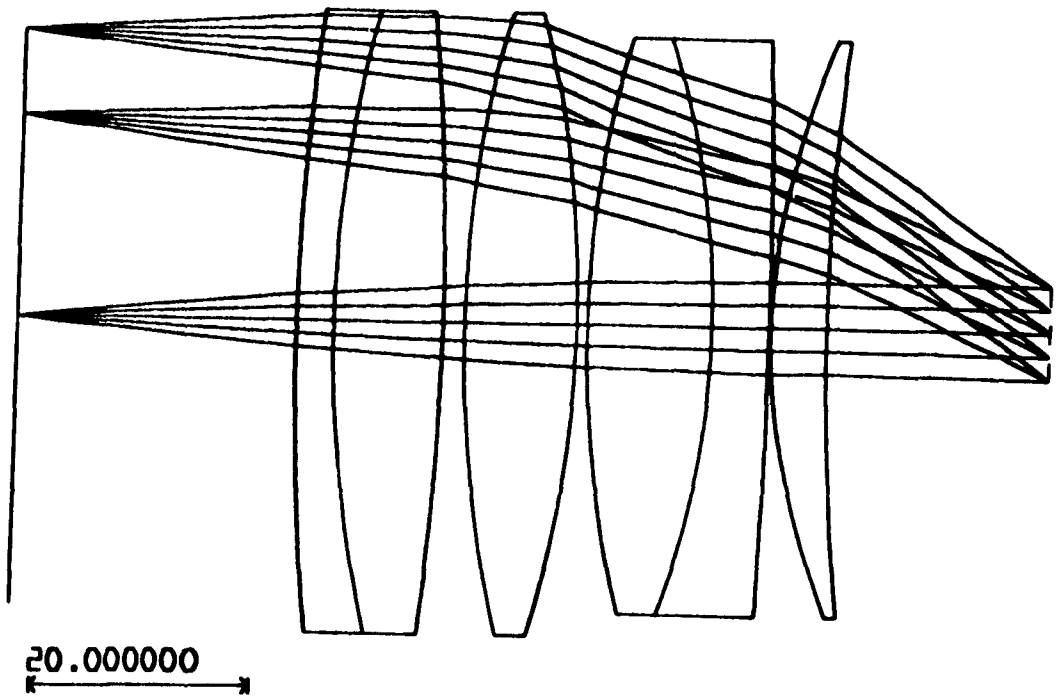


Figure 4.7: The starting system of the BLHMD straight optical design.

has to be reduced to keep the power of the system. In view of these considerations, we assign a power of 0.19 to the field lens and of 0.40 to the middle lens. The eye lens (including one doublet and one meniscus lens) comes out to have a power of about 0.47 for an overall focal length of 1.0. It is possible that other distributions might be better.

The central thicknesses of the lenses were chosen such that 1) the central thickness of each lens is larger than 3 mm, and 2) the edge thickness at a radius of 25 mm is larger than 2 mm. In the selection of optical glasses, two issues have to be considered: chromatic aberration and cost. Since the two doublets are positive, according to the discussion in Chapter 2, the positive elements in the doublets must use a low-relative-dispersion glass (BK7 and KF50 were chosen here), and the negative elements must use a high-relative-dispersion glass (we selected F4). The middle biconvex lens and the meniscus lens are chosen as BK7, a very common and high quality optical glass.

Table 4.2 summarizes the parameters of this starting system. The system layout is illustrated in Figure 4.7.

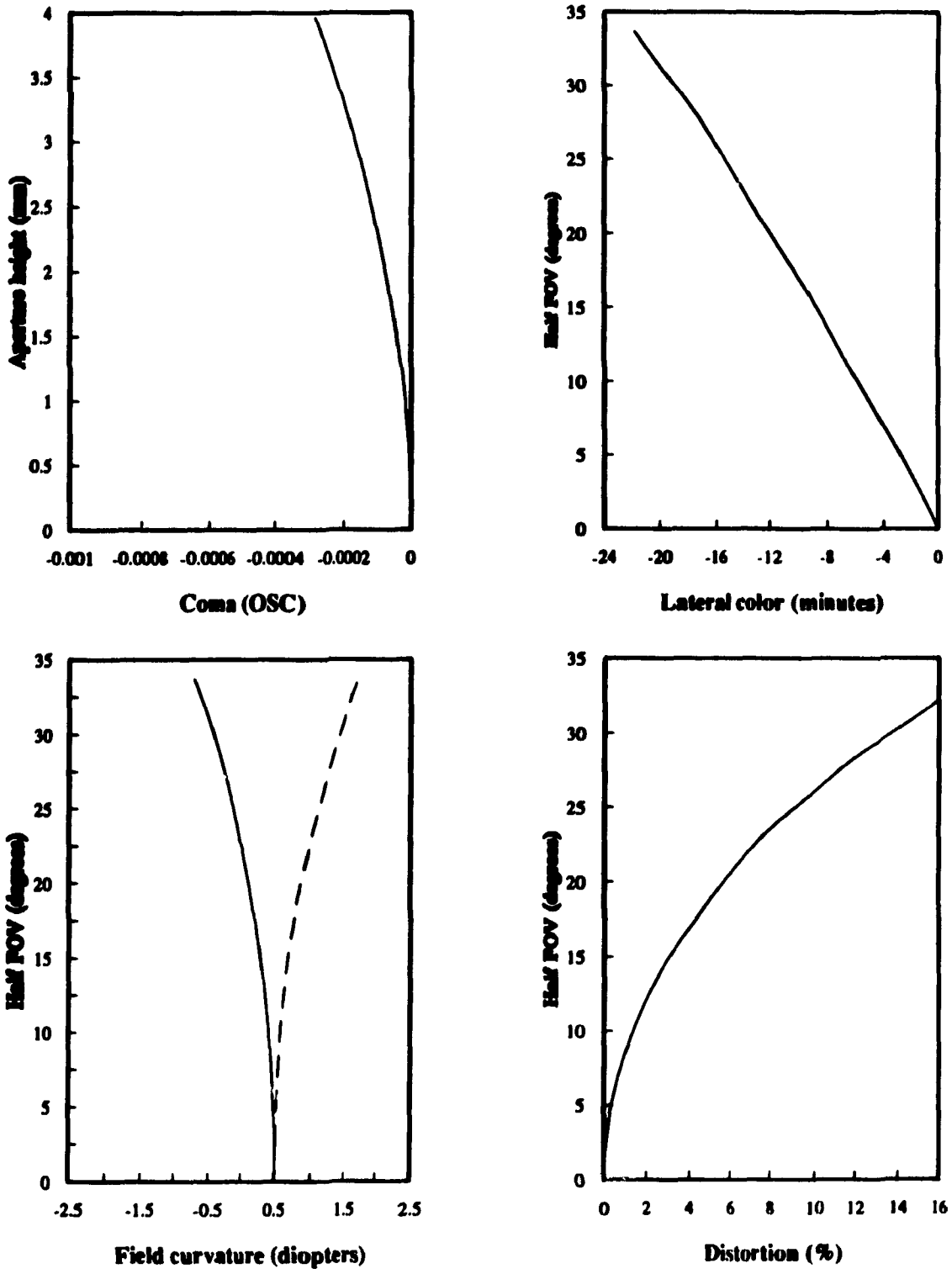


Figure 4.8: The aberration analyses of the starting system of the straight optical design.

Surface	Radius (mm)	Thickness (mm)	Medium	Index	Clap (mm)
obj	0.000	25.200	AIR	1.000	
1	250.00	3.000	F4	1.616591	27.00
2	100.00	10.00	KF50	1.530881	27.00
3	-250.00	2.00	AIR	1.000000	27.00
4	100.00	10.00	BK7	1.516799	27.00
5	-100.00	1.00	AIR	1.0000	27.00
6	95.00	11.00	BK7	1.516799	25.00
7	-75.00	5.00	F4	1.616591	25.00
8	-600.00	0.10	AIR	1.000000	25.00
9	60.00	5.00	BK7	1.516799	25.00
10	200.00	20.00	AIR	1.000000	25.00
11	0.00	-1228.895	AIR	1.000000	4.00
12	0.00	0.00	AIR	1.000000	

Table 4.2: Structural parameters of the starting system of the BLHMD straight optical relay.

4.3.3 Computer Optimization

The aberration analyses in Figure 4.8 show the characteristics of this starting system and the necessity to change the structural parameters of this system to reduce aberrations.

Since the basic structure of the optical relay is determined, now the design problem becomes: 1) how to construct a merit function consisting of all the design criteria and 2) how to minimize the merit function by varying the structural parameters. The most important design variables are the shapes and the thicknesses of the lenses, and the air spaces between lenses.

Merit Function

The merit function for the BLHMD consists of

1. the four Seidel aberrations (lateral color, coma, distortion, and field curvature) at zonal and/or marginal field;

2. the penalty of violation of geometrical constraints, such as the edges of lenses dropping below a desired value (1.5 mm), two adjacent lens surfaces intersecting, and the IFOV

becoming less than the assigned value; and

3. the departure of the location of the final virtual image from the desired location. This helps to hold the power of the optical relay, since the position of the object is fixed.

The four aberrations are found according to their definitions in Chapter 2 by tracing necessary rays through the system as follows:

- Lateral color: trace both C ($\lambda = 656.3\,nm$) and F ($\lambda = 486.1\,nm$) principal rays from the zonal object point (Y = 0.707 object height) through the system, and obtain the slope angles θ_c and θ_f between rays and optical axis at the exit pupil (both rays must go through the center of the exit pupil, since they are principal rays.). The difference between these two angles is the lateral color for the zonal point. The lateral color for the marginal point (Y = object height) is obtained in the same way.

- Distortion: trace a marginal principal ray, and find the height of the intersection point on the final virtual image plane. The departure of this value from the paraxial principal ray height on the same final image plane is the distortion at the marginal field. The percentage distortion is the ratio of the distortion to the paraxial principal ray height.

- Coma: trace two rays from the zonal object point through the upper and lower rim of the exit pupil separately, and obtain their points of intersection with the image plane. The distance of the middle point of these two points (these two points may coincide) above or below the principal ray is a direct measure of tangential coma at the zonal field. The marginal tangential coma may be found by the same way. Sagittal comas can be obtained by tracing sagittal rays through the front and rear rim of the exit pupil. OSC is equal to the sagittal coma divided by the principal ray height on the image plane.

- Field curvature: trace two rays from the same object point close to the principal ray through the whole system, then calculate their points of intersection in the image space from their slope angles and positions at the exit pupil. The reciprocal of the longitudinal distance in meters along the optical axis between the image point and the exit pupil is the field curvature in diopters.

The violations of geometrical contraints and the departure of the location of the final image are regarded as aberrations, with different weights.

NAME	SURF(S)	MODE	VALUE	TARGET	WEIGHT	FMTC
CFG 1						
IMD	0 TO 12	COR	1228.895432	1228.895432	1.000	0.00E+00
EDGE	3	BYP	7.176244	1.500000	5.000	0.00E+00
EDGE	5	BYP	8.631587	1.500000	1.000	0.00E+00
EDGE	1	BYP	5.251691	2.000000	1.000	0.00E+00
EDGE	2	BYP	4.823756	1.500000	1.000	0.00E+00
EDGE	4	BYP	2.572064	1.500000	1.000	0.00E+00
EDGE	6	BYP	3.362192	1.500000	1.000	0.00E+00
EDGE	9	BYP	1.112212	1.500000	1.000	0.00E+00
PUPIL	11	BYP	4.000000	0.000000	1.000	0.00E+00
YA2	12	BYP	-0.589005	0.000000	1.000	0.00E+00
VWA	12	COR	-0.593481	-0.550000	100.000	1.89E+01
YA4	12	BYP	-0.598748	0.000000	1.000	0.00E+00
TCOM1		COR	1.766141	0.000000	2.000	1.25E+01
TCOM7		COR	0.102722	0.000000	2.000	4.22E-02
TAFC1		COR	-8.141370	0.000000	2.000	2.65E+02
TAFC7		COR	-6.649082	0.000000	2.000	1.77E+02
SCOM1		COR	0.256889	0.000000	2.000	2.64E-01
SCOM7		COR	-0.134102	0.000000	2.000	7.19E-02
SAFC1		COR	-20.929600	0.000000	1.000	4.38E+02
SAFC7		COR	-25.549946	0.000000	1.000	6.53E+02
LCM		COR	22.171365	0.000000	1.000	4.92E+02
LCZ		COR	14.351964	0.000000	1.000	2.06E+02
DISF		COR	18.205535	0.000000	1.000	3.31E+02

FIGURE OF MERIT = 2.59357E+03

Table 4.3: Merit function of the starting system of the BLHMD straight optical relay.

The final figure of merit (FMT) is the weighted sum of squares of all the aberrations. The smaller the FMT, the better the design. Table 4.3 summaries the merit function for the starting system In the table,

IMD is the distance between the exit pupil and the final virtual image,

VWA is the half viewing angle in radians (the minus sign comes from conventions of ray tracing in ACCOS V),

PUPIL is the diameter of the exit pupil in mm,

TCOM1 is the scaled tangential coma at the marginal field,

TCOM7 is the scaled tangential coma at the zonal field,

SCOM1 is the scaled sagittal coma at the marginal field,

SCOM7 is the scaled sagittal coma at the zonal field,

TAFC1 is the scaled tangential field curvature in diopters at the marginal field,

TAFC7 is the scaled tangential field curvature in diopters at the zonal field,

SAFC1 is the scaled sagittal field curvature in diopters at the marginal field,

SAFC7 is the scaled sagittal field curvature in diopters at the zonal field,

LCM is the lateral color in minutes at the marginal field,

LCZ is the lateral color in minutes at the zonal field, and

DISF is the percentage distortion at the marginal field.

Optimization

The variables, their low and high limits, and their weights must be specified before optimization. Which parameters to be chosen as variables indicates which optimization path to go on and which local minimum to be found. This depends on which aberrations are important in the current system and which parameters are sensitive to these aberrations. The answers to the latter problem is usually found by analyses of the system or by the designer's experience. It is not a good idea at the first design stage to let all available parameters vary at the same time, because this not only takes a huge amount of computing time, but also is not sure to reach a better design.

ACCOS V provides three kinds of optimization routines.

- Damped-least-squares (DLS) optimization (Levenberg-Marquardt algorithm). Automatic adjustment of the damping factor is provided. Solutions for which the FMT fails to improve are rejected. This leads to one local minimum.

- Pseudo-global search (PGS) optimization. This is different from DLS optimization in that a diverging solution will be accepted a certain percentage of the times that the damping factor would normally have been increased and the cycle repeated. Any diverging solutions whose FMT is greater than a certain percentage of the starting FMT will be rejected.

- Random search for a minimum FMT. Each variable is randomly varied between low and high limits, and the FMT is calculated. The lens system with the smallest FMT is saved.

NAME	SURF(S)	MODE	VALUE	TARGET	WEIGHT	FMTC
CFG 1						
IMD	0 TO 12	COR	1229.731079	1229.731079	1.000	0.00E+00
EDGE	3	BYP	5.422181	1.500000	5.000	0.00E+00
EDGE	5	BYP	11.963428	1.500000	1.000	0.00E+00
EDGE	1	BYP	8.869448	2.000000	1.000	0.00E+00
EDGE	2	BYP	3.428364	1.500000	1.000	0.00E+00
EDGE	4	BYP	3.687344	1.500000	1.000	0.00E+00
EDGE	6	BYP	-0.231663	1.500000	1.000	0.00E+00
EDGE	9	BYP	2.273026	1.500000	1.000	0.00E+00
PUPIL	11	BYP	3.500026	0.000000	1.000	0.00E+00
YA2	12	BYP	-0.509874	0.000000	1.000	0.00E+00
VWA	12	COR	-0.526139	-0.550000	100.000	5.69E+00
YA4	12	BYP	-0.542346	0.000000	1.000	0.00E+00
TCOM1		COR	0.549484	0.000000	2.000	1.21E+00
TCOM7		COR	3.892575	0.000000	2.000	6.06E+01
TAFC1		COR	-11.157120	0.000000	2.000	4.98E+02
TAFC7		COR	-10.685508	0.000000	2.000	4.57E+02
SCOM1		COR	0.584877	0.000000	2.000	1.37E+00
SCOM7		COR	1.703292	0.000000	2.000	1.16E+01
SAFC1		COR	-0.782738	0.000000	1.000	6.13E-01
SAFC7		COR	1.683272	0.000000	1.000	2.83E+00
LCM		COR	-1.844433	0.000000	1.000	3.40E+00
LCZ		COR	2.700431	0.000000	1.000	7.29E+00
DISF		COR	1.886170	0.000000	1.000	3.56E+00

FIGURE OF MERIT = 1.05283E+03

Table 4.4: Merit function of the optimized BLHMD straight optical relay.

We mainly employed DLS and PGS optimizations. After long trial-and-error, we came to a final design, whose merit function is summarized in Table 4.4.

4.3.4 Design Results and Analyses

Table 4.5 summarizes the structural parameters in our final design of the optical relay. This multi-spherical-lens system consists of two doublets, one biconvex lens, and one meniscus lens. All lens diameters are less than or equal to 25 cm. The design was optimized for a full unvignetted 10 mm exit pupil, a 20 mm eyerelief, and a 60° IFOV. The layout of this system is illustrated in Figure 4.9.

The aberrations of this system were analyzed when the system forms a virtual image at a distance of 1.23 meters away from the exit pupil. Figure 4.10 illustrates the analysis

Surface	Radius (mm)	Thickness (mm)	Medium	Index	Clap (mm)
obj	0.000	26.00	AIR	1.000	
1	105.471	2.000	F4	1.616591	25.00
2	36.582	15.00	KF50	1.530881	25.00
3	-185.055	1.00	AIR	1.000000	25.00
4	116.011	9.00	BK7	1.516799	25.00
5	-122.092	1.00	AIR	1.0000	25.00
6	41.495	11.645	BK7	1.516799	20.00
7	-33.048	3.00	F4	1.616591	20.00
8	74.531	0.10	AIR	1.000000	20.00
9	26.186	5.041	BK7	1.516799	16.00
10	48.958	20.00	AIR	1.000000	16.00
11	0.00	-1229.73	AIR	1.000000	4.00
12	0.00	0.00	AIR	1.000000	

Table 4.5: Structural parameters of the optimized BLHMD straight optical relay.

results.

The overall characteristics of this final design are satisfactory. The distortion is less than 2.5% on the whole field. Lateral color is less than 3.5 minutes. OSC is less than 0.001 within a 4 mm diameter exit pupil (the same size of the human pupil under daytime viewing conditions), and less than 0.0025 within a 10 mm diameter exit pupil. The tangential field curvature is quite flat, less than 0.8 diopters on the full field. The sagittal field curvature is reasonably good within a $\pm 20^{o}$ field, and quite curved beyond this field. Hence, the outer field is useful only for identifying the presence of a possible target.

The characteristics of the field curvatures may be explained as follows. Due to the constraint of the total length from the display to the exit pupil and the requirement of a wide IFOV, the focal length of the optical relay must be small. This makes the Petzval sum large, i.e., the Petzval surface is unavoidably quite curved. In order to obtain a flat tangential field, overcorrected astigmatism was introduced. This makes the sagittal field more curved.

This eyepiece approach has three disadvantages:

1. all weights of the lenses and the displays are located in front of the head, which may cause fatigue in the operator;

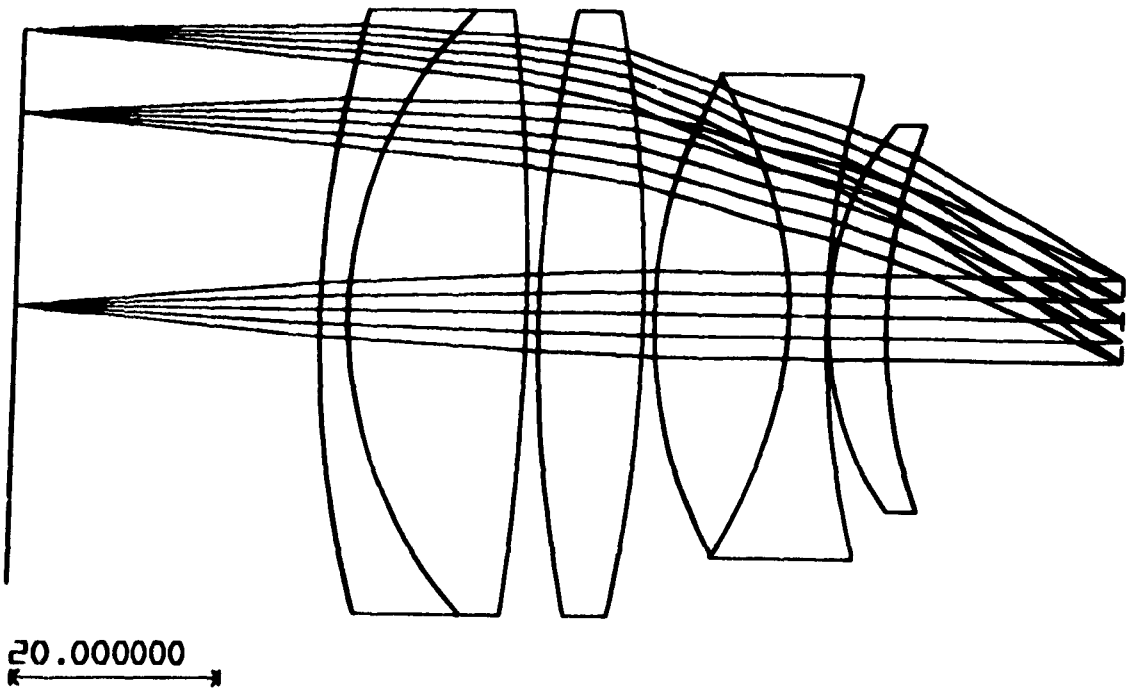


Figure 4.9: The Y-Z layout of the optimized straight optical relay.

2. there is a limitation on the size of the displays set by the human IPD; and

3. some aberrations cannot be corrected by the eyepiece alone.

4.4 Folded Design

In order to adjust the center of gravity of the lenses and the LCDs, and to remove the limitation of the human IPD on the size of the displays, we came up with an alternate optical relay: folding the optical relay and display around head. The approach consists of an eyepiece, an objective, and two folding mirrors.

Another motive of the folded design is the possibility of aberration balance between the objective and the eyepiece.

4.4.1 Starting System

A Pentac-Heliar type objective (shown in Figure 4.11) is chosen as the one in the folded optical relay based on the following considerations:

- to reduce the Petzval sum by utilizing two air separations between the positive and the negative components,

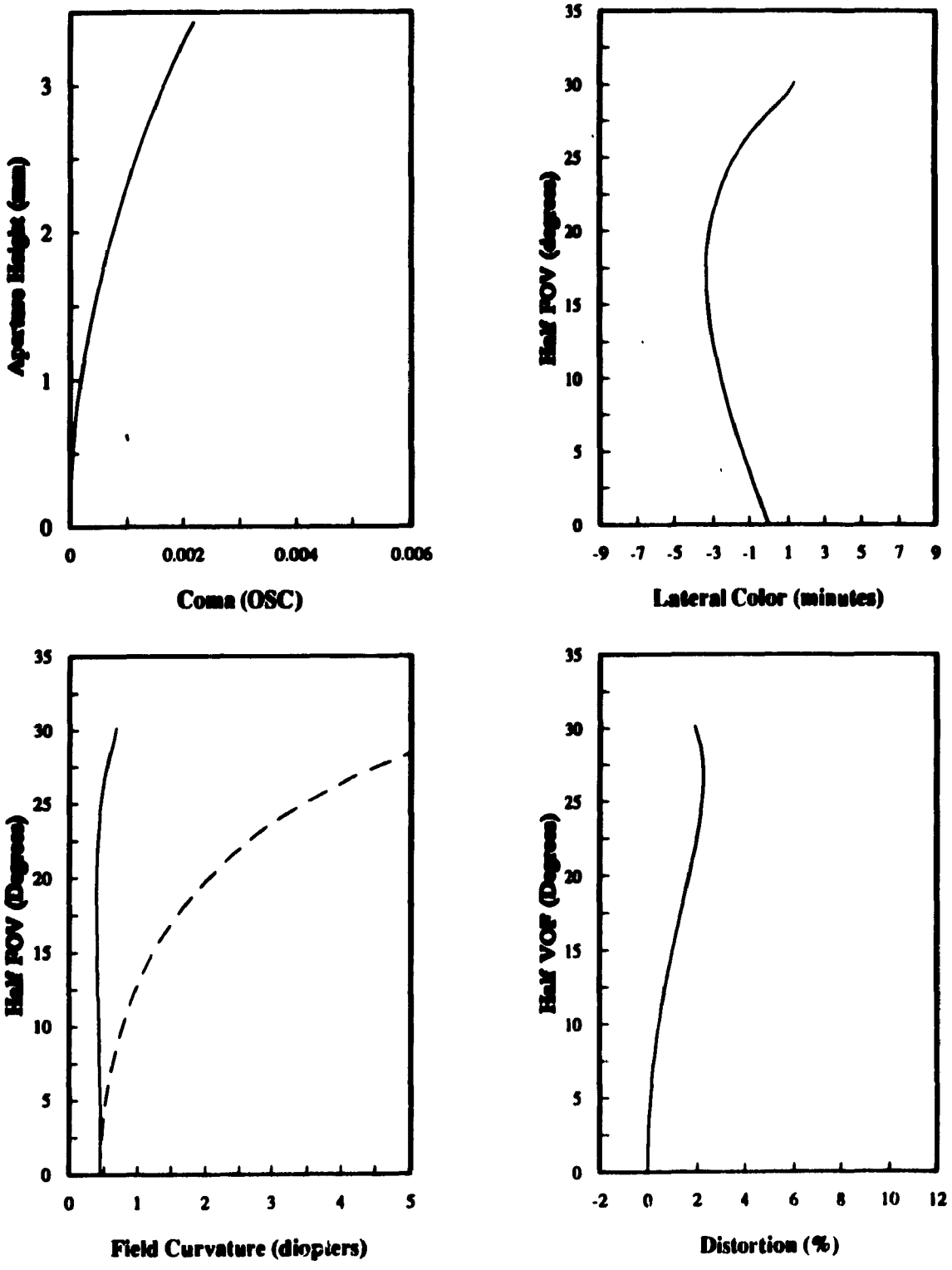


Figure 4.10: The aberration analyses of the optimized straight optical design.

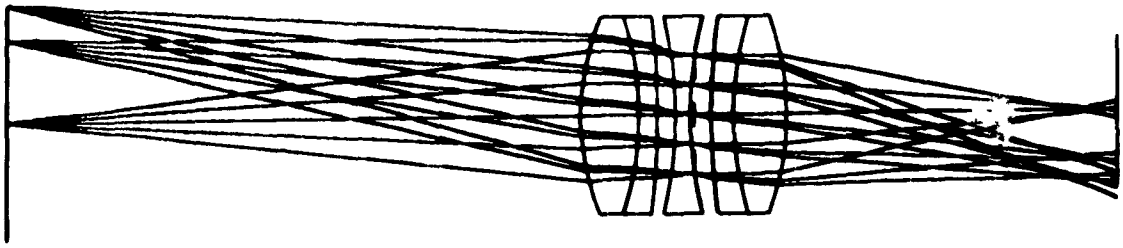


Figure 4.11: The layout of the starting objective of the folded optical relay.

Surface	Radius (mm)	Thickness (mm)	Medium	Index	Clap (mm)
obj	0.0000	1.00E+10	AIR	1.000000	
1	65.0000	15.0000	SCHOTT LAK31	1.696732	25.00
2	-80.0000	5.0000	SCHOTT SF1	1.717360	25.00
3	-200.000	5.0000	AIR	1.0000	25.00
4	-100.000	4.0000	SCHOTT BK7	1.516799	25.00
5	100.0000	1.0000	AIR	1.000000	25.00
6	0.000	4.00	AIR	1.000	15.00
7	200.000	5.000	SCHOTT SF1	1.71736	25.00
8	80.000	15.000	SCHOTT LAK31	1.696732	25.00
9	-65.000	44.0068	AIR	1.000000	25.00
10	0.000000	0.000	AIR	1.000000	

Table 4.6: Structural parameters of the objective of the BLHMD starting folded optical relay.

- to have a symmetrical structure, and
- to have more design freedoms. The compounding of doublets provides additional freedom which may be regarded as simply a means of artificially generating an unavailable glass type by combining two available glasses; alternatively, the refractive characteristics of the cemented interface may be utilized to control the course of the upper rim ray, which is affected strongly by these two surfaces. This reduces high order coma and off-axis spherical aberration.

The objective has a 66 mm effective focal length and a 44 mm back focal length. The structural parameters of the objective are shown in Table 4.6.

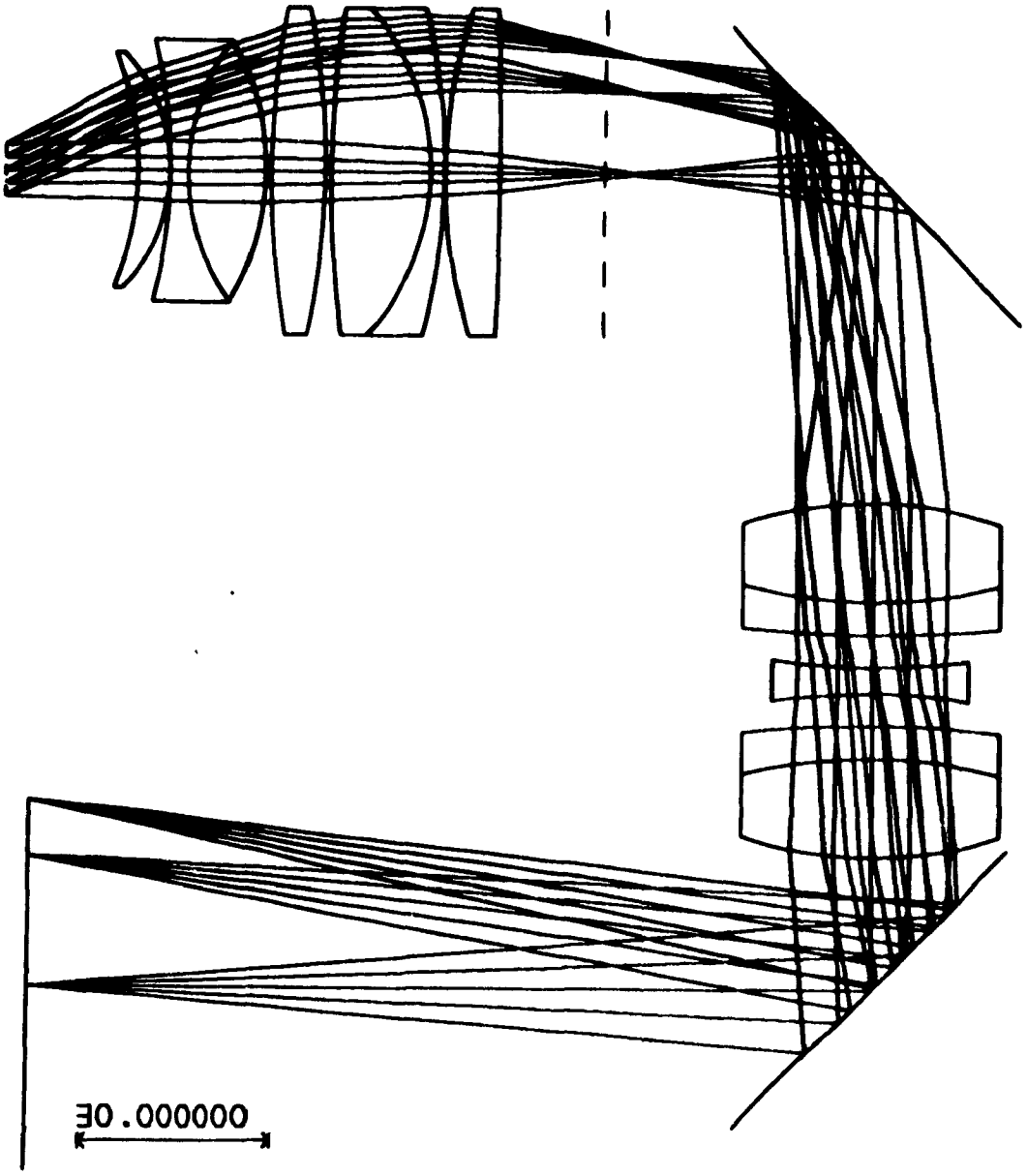


Figure 4.12: The starting system layout of the folded optical relay design.

We choose the optimized straight design as the eyepiece in this folded starting system. One field lens was added in front of the eyepiece in order to collect marginal rays. The system layout is illustrated in Figure 4.12.

4.4.2 Design Procedure

An objective and an eyepiece are all positive lenses, but they differ in that the entrance and exit pupils of an eyepiece are outside the eyepiece, and usually the objective has a longer focal length and an aperture whose position can be adjusted as a design freedom. Consisting of an eyepiece and an objective, the folded BLHMD optical relay has an exit

pupil outside the system (same with an eyepiece) and an intermediate image formed by the objective.

The system aberrations may be corrected by the objective and/or by the eyepiece. One class of aberrations have to be corrected by objective and eyepiece independently, such as the Petzval curvature and lateral color; the other aberrations can be compensated between the objective and the eyepiece, such as coma, astigmatism, and distortion. The Petzval sum of the objective and the eyepiece must be kept as low as possible.

The Objective

The first step is to optimize the objective by evaluation of the intermediate image. The same types of aberrations are included in the merit function, but more weight are put on the Petzval curvature and lateral color.

The optimized objective is inserted into the folded system with the well-corrected eyepiece.

The Whole System

The next step is to optimize the whole system. The method to calculate and form the merit function is the same as the straight design.

Since the system has more variables than the straight design, it took a longer time to find a minimum. The structural parameters of the best solution we found is summarized in Table 4.7.

4.4.3 Design Results and Analysis

This folded design was optimized for a full unvignetted 8 mm exit pupil, a 20 mm eyerelief, and a little larger than 60° IFOV. The lens diameters are all less than or equal to 25 cm. The layout of this system is illustrated in Figure 4.13.

The aberrations of the final folded optical relay are analyzed in Figure 4.14. The lateral color on the full field is less than 2.5 minutes, which is smaller than the straight optical relay. This is achieved by double corrections of the objective and the eyepiece. The coma, evaluated within a 4 mm diameter exit pupil, has the same value as that of the straight system. The distortion is a little worse and has a maximum of 4.7% at the margin field. The field curvatures are worse. A part of the tangential field is beyond infinity, which is hard to accommodate for most people. The astigmatism is smaller.

The overall characteristics of the system are not as good as expected. It is understandable that the folded system has a worse field curvature, because the objective and the

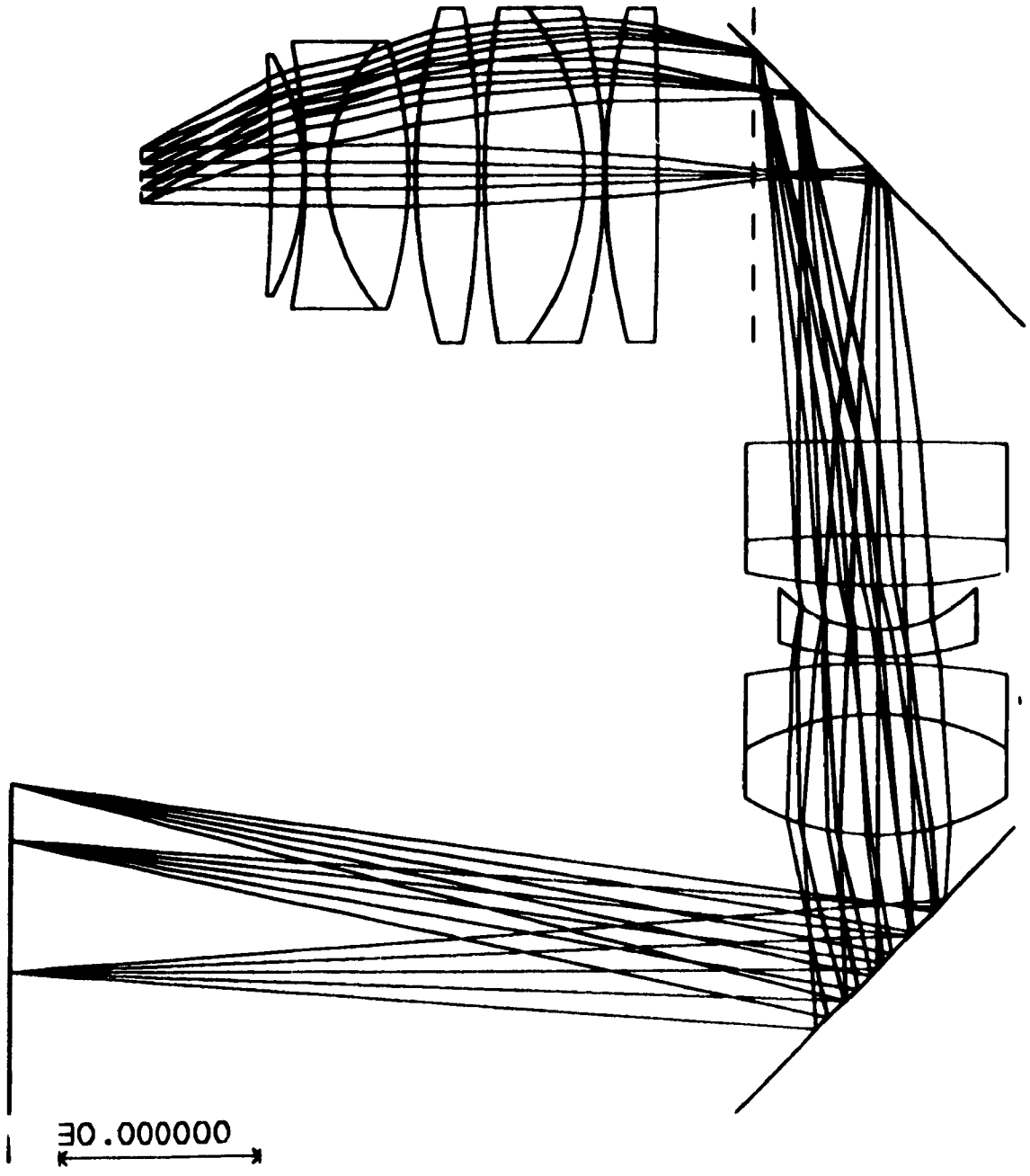


Figure 4.13: The layout of the optimized folded optical relay.

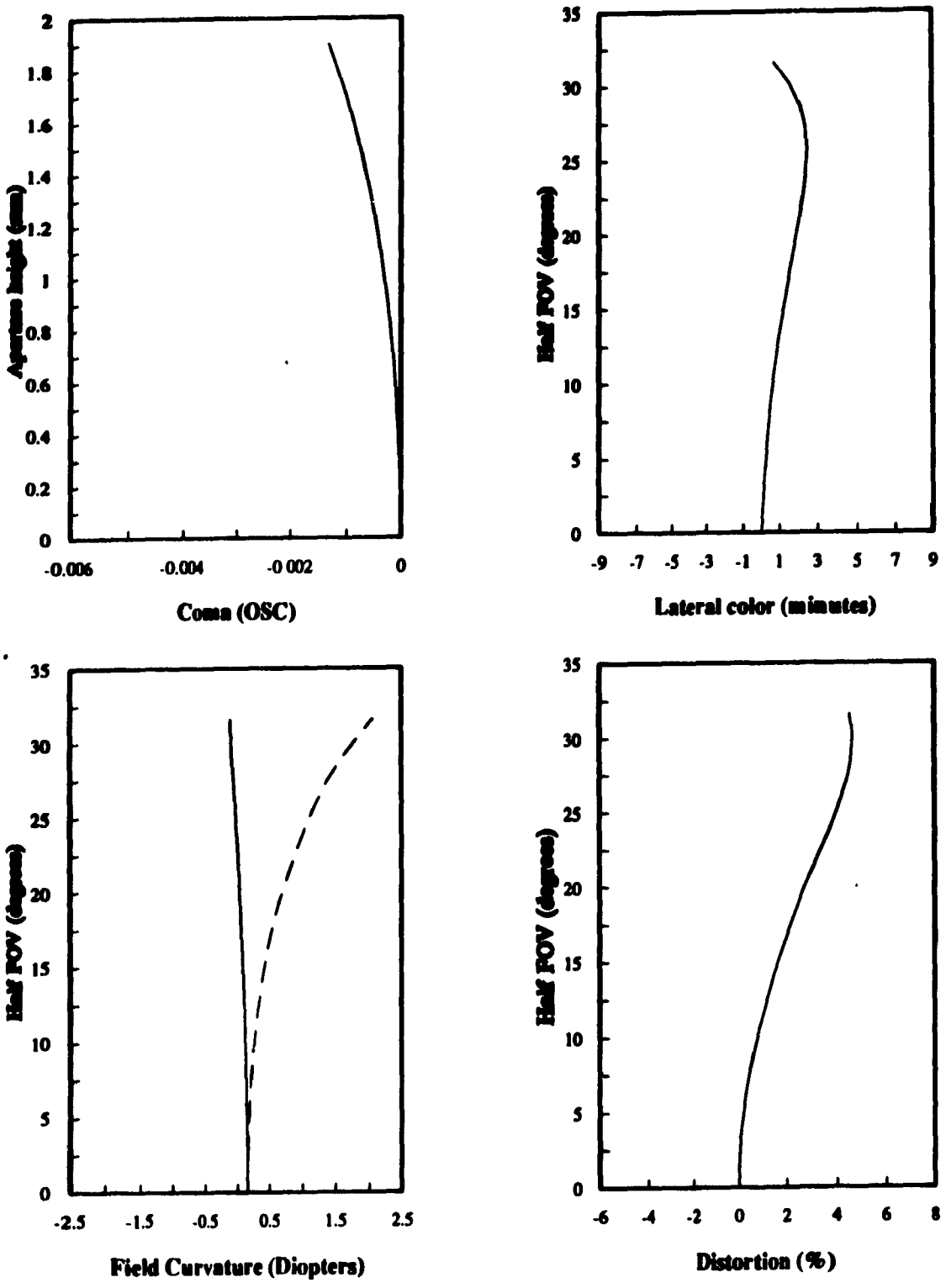


Figure 4.14: The aberration analyses of the optimized folded optical design.

eyepiece are positive lenses and contribute the same sign of field curvature to the system. The only thing we can do for this aberration is to keep the Petzval sum for both of them as low as possible. Coma and distortion were expected to be smaller, but it is not the case. The reasons might be:

1. the whole system is not symmetrical about the intermediate image, and
2. the system is too complex, and the best solution we found is just a local minimum.

The center of the folded optical relay and the LCDs can be adjusted by adjusting the positions of two mirrors. The orientations of the mirrors have to be kept the same. The size of the LCD screen can be larger. All those are achieved at the cost of more complex structure, more weight, and higher cost.

4.5 Summary

In Chapter 3, we derived the desired characteristics of the BLHMD optical relay. The present chapter has presented the structural parameters of two different optical relays to meet the desired characteristics. One optical relay has a straight form, and the other has a folded layout.

The biggest problem in either the straight or the folded optical relay design is the conflict among the requirement of the wide IFOV, the limit on the size of lenses, and the aberration tolerances of the human eye. The other conditions which complicate the design task are (1) a relatively large eye relief and (2) a large exit pupil.

The straight optical relay is a complete eyepiece consisting of 10 surfaces. The characteristics of this optical relay are satisfactory after its optimization. This optical relay is relative simple compared to the folded design. The major drawbacks are that the center of gravity of the relays and LCD screens is in front of the operator's head, and there is a limit on the size of LCDs due to the human operator's IPD.

The folded optical relay locates the LCDs on the two sides of the head, hence allows the use of larger size LCDs. The other advantage is that the center of gravity of the optical relay and LCDs can be adjusted. Although the effect of aberration balance between the objective and the eyepiece is not as obvious as we expected, the aberrations are all smaller than the design criteria except the field curvature. This is due to the fact that the objective and the eyepiece are all positive, and the field curvature introduced by the objective is enhanced by the eyepiece, in lieu of being compensated. The folded optical relay has more lenses, hence has more weight than the straight optical relay. If more time is taken on searching for an alternative folded structure, i.e., choosing different starting structures for the objective and

the eyepiece, and on the optimization, the folded form optical relay may hold the promise of better aberration correction.

Tolerances on the lens parameters of each optical relay, which are not discussed in this thesis, should be calculated before a prototype is made. Although practical tolerances may be expected to be significantly larger than the calculated tolerances [8], the calculated ones provide guidelines for setting practical tolerances.

Surface	Radius (mm)	Thickness (mm)	Medium	Index	Clap (mm)
obj	0.000000	120.000000	AIR	1.000000	
1	0.000000	0.000000	REFL	1.000000	30.000000
2	0.000000	-20.000000	AIR	1.000000	
3	-41.761055	-18.000000	SCHOTT LAK23	1.668815	25.000000
4	71.473247	-4.000000	SCHOTT SF1	1.717360	25.000000
5	352.219345	-0.500000	AIR	1.000000	25.000000
6	-132.056300	-8.000000	SCHOTT BK7	1.516799	22.000000
7	-30.316313	-6.895022	AIR	1.000000	22.000000
8	-50.415401	-7.061731	SCHOTT SF1	1.717360	25.000000
9	-690.751553	-14.000000	SCHOTT LAK31	1.696732	25.000000
10	-817.205693	-40.000000	AIR	1.000000	25.000000
11	0.000000	0.000000	REFL	1.000000	30.000000
12	0.000000	23.035298	AIR	1.000000	
13	0.000000	14.122293	AIR	1.000000	25.000000
14	94.137016	7.209360	SCHOTT SF8	1.688931	27.000000
15	-543.642373	0.100000	AIR	1.000000	27.000000
16	-217.455927	3.000000	SCHOTT F4	1.616591	27.000000
17	34.747895	15.000000	SCHOTT KF50	1.530881	24.000000
18	-62.106541	1.000000	AIR	1.000000	24.000000
19	79.377106	10.000000	SCHOTT BK7	1.516799	24.000000
20	-115.108893	1.000000	AIR	1.000000	24.000000
21	34.040843	12.000000	SCHOTT BK7	1.516799	21.000000
22	-45.208412	3.000000	SCHOTT F4	1.616591	21.000000
23	54.916834	0.100000	AIR	1.000000	20.000000
24	24.334637	5.060000	SCHOTT BK7	1.516799	16.000000
25	38.665779	20.000000	AIR	1.000000	16.000000
26	0.00000	-1159.527273	AIR	1.0000	4.000000
27	0.000000	0.000000	AIR	1.000000	60.000000

Table 4.7: Structural parameters of the optimized BLHMD folded optical relay

Chapter 5

Conclusions

5.1 Summary

The objectives of this thesis were:

1. to derive optical design criteria with the considerations of the human factors and optical constraints; and

2. to design an optical relay for use in the BLHMD system. The optical system has an IFOV greater than 60^o and well-corrected aberrations, and satisfies human factor constraints.

This thesis reviewed the alternate approaches to three HMD components: displays or image sources, optical relays, and combiners. The optical relay is the key component if a wide IFOV is required. We obtained two different relay lens systems, a straight form and a folded form. Each of them has an IFOV greater than 60^o. The design procedures and aberration analyses were also presented.

HMDs have been designed as a possible means of man-machine interface. To obtain the design criteria, we reviewed and discussed the major factors affecting the human visual acuity, the human visual requirements, optical limitations, and the particular application of the BLHMD. The optical design criteria we obtained include human factor constraints, aberration tolerances, and other constraints.

A brief review of first order optics, aberration theory, general design principles, and computer aided lens design was given.

5.1.1 BLHMD System

Based on the reviews of the alternate approaches to the HMD components and the discussions about the HMD's application conditions for telerobotics, we decided on an approach to the components in the BLHMD system. Our BLHMD will have the following features:

- Color LCDs,
- Stereo vision,
- Wide IFOV: larger than 60°, and
- Computer combiner.

5.1.2 Design Criteria and Design Characteristics

Table 5.1 summarizes the design criteria and the characteristics of the two optical relay, in comparison with the design targets. The straight optical relay consists of two doublets,

Parameters	Design Target	Straight Design	Folded Design
IFOV	$> 60°$	$62°$	$61°$
Lens diameter (mm)	< 55	< 54	< 54
Eyerelief (mm)	$> 18 \sim 20$	20	20
Exit pupil (mm)	10	10	8
Distortion (%)	$< 5 \sim 10$	< 2.5	< 4.5
Lateral color (min)	$< 5 \sim 10$	< 3.5	< 2.5
Coma (OSC)	< 0.0025	< 0.001	< 0.0015
TFC (diopter)	< 0.9	< 0.8	< 0.7, part of field is beyond ∞
SFC (dio.)	< 3	< 3 within $\pm 20°$ FOV	< 2.5
Astigmatism (dio.)	< 2	< 2 within $\pm 20°$ FOV	< 2
Weight	as light as possible	light	heavy
Shape of lens	spherical	spherical	spherical

Table 5.1: Summary of the design criteria and the characteristics of the two optical relays. TFC is the tangential field curvature. SFC is the sagittal field curvature.

one biconvex lens, and one meniscus lens. Its characteristics are satisfactory. But the sagittal field is quite curved in the field beyond $\pm 20°$. Hence the outer field is useful only for identifying the presence of a possible target. The folded system consists of an objective, an eyepiece, and two mirrors. Because of the complexity of the system, this system was optimized within a 8 mm exit pupil, which is smaller than the desired 10 mm.

All the aberrations are smaller than the design criteria except that a part of tangential field curvature is beyond infinity.

Compared to the straight optical relay, the folded system can adjust the center of gravity of the system and the LCD, allows the use of larger displays. But the folded relay is more complex, more heavy and more expensive to be made.

5.2 Limitations and Recommendations for Future Research

5.2.1 Structure of Optical Relays

For reasons of manufacture and cost, only spherical lenses were used in our optical design. In order to control aberrations under our desired levels, to have a wide IFOV, and to keep diameters of lenses smaller than 25 cm, both straight and folded optical relays employed complex structures: the straight relay has 10 surfaces and the folded relay has 22 surfaces.

To simplify the structures or to reduce the number of surfaces, there are three methods:

- use of aspherical lenses,
- use of gradient index lenses, and
- computer aberration compensations.

5.2.2 Aberrations

There are several alternate approaches to the aberration corrections.

Optical Fiber Field Flattening Lens

A wide IFOV and the constraint of total length of relay require a short focal length; a short focal length usually means a large Petzval sum. To flatten the field-of-view, over-corrected astigmatism has to be introduced. But too much of this will badly blur the outer image. Another way to achieve a flat field is by employing an optical fiber field flattening lens. One example is to insert this lens between the objective and the eyepiece of the folded relay, and let the two surfaces of the optical fiber lens have the same curvature with image planes of the objective and the eyepiece, which are all concave (Figure 5.1).

Symmetric Structure

The aberration analyses of the folded optical relay showed that the aberration correction was not as good as we expected. One important reason is that the whole system (including

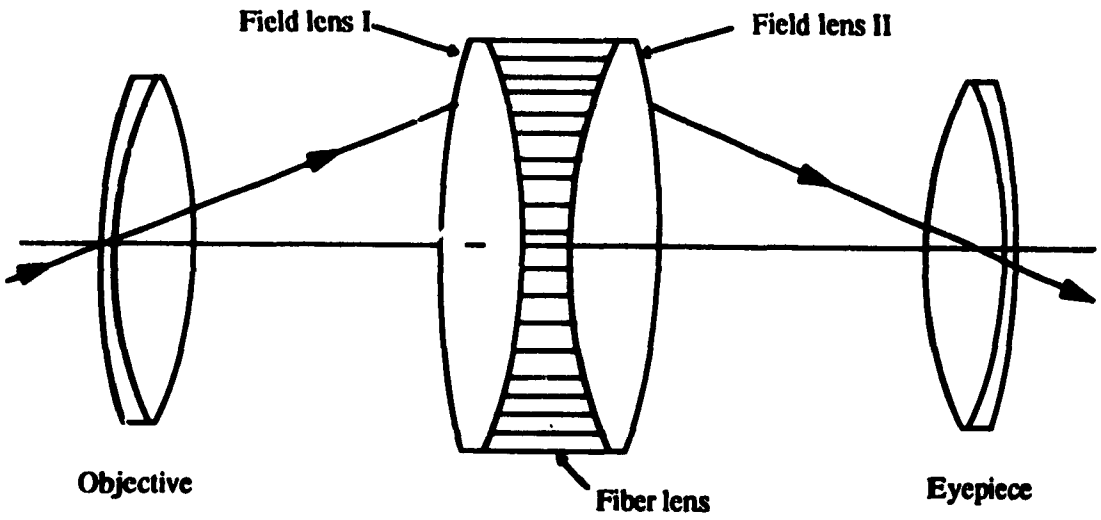


Figure 5.1: The optical fiber lens to flatten the field-of-view. From [32].

the objective and the eyepiece) is not symmetrical about the intermediate image plane. The starting objective was symmetrical about itself and the eyepiece was not. One possible way to have a better folded design is to find a symmetric, or approximately symmetric, starting system, and then find the best solution from this system by following the same optimization procedure.

Computer Aberration Compensation

For most optical designs, all aberrations in a system are corrected by adjusting structural parameters of the system. The other approach is using computers to introduce opposite aberrations to compensate for the system aberrations. It is easy to compensate when there is only one aberration in the system, such as the distortion [25]. In practice one is far more likely to encounter aberrations in combination than singly. It is a challenging research problem to find a computational mode to compensate for the mixed aberrations.

Bibliography

[1] Military standardization handbook: Optical design (mil-hdbk-141). (Army Contract No. DA-36-038-ORD-020690) Department of Defense, Washington DC, USA, 1962.

[2] K. E. Baker. Some variables influencing vernier acuity. *J. Optical Society of America*, 39(7):567–576, 1949.

[3] K. Boff and J. Lincoln. *Engineering Data Compendium: Human Perception and Performance*. AAMRL, Wright-Patterson AFB, Ohio, 1988.

[4] J. Bridenbaugh, W. Kama, and L. Task, II. The helmet-mounted hud: a change in design and applications approach for helmet-mounted displays. In *AGARD Conference Proceeding No 329*, pages 30-1 -30-5, 1982.

[5] J. Brown. The structure of the visual system. In C. Graham, editor, *Vision and Visual Perception*, pages 39–59. Wiley, New York,U.S.A., 1966.

[6] G. Burton and N. Haig. Effects of the Seidel aberrations on visual target discrimination. *J. Optical Society of America*, 1:373–385, 1984.

[7] F. Campbell and D. Green. Monocular versus binocular visual acuity. *Nature*, 208:191 192, 1965.

[8] A. Damon, H. Bleibtreu, O. Elliot, and G. E. Predicting somatotype from body measurements. *American Journal of Physical Anthropometry*, 20:461 473, 1962.

[9] G. De Vos and G. Brandt. Use of holographic optical elements in hmds. In R. Lewandowski, editor, *Helmet-Mounted Display (II)*, pages 2-8, Orlando, Florida, April 1990. SPIE- the International Society for Optical Engineering.

[10] J. G. Droessler and D. Rotier. Tilted cat helmet-mounted display. *Optical Engineering*, 29(8):849–854, 1990.

[11] H. Erfle. *U.S. Patent 1,478,704 filed on August 1921.*

[12] R. Farrell and J. Booth. *Design Handbook for Imagery Interpretation Equipment.* Boeing Aerospace Co., Seattle, WA, 1984.

[13] M. Griot. *Optics Guid 5*, chapter One. Melles Griot Inc., Irvine, CA, USA, 1991.

[14] J. Heard, D. Hayes, J. Ferrer, and A. Zilgalvis. Design of an airborne helmet-mounted display. Technical report, Huges Aircraft Co. under contract to US Aerospace Medical Research Laboratory, 1969. AMRL-TR-74-3.

[15] H. Hertzberg, G. Daniels, and E. Churchill. Anthropometry of flying personnel-1950. Technical report, Wright Air Development Center, 1954. WADC-TR-52-321.

[16] W. Kim, A. Liu, K. Matsunaga, and L. Stark. A helmet mounted display for telerobotics. *IEEE J. Robotics and Automation*, pages 543 – 547, 1988.

[17] R. Kingslake. *Lens Design Fundamentals.* Academic Press, New York, USA, 1978.

[18] R. Kingslake. *Lens Design Fundamentals*, chapter 15. Academic Press, New York, USA, 1978.

[19] M. Millodot, C. Johnson, A. Lamont, and H. Leibowitz. Effect of dioptrics on peripheral visual acuity. *Vision Research*, 15:1357–1362, 1975.

[20] Optikos. *ACCOS V Manual*, 1990.

[21] E. Pel. Visual issues in the use of a head-mounted monocular display. *Optical Engineer*, 29:883 892, 1990.

[22] R. A. Perez. *Electronic Display Devices.* TAB Professional and Reference Books, P.O. BOx 40, Blue Ridge Summit, PA, USA, 1988.

[23] P. Reeves. The rate of pupillary dilation and contraction. *Psychological Review*, 25:330–340, 1918.

[24] A. Reichert and B. Cohen. Prismatic combiner for head-up display. In H. Assenheim, editor, *Display System Optics(II)*, pages 197–202, Orlando, Florida, March 1989. SPIE - the International Society for Optical Engineering.

[25] W. Robinett and J. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence*, 1:45–62, 1992.

[26] P. Rogers. Visual optical systems. Material for Lens Design Summer Course, University of Rochester, June 1991.

[27] S. Rosin. *Eyepieces and Magnifiers*, volume III, chapter 9, pages 331 361. Academic Press, New York, U.S.A., 1965.

[28] S. Shlaer. The relation between visual acuity and illumination *Journal of General Physiology*, 21:165–188, 1937.

[29] W. Smith. *Modern Optical Engineering.* Optical and Electro-Optical Engineering Series. McGraw-Hill, Inc., New York, USA, second edition, 1990.

[30] E. Tayler. The inverting eyepiece and its evolution. *J. Sci. Instr.*, 22:43 48, 1945.

[31] M. Thomas, W. Siegmund, and S. E. Antos. Fiber optic development for use on the fiber optic helmet-mounted display. *Optical Engineering*, 29(8):855 862, 1990.

[32] Z. Wang. *Geometrical Optical and Optical Design.* Zhejiang University Press, Zhejiang, P.R. China, 1989.

[33] Z. Wang. *Geometrical Optical and Optical Design*, chapter 15. Zhejiang University Press, Zhejiang, P.R. China, 1989.

[34] M. Weinstock, W. Pishtey, J. LaRussa, and C. Tritsch. A holographic helmet mounted display application for the extravehicular mobility unit. In R. Lewandowski, editor, *Helmet-Mounted Display (II)*, pages 2–8, Orlando, Florida, April 1990. SPIE- the International Society for Optical Engineering.

[35] M. J. Wells and M. Venturino. Performance and head movements using a helmet-mounted display with different sized field-of-view. *Optical Engineering*, 29:870 877, 1990.

[36] G. Wilson and R. McFariane. The development of an aviators helmet mounted night vision goggle system. In R. Lewandowski, editor, *Helmet-Mounted Display (II)*, pages 128–139, Orlando, Florida, April 1990. SPIE- the International Society for Optical Engineering.