Meshfree Approximation Methods for Free-Form Surface Representation in Optical Design with Applications to Head-Worn Displays

Ozan Cakmakci[1], Gregory E. Fasshauer[2], Hassan Foroosh[3], Kevin P. Thompson[4] and Jannick P. Rolland[1-3]

[1]CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida

[2] Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois

[3] School of Computer Science, University of Central Florida, Orlando Florida

[4]Optical Research Associates, Pasadena, California

ABSTRACT

In this paper, we summarize our initial experiences in designing head-worn displays with free-form optical surfaces. Typical optical surfaces implemented in raytrace codes today are functions mapping two dimensional vectors to real numbers. The majority of optical designs to date have relied on conic sections and polynomials as the functions of choice. The choice of conic sections is justified since conic sections are stigmatic surfaces under certain imaging geometries. The choice of polynomials from the point of view of surface description can be challenged. The advantage of using polynomials is that the wavefront aberration function is typically expanded in polynomials. Therefore, a polynomial surface description may link a designer's understanding of wavefront aberrations and the surface description. The limitations of using multivariate polynomials are described by a theorem due to Mairhuber and Curtis from approximation theory. In our recent work, we proposed and applied radial basis functions to represent optical surfaces as an alternative to multivariate polynomials. We compare the polynomial descriptions to radial basis functions using the MTF criteria. The benefits of using radial basis functions for surface description are summarized in the context of specific magnifier systems, i.e., head-worn displays. They include, for example, the performance increase measured by the MTF, or the ability to increase the field of view or pupil size. Full-field displays are used for node placement within the field of view for the dual-element head-worn display.

Keywords: Optical system design, alternative surface representation, radial basis functions, head-worn displays

1. INTRODUCTION TO SURFACE REPRESENTATIONS IN OPTICAL DESIGN

The case of stigmatic imaging can be covered using conic sections for certain imaging geometries. A parabola will image an object at infinity to its focal point perfectly within the geometrical optics approximation. An ellipse will image perfectly, within a geometrical optics approximation, an object placed at one of its foci to the other focus. Two aspheric

Novel Optical Systems Design and Optimization XI, edited by R. John Koshel, G. Groot Gregory, James D. Moore Jr., David H. Krevor, Proc. of SPIE Vol. 7061, 70610D · © 2008 SPIE
CCC code: 0277-786X/08/$18 · doi: 10.1117/12.798351

surfaces can be designed for stigmatic imaging with the sine condition imposed by integrating the Wasserman-Wolf differential equations, which generate the aspheric surface profiles[1]. Wasserman-Wolf is a non-iterative technique generating a pair of differential equations for stigmatic imaging of a point object. Coma correction is obtained by satisfying the sine condition. Essentially, by rewriting a vectorial version of Snell's law at each aspheric interface, Wasserman-Wolf generates a slope field representing the tangents on each point of each aspheric surface, the slope field is integrated using a Runge-Kutta or Adams method to get the surface profile. The original paper deals with the case of rotationally symmetric aspheres. David Knapp's thesis work generalized the Wasserman-Wolf technique removing the axial symmetry assumption in the original work[2].

In the case of imaging extended objects, the exact functional form of the optical surfaces are usually unknown a priori. Therefore, often times an approximation to a function representing the ideal surface is used, where the approximation is represented as a linear combination of basis functions

$$s(x,y) = \sum_n \phi_n(x,y) w_n \tag{1}$$

where the ϕ_n represent the basis functions and the w_n represent the weights. We use the term approximation to mean that we are given a set of data sites $\{\boldsymbol{x}_1,\boldsymbol{x}_2,..\boldsymbol{x}_n\}$ and a set of scalar data values $\{z_1,z_2,..,z_n\}$, we are looking for the function s that is approximately equal to the data values at each data site, i.e., $s(\boldsymbol{x}_i)\approx z_i$.

We will now briefly discuss the three standard choices for the basis functions (ϕ_n), namely polynomials, trigonometric functions and rational functions.

1. 1 POLYNOMIALS AND PIECEWISE POLYNOMIALS AS THE CHOICE OF BASIS

The standard way to describe aspheres in imaging optics has been to add polynomials to a base conic as

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

where c represents the curvature, r is $\sqrt{x^2+y^2}$, and k is the conic constant.

It is known that in 1D there will be a unique polynomial of degree N-1 that will interpolate N distinct points. Interpolation using polynomials with equally spaced points suffers from the Runge phenomenon. The Runge phenomenon can be observed by considering a set of equally spaced points on the interval [-1,1] and fitting a polynomial to the Runge function $\frac{1}{1+25x^2}$. Increasing the polynomial order does not yield better approximations with a uniform spacing of the interpolation points. In fact, this introduces huge oscillations near the end points of the interval. Changing the point distribution from uniformly spaced to Chebyshev points (i.e., points equally spaced on a circle and projected to an axis) will improve the fit to the Runge function. Increasing the polynomial order, with a Chebyshev point distribution, helps achieving better fits in the case of the Runge function. The difficulty of choosing a basis in two or higher dimensions is stated by the Mairhuber-Curtis fact, which says that a multivariate basis, polynomial or otherwise, with a finite number of basis functions and a set of distinct interpolation points, may not lead to an invertible interpolation matrix, if the basis functions are independent of the data. The formal definition of the Mairhuber-Curtis fact is given by Fasshauer[3].

An alternative approach to using high order polynomials with a Chebyshev point distribution is to break up the domain into pieces. Low order polynomials, for example cubics, can be used in each sub-interval to achieve better approximations to the target function, and this approach is known as the spline approach. There are several kinds of splines (Bezier splines, B-splines, non-uniform rational B-splines (NURBS), for example). Stacy used cubic splines

within the ACCOS-V code as general surfaces in the design of unobscured reflective optical systems[4]. The basic aim in Stacy's work was to compare spline based designs to the classically designed Galileo narrow angle camera. Stacy found that the cubic spline designs allowed wider field angles and higher transmission than the Galileo with lower but still acceptable image quality. Chase did some initial work on the application of non-rational Bezier curves to the description of rotationally symmetric asphere design. Chase compared the spherical aberration correction of a non-rational Bezier curve against an even and odd polynomial in a Cassegrain system[5]. Davenport investigated third degree non-rational Bezier splines (section 3.3 and appendix A of his thesis detail the interpolation process) as a tool for creating an incoherent uniform circular illuminance distribution on a target plane[6].

Greg Forbes recently proposed a sum of Jacobi polynomials to represent axisymmetrical aspheres[7]. Forbes emphasizes the use of Jacobi polynomials, which are global and orthogonal. Forbes' representation has the property that the mean square slope of the normal departure from a best-fit sphere is related to the sum of squares of individual coefficients of Jacobi polynomials. This property facilitates the enforcement of fabrication constraints.
Scott Lerner introduced a novel explicit superconic surface[8]. Lerner also explored parametrically defined optical surfaces and implicitly defined optical surfaces. A truncated parametric Taylor surface and an *xyz*-polynomial were shown to be general surface descriptions. Representations were compared for ray tracing speed, optimization complexity, the ability to correct highly aspheric wavefronts, and the ability to represent steeply sloped surfaces. Specifically, Table 5.2 in Lerner[8] lists a superconic, explicit superconic, truncated parametric Taylor, and an *xyz* polynomial as the surfaces that were compared.

1. 2 RATIONAL FUNCTIONS AS THE CHOICE OF BASIS

The most flexible splines are the non-uniform rational B-splines (NURBS). A *p*-th degree NURBS curve is fully defined using control points, knots, and weights, see Piegl[9] for the mathematical formula. The NURBS curve is essentially a rational basis. The NURBS curve can be extended to two dimensions using the ideas of a tensor-product or triangulation over a domain. The set of control points in two dimensions is called a control mesh. Generation of these meshes is a cumbersome task in three dimensions and higher.

1. 3 TRIGONOMETRIC FUNCTIONS AS THE CHOICE OF BASIS

Michael Rodgers considered alternative aspheric representations to the standard method of adding a power series to a base conic[10]. Rodgers' thesis discussed nonpolynomial basis functions added to a conic in order to yield perfect axial imagery. Examples of nonstandard aspheric functions are given in Table 2.1 of Rodgers' thesis to be hyperbolic cosine, logarithm, secant, inverse sine, tangent and a Gaussian. Convergence properties of such alternative representations were studied. Rodgers' dissertation considered rotationally symmetric systems only. Rodgers' thesis studied global and compact representations for optical surfaces. Rodgers' thesis includes noniterative as well as iterative approaches to generating aspheric profiles. The noniterative surface generation technique was based on Wasserman-Wolf. Imaging and energy redistribution examples of two, three, and four mirror reflective designs were given. These examples illustrated that basis functions like the hyperbolic cosine, logarithm, and the secant can represent a surface shape significantly better than the same number of power series terms. Lerner studied the possibility of a Fourier series based on cosine functions for rotationally symmetric surfaces[8]. Lerner found that a Fourier series is not suitable as an optical surface and this was illustrated in page 39 of his thesis with a plano-convex lens having a single field. Lerner shows designs with 10, 20, 60, 100, 200 terms in the Fourier series and comments that the residual aberration is too large to be acceptable.

2. RADIAL BASIS FUNCTION APPROXIMATION

Our emphasis in this paper is on a local representation of shape that is suitable for the optimization of rotationally non-symmetric free-form systems as well as rotationally symmetric systems. We are using the term free-form in reference to surfaces that are rotationally non-symmetric. Advances in fabrication technology (e.g., diamond turning) are making the design and use of free-form optical surfaces feasible in imaging and non-imaging optics. Benefits of free-form surfaces have been summarized by Rodgers and Thompson as ``greater control of the location of nodes in the aberration field, and potentially the larger number of nodes in the field"[11]. The potential of radial basis functions as a surface representation has been recognized by A.K. Chan et al.[12], however such representation was not proposed or investigated in solving lens

design optimization problems. Greynolds used a linear combination of 1-dimensional Gaussians to represent and propagate wavefronts[13].

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

$$s(\boldsymbol{x}) = \sum_n \phi\left(\left\|\boldsymbol{x} - \mathbf{c}_n\right\|\right)\mathbf{w}_n \quad . \tag{3}$$

where $\boldsymbol{x}$ is an evaluation point, $\mathbf{c}_n$ are the centers of the basis functions, and $\mathbf{w}_n$ are the weight vectors. The translated basis functions in equation (3) take the data into account by shifting the basis functions to the data sites (and thus there are no problems with the implications of the Mairhuber-Curtis theorem). This choice of basis does not require a mesh to generate a surface. The basis centers are placed within a domain $\Omega \subset R^s, s \geq 1$, and a discrete surface is approximated at the points of interest (i.e., evaluation points). The placement of the basis centers can be uniform, according to Chebyshev points, or even completely random. In our numerical experiments $\|.\|$ denotes the 2-norm. The impact of different *p*-norms in our application, specifically, the *p=1* and infinity norms, will be studied in future work. Composing the basis functions with the 2-norm makes these basis functions spherically radial.

Using the approximation condition $s(\boldsymbol{x}_i) \approx z_i$, where $i=1,2,\ldots,m$, at m data points and using n basis functions for $s(\boldsymbol{x})$ in equation (3), we can express the approximation at each point $\boldsymbol{x}_i$ as a dot product. It is possible to express the collection of m products with a matrix-vector multiplication and arrive at the following m by n linear system

$$\mathbf{\Phi}_{mxn}\mathbf{w} = \mathbf{Z} \tag{4}$$

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

$$\mathbf{\Phi} = \begin{bmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_n(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_m) & \phi_1(\mathbf{x}_m) & \cdots & \phi_n(\mathbf{x}_m) \end{bmatrix} \quad . \tag{5}$$

Therefore, the question now becomes the conditions on the basic function ϕ, translates of which generate the $\mathbf{\Phi}$ matrix. A popular point of view is to assume that $\mathbf{\Phi}$ is positive definite (and $m=n$). In this case, the connection between the choice of ϕ and a desired positive definite interpolation matrix $\mathbf{\Phi}$ is provided by the Bochner theorem[3]. The most important practical implication of the Bochner theorem is that the translated versions of a positive definite function ϕ indeed generate a positive definite $\mathbf{\Phi}$ matrix. A function is positive definite if it has a nonvanishing Fourier transform. A matrix A is positive definite if $u^T\mathbf{A}u \geq 0$ for all $u \in R^s$. Positive definite matrices have positive eigenvalues, and therefore positive definite matrices are invertible. The theory developed by Micchelli, using the results of Bochner, covers the case where the basis centers and the data points coincide (i.e, symmetric and positive definite interpolation matrix $\mathbf{\Phi}$). There is not much theory available for the case when the data sites and the centers do not coincide[3]. The designs discussed in Section 3.4 and Section 3.5 did not require a user-defined error function which would trace rays only at the basis centers during optimization, which means that the basis centers and the data points in those experiments are not coincident.

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

Given a particular surface Z, choice of a radial basis function, choice of basis centers and data points, and a shape factor in the case of a Gaussian or a multiquadric, the weights can be found using least squares by

$$w = \left(\Phi^T \Phi\right)^{-1} \Phi^T Z \tag{6}$$

Figure 1 illustrates the approximation of a sphere using a linear combination of Gaussian radial basis functions.

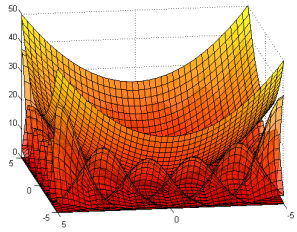


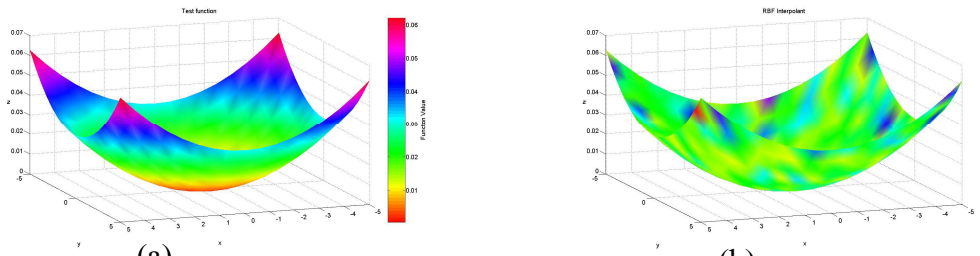
Fig. 1. Illustration of a linear combination of Gaussian basis functions approximating a sphere. Original sphere and the weighted basis functions are shown. The 2D Gaussians are spaced uniformly with means centered on a 6x6 grid. Each 2D Gaussian has a variance of unity along the x and y dimensions. The weighted Gaussian basis functions shown underneath the sphere are found through least squares.

We would like to note that except for the thin-plate spline, the Gaussian and the multiquadric functions contain a shape factor ε. In the case of the Gaussian, the shape factor is related to the variance as $\varepsilon^2 = \frac{1}{2\sigma^2}$. The impact of the choice of ε on the accuracy of function approximation, in the case of coincident basis centers and data points, is visualized in Fig. 2. Figure 2 (a) is a plot of a parabolic function. A 6x6 uniformly distributed Gaussian RBF approximation of a paraboloid with ε=0.1 is shown in Fig. 2(b). The RMS error is on the order of 10^{-9}. Therefore, a plot of it is not shown. A 6x6 uniformly distributed Gaussian RBF approximation with ε=1 is shown in Fig. 2(c). The RMS error is on the order of 10^{-3} and is plotted in Fig. 2(d). A 6x6 uniformly distributed Gaussian RBF approximation with ε=10 is shown in Fig. 2(e). The RMS error is on the order of 10^{-2} and is plotted in Fig. 2(f). We observe that when the basis functions become "too peaky", the quality of the approximation is reduced.

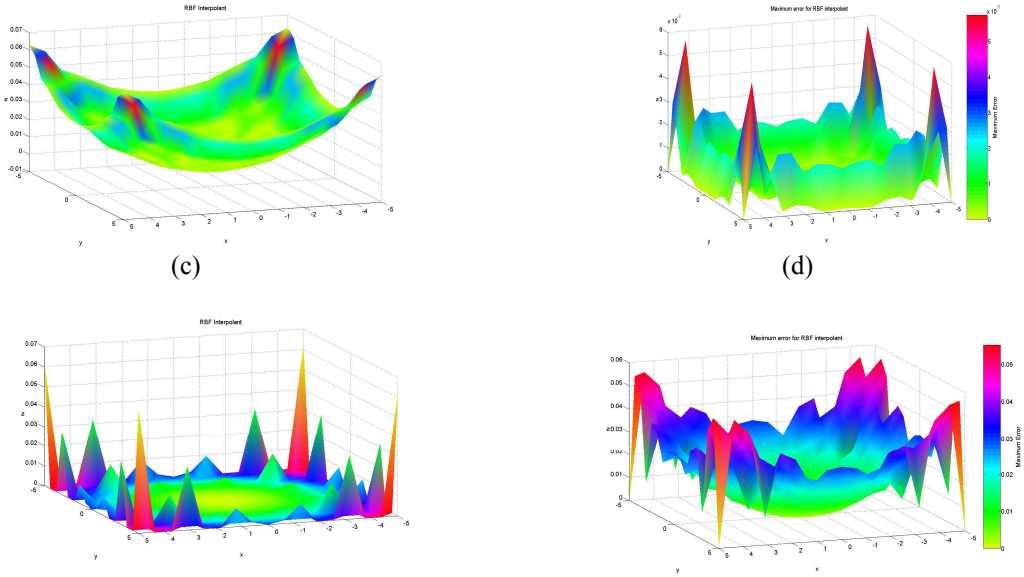
A few natural model choices that come up in the application of radial basis functions to optical surfaces include the choice of the basic function, the number of basis functions to use in a particular problem, the spatial distribution of the basis centers, and the choice of the shape factor. To our knowledge, there is no theoretical guidance on choosing the number of basis functions for a particular optical design problem. The minimum number of basis functions that provides an adequate approximation is found through experiments and is application specific. In terms of the distribution of points, Chebyshev distribution of points is likely to be more desirable over a uniform distribution of points.

Table 1. Definition of the Gaussian, thin-plate spline, and the multiquadric radial basis functions

Gaussian	Thin-plate spline	Multiquadric
$\phi(x) = e^{-\varepsilon^2 \|\|x\|\|_2^2}, x \in R^s$	$\phi(x) = \|\| x \|\|^{2\beta} \log(\|\| x \|\|), x \in R^s, \beta \in N$	$\phi(x) = (1+\varepsilon^2 \|\| x \|\|^2)^{-\beta}$ $x \in R^s, \beta \in R \setminus N_0$



(a) (b)



(c) (d)

(e) (f)

Fig. 2 Impact of the shape factor on the accuracy of the approximation (a) Plot of the function (paraboloid) being approximated (b) Plot of the 6x6 uniformly distributed Gaussian RBF approximation with ε=0.1 (c) Plot of the 6x6 uniformly distributed Gaussian RBF approximation with ε=1 (d) Plot of the error $|f_{approx.} - f|$ (e) Plot of the 6x6 uniformly distributed Gaussian RBF approximation with ε=10 (f) Plot of the error $|f_{approx.} - f|$

3. OPTICAL SYSTEM DESIGN EXAMPLES: HEAD-WORN DISPLAYS

A variety of visual displays, both on-body and in the environment, are being developed to support mobile augmented reality (AR) applications. A review of displays in the environment is provided in [12]. Despite several challenges, head-worn displays (HWD) are becoming a viable visual display option for mobile augmented reality[14]. There is a general demand expressed by several AR users, developers and researchers that a mobile AR display should look like eyeglasses as much as possible. A top-down requirement such as the eyeglass form-factor restricts the volume of the optics, electronics and the mechanics assembly to the order of 20 to 30 cm^3 per eye. This volume is estimated from a laser scanned CAD model of a human head[15] and close fitting sunglasses. The whole display assembly can be broken down into sub-assemblies and certain sub-assemblies can be placed elsewhere on the human body. As an example, fiber bundles have been used to decouple the drive electronics and the image sources from the optical sub-assembly coupling to the human visual system. However, the optical sub-assembly coupling to the human eye must reside approximately within this volume. Such a volume constrains the geometry and the number of elements of the optics.

In general, there are two design forms that can be pursued for the optical design of eyeglass displays: pupil forming and non-pupil forming designs. Pupil forming designs include eyepieces that magnify an intermediary image while imaging the pupil of the optical system forming the exit pupil. Other examples of pupil forming designs include projection optics, and retinal scanning displays that can optically be considered as scanning projection systems. Non-pupil forming designs include magnifiers. Once a microdisplay suitable for the task is chosen, the role of the optical system is to relay a magnified virtual image of the microdisplay to the human visual system. The gain in apparent size is typically quantified by comparing the image seen through the optical system to the image seen with the unaided or naked eye. For the unaided human eye, the apparent size of a microdisplay can be increased by bringing the microdisplay closer to the eye up to the accommodation limit. Compact and lightweight optical systems are necessary in many applications, especially when the users are mobile. Compactness of the microdisplay necessitates a magnification requirement since the microdisplay typically is too small to view with the unaided eye. In HWDs, the virtual image is often placed within about 1m for near-field tasks such as medical, maintenance and repair, or beyond 1m for far-field tasks[16,17]. The pupil of

the observer becomes the limiting aperture and is the exit pupil in a magnifier. Magnifiers are typically designed to accommodate a range of eye movements such as translations and rotations while observing the virtual image formed by the magnifier.

In section 3.1, we will briefly summarize related work in eyeglass displays. In section 3.2, we will summarize our dual-element system using a free-form optical mirror and lens with a diffractive optical element. In section 3.3, we will summarize our single mirror solution. Both the single and dual-element systems presented in sections 3.2 and 3.3 have been fabricated and both systems have relied on *x-y* polynomials to describe the free-form optical mirrors. In section 3.4, we provide one approach to addressing the question of "what is the optimal mirror shape in a magnifier with 15mm eye clearance, a 3mm pupil and a 24 degree full field-of-view?".

3. 1 BRIEF SURVEY OF RELATED WORK

Upton, in the mid 60's and 70's, integrated display systems within eyeglasses for applications in speech interpretation assistance. Initial prototypes were based on energization of small lights or lamps mounted directly on the surface of an eyeglass lens[18]. A later prototype from around the early 70's embodying small reflecting mirrors on the lens of the eyeglasses and moving the direct mounting of the light sources away from the lens resulted in being less noticeable and less obstructive to the wearer's vision[19]. In the late 80's, Bettinger developed a spectacle mounted display in which the spherical reflective surface of a partially transparent eyeglass lens was employed[20]. There was work in the late 90's embedding the mirrors into the eyeglasses lens by Spitzer and colleagues[21]. Spitzer et al. decided that based on the ~20x practical magnification of a single lens and their image goals of 28x21cm at 60cm, they would need a 0.7" display, which they concluded would be too large for concealment in eyeglasses. Therefore, they concentrated on a relay system built into the eyeglasses' frame to move the microdisplay away from the eyeglasses in their initial prototype. They have demonstrated a system with an overall thickness of the eyeglasses lens of less than 6.5mm, which fits in the commercial eyeglass frame. Holographic optical elements (HOE) were applied to the eyeglass-based display problem. An HWD with a 3mm exit pupil, 27x10 degrees of field of view, operating at the single wavelength of 532nm was designed based on a HOE and fabricated[22]. Sony recently proposed a 3mm thick lightpipe with two in and out coupling hologram elements as an approach to achieving a full-color display with a 20 degree field of view (pupil size was not reported) by using volume holograms[23]. Color crosstalk is eliminated by separating the red and blue hologram layers from the green hologram layer. Color uniformity is improved by tilting the microdisplay and using a color correction transform in the drive electronics. A second alternative in the design of compact optical magnifiers is to use a laser source and to temporally scan the image onto the retina. However, as can be explained by the Lagrange invariant, the exit pupil of such a scanner is small and pupil expansion mechanisms are required. Pupil expanders add additional complexity and size to the whole system, which deviate from the goal of compact magnifiers with high image quality.

3. 2 DUAL-ELEMENT OFF-AXIS HEAD-WORN DISPLAY

The optical layout is shown in Fig 3(a) and consists of an 8 mm exit pupil, a free-form mirror, a refractive and diffractive lens (referred to as a hybrid lens in Fig. 3 (a)), a flat fold mirror and a transmissive microdisplay. The system shown in Fig. 3 (a) is designed with a 20° diagonal field of view and a 15mm eye clearance. The line connecting the vertex and the center of curvature of the freeform mirror is tilted at 34 degrees in the *x-z* plane (e.g., "top view") with respect to the optical axis of the human eye. We found this angle to be the minimum tilt angle that can provide the necessary clearances around a human head based on a database of publicly available CAD models of human heads. Typically, in an HWD, the pupil of the human eye is the aperture stop of the system. Therefore, the aperture stop and the exit pupil coincide in this system. A user would place their pupil at the exit pupil to view the computer generated images. Even though magnifiers do not form exit pupils, it is customary to optimize these systems across a finite pupil size that is larger than that of the human eye to accommodate natural eye movements. Since under photopic illumination the pupil of the human eye is about 3mm, it is customary to conduct the analysis for a 3 mm eye pupil for both centered and decentered pupils.

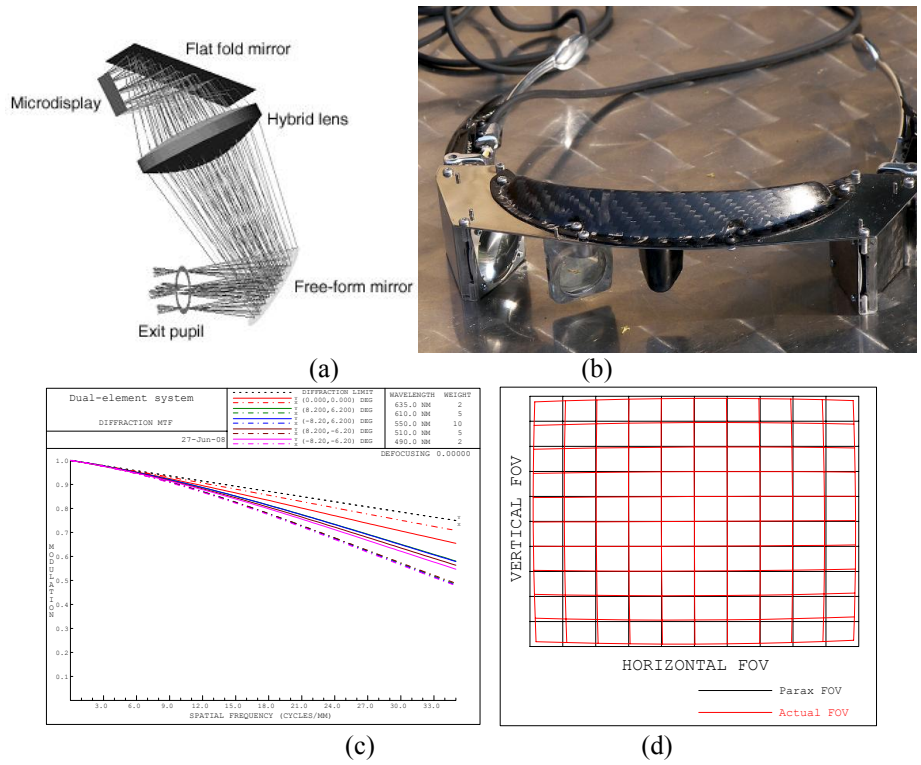


Fig. 3. Dual element magnifier: (a) Optical layout; (b) Gen-II dual-element eyeglass display; (c) Modulation transfer function plotted out to 35 cycles/mm (evaluated with a 3mm pupil); (d) Distortion grid.

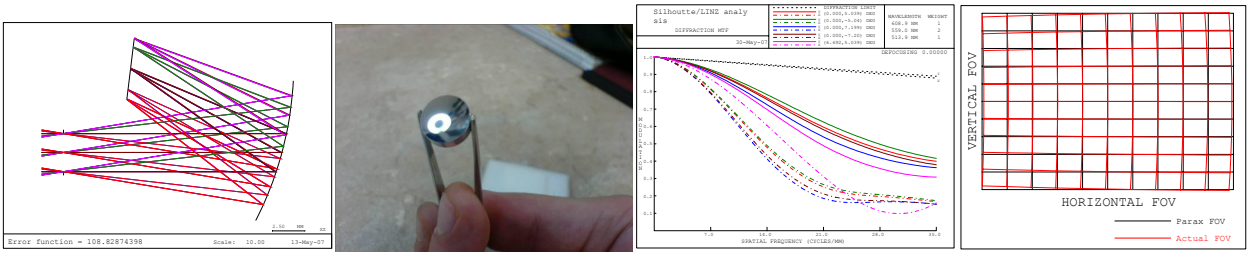
The polychromatic MTF evaluated for a centered 3mm pupil is plotted in Fig.3(c) for the on-axis field and the performance limiting fields. The performance of the MTF across decentered pupils up to 6 mm in various directions displayed similar performance to the centered pupil MTF. A red, blue, and green microdisplay pixel triplet is 14.1 µm in size, which yields a Nyquist frequency of ~35 cycles/mm. Thus the polychromatic MTF was plotted up to 35 cycles/mm. In reading the MTF plots, one desires all curves (each one corresponding to a selected design field point), to the extent possible, to be overlapped and provide ~20% contrast (i.e., modulation transfer function value) at the cut-off spatial frequency set by the microdisplay. Cutting off the spatial frequency of the microdisplay prevents pixelization effects that are visually resolvable. In the current prototype, the microdisplay has a 0.44" diagonal and contains 640x480 pixels. In visual space, the display provides 1.5 arcminutes resolution as limited by the pixel spacing on the microdisplay. It may be noted that a calculation of Nyquist frequency, given a 2.5µm cone spacing of the human visual system, yields a 1 arcminute resolution limit for the human visual system. The magnitude of maximum distortion occurs at (x=8.2°, y=-6.2°) in the field and was measured in simulation to be -3.8%. Fig.3(d) shows a distortion plot with the real rays compared against the paraxial rays. The Gen-II system is shown in Fig.3(b). The weight of the Gen-II system is 124 grams. The Gen-I system was described in our earlier work[24].

3. 3 SINGLE-ELEMENT OFF-AXIS HEAD-WORN DISPLAY

An ideal solution for an HWD would be a single surface mirror design. A single surface mirror does not have dispersion. Therefore, color correction is not required. A single surface mirror can be made see-through by machining the appropriate surface shape on the opposite side to form a zero power shell. From a system perspective, the system weight is reduced to a minimum.

We found a solution to the single mirror magnifier problem using an x-y polynomial surface. The layout of the single mirror magnifier is shown in Fig. 4(a). The system shown in Fig. 4(a) has a 24° degree diagonal full field-of-view with an eye clearance of about 15mm and a 3mm exit pupil. Increasing the exit pupil size is a challenge in this configuration. The system is full-color operating across the visible regime (450-650nm). The mirror has a tilt angle of ~15° in the x-z plane with respect to the optical axis of the human eye. The fabricated mirror prototype is shown in Fig. 4(b) and it is currently being integrated into a pair of sunglasses. The modulation transfer function is plotted out to 35 cycles/mm and

shown in Fig.4(c). The Nyquist frequency of the modulation transfer function plot is based on a 15μm pixel spacing. This system has fewer degrees of freedom compared to the system in Fig. 4(a) which explains the different modulation transfer function curves. The distortion grid is shown in Fig. 4(d) and the system has a maximum distortion of about 2.5%.



(a) (b) (c) (d)

Fig. 4. Single element magnifier: (a) Optical layout; (b) Fabricated mirror prototype with a 13mm diameter; (c) Modulation transfer function plotted out to 35 cycles/mm (evaluated with a 3mm pupil); (d) Distortion grid.

3. 4 APPLICATION OF RADIAL BASIS FUNCTIONS TO FREE-FORM SURFACE DESCRIPTION

The single element design and prototype shown in Fig. 4 serves as an ideal testbed for the following question: what is the optimal surface shape in the case of a single element magnifier for our specific problem? Physically, the optical design problem in the case of a single element magnifier reduces to adjusting surface normals of the mirror at each point along the mirror surface. Therefore a notion of optimality has to enter the picture. Each point along the mirror surface receives rays from multiple fields, which means that a single position for the surface normal will not exactly satisfy the demands for each field. Therefore, the problem turns into an optimization problem where we minimize the error for each field and balance the demands from each individual field. The case of a single field can be solved by integrating differential equations of Wasserman-Wolf. Mathematically, we think of an optical mirror surface as a function. We choose to approximate the function describing the optical mirror using radial basis functions.

Our implementation and initial experimentation have been with the Gaussian radial basis function listed in Table 1. Gaussians possess several desirable properties from the point of view of optical design. First, Gaussians are smooth functions (C^{∞}) having derivatives of all orders providing a desirable property given that optical surfaces are smooth. In addition, smoothness is desirable from a fabrication point of view. Second, theoretically, Gaussians are not local functions. However, practically they can be considered local since the value of a Gaussian outside of 3σ is small (<0.011, for a unit amplitude Gaussian). Third, the Fourier transform of a Gaussian is a Gaussian, which provides an analytical description for the power spectral density (PSD) of the surface.

We will next describe the optimization procedure. The first step is to set the grid size. Optimal setting of the grid size for a particular problem requires further study and has not been considered yet. For the magnifier example in Fig. 5, we have found a 17x17 grid to perform well when compared against relatively general surface representations such as *x*-*y* polynomials or Zernike polynomials. The second step was to initialize the starting point. Our raytrace software requires a base sphere in order to perform paraxial image calculations. Therefore, we add a base conic to equation (1).

Given a number of basis functions, we divided the aperture into x_{num} pieces in the *x*-dimension and y_{num} pieces in the *y*-dimension. The number of columns in the Φ matrix was set by the product of x_{num} and y_{num}. The number of rows in the Φ matrix is set by the number of evaluations points. We make a rectangular aperture assumption in this case. However, the representation of the surface as a linear combination of basis functions accommodates any aperture shape since the Gaussians can be moved spatially using their means, and placed anywhere within the aperture. We divided the aperture diameter into x_{num} pieces in the *x*-dimension and uniformly placed each *x*-mean $1/x_{num}$ apart from each other. Similarly, we divided the aperture into y_{num} pieces in the *y*-dimension and placed each *y*-mean $1/y_{num}$ apart from each other. Chebyshev point distribution is known to help reduce errors near the boundary[3]. However, we have not experimented extensively with this within the Code V implementation as of now. Preliminary experimentation also suggests that solving the problem on a slightly larger domain will help with errors near the problem boundary. The variances in the covariance matrix were set to 1.

Unlike the function fitting example given in Fig. 1 where we have a surface *S* to fit, in the context of the optical design problem, the surface *S* is unknown a priori, and the goal of an iterative optimizer is to adjust the weights *w* in equation

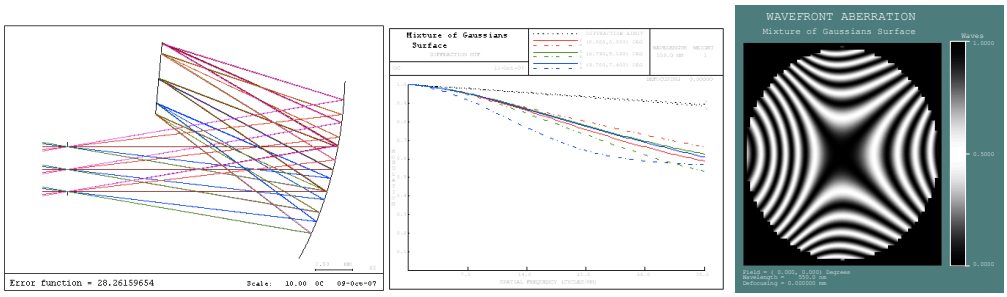
(2) with the goal of reaching a minimum of the merit function given a starting point. We used the transverse error in the image plane, which is the sum of squares of the deviations of the rays from their respective reference wavelength chief rays, as our merit function. The weight vector is found using the built-in damped least squares optimizer in Code V.

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

We designed four systems to address the question of optimal shape for the off-axis magnifier problem. A 10th order anamorphic sphere, an x-y polynomial, a 10th order Zernike polynomial and a linear combination of Gaussians were compared. Each system under comparison had a >15mm eye clearance, 3mm pupil, 24 degree diagonal full field of view (9.6° x semi-field and 7.2° y semi-field), and a ~15° mirror tilt angle. Each system was optimized with the minimum set of constraints such as the effective focal length (14.25mm), real ray based distortion constraints, and the field weights. Each system had 17 field points defined. The variables in each system included the surface coefficients, image plane defocus and tilt. The distance from the pupil to the mirror vertex was 16.9mm. The distance from the vertex of the mirror to the image plane was kept around 13mm since it is undesirable to have a microdisplay physically close to a human eye. The image plane has a rectangular aperture with a size of 4.8mm by 3.6mm in the x and y dimensions, respectively. The image plane needs to stay clear out of the ray path between the pupil and the mirror. 140 rays across the pupil were traced in each system during optimization. Figure 5(a) shows the layout of the optimal off-axis magnifier. The modulation transfer function (MTF), evaluated at λ =550nm (no dispersion), for the optimized linear combination of Gaussians surface is shown in Fig. 5(b). The interferogram of the surface with the base sphere subtracted is shown in Fig. 5(c). Distortion characteristics exhibit similar behavior with each design having a maximum of about 3 to 4%. Table 2 shows a comparison of the surface representation proposed and implemented in this paper against an anamorphic asphere, a Zernike polynomial up to and including 10th order, an x-y polynomial up to and including 10th order (good balancing achieved with order 5) with the maximum distortion and the average MTF across 17 field points as the comparison metrics. The sum of local basis representation proposed in this paper achieves the highest MTF performance averaged across 17 field points by 18.5% in the field with an acceptable level of maximum distortion, among the functions that were compared [25].

Table 2. Comparison of the transverse error function value and 17 average tangential and sagittal MTF values between an anamorphic asphere, an x-y polynomial, a Zernike polynomial, and a linear combination of Gaussians surface type.

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



(a) (b) (c)

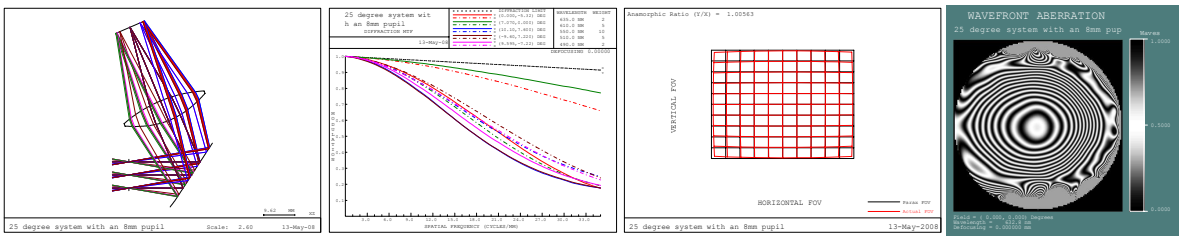
Fig. 5. (a) Optical layout of the off-axis magnifier. (b) MTF evaluated on-axis, 0.7 in the field, and at the maximum field. (c) Interferogram of the surface represented with a linear combination of Gaussian radial basis functions

3. 5 DUAL-ELEMENT DESIGN REVISITED: PUPIL SIZE AND FIELD OF VIEW LIMIT

In this section, guided by the results given in Table 2, we apply the radial basis function framework to the description of optical mirror shapes in a dual-element magnifier and study field of view[26] and pupil size[27] limits in this section.

3. 5.1 FIELD OF VIEW LIMIT WITH AN 8MM PUPIL

This subsection shows our results on the field of view limit for a dual-element magnifier designed with an 8mm pupil. Figure 6(a) shows the optical layout of the 8mm pupil, 15mm eye clearance, 25 degree diagonal full field of display. The modulation transfer function plotted out to 35 cycles/mm and evaluated at a 3mm pupil is shown in Fig. 6(b). The distortion grid is shown in Fig. 6(c). Our initial results indicate a 25° field of view limit while having an 8mm pupil and ~15 mm eye clearance. The interferogram of the optimized free-form surface at the He-Ne wavelength of 632.8nm is shown in Fig. 6(d).

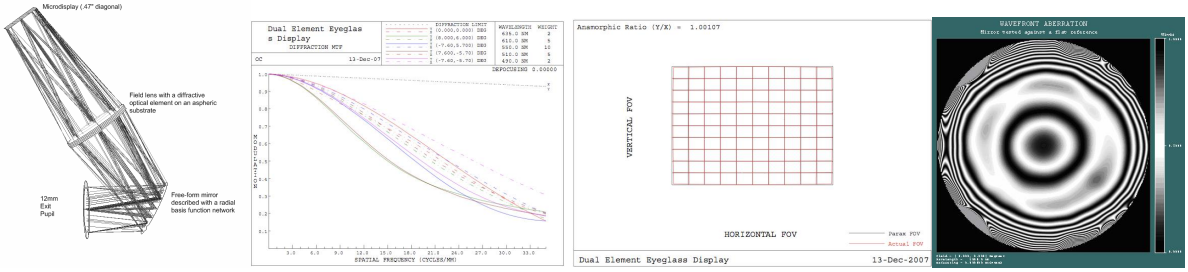


(a) (b) (c) (d)

Fig. 6. 25 degree magnifier designed with an 8mm pupil (a) Optical layout (b) Modulation transfer function plotted out to 35 cycles/mm evaluated with a 3mm pupil (c) Distortion grid (d) Interferogram of the free-form surface at the He-Ne wavelength of 632.8nm

3.5.2 PUPIL SIZE LIMIT WITH A 20 DEGREE FIELD OF VIEW

Fig. 7(a) shows the layout of the dual-element off-axis catadioptric magnifier where the mirror is described with Gaussian radial basis functions. A red, blue, and green microdisplay pixel triplet is 15 μm in size, which yields a Nyquist frequency of about 33 cycles/mm. The polychromatic MTF evaluated for a centered 12 mm pupil is plotted in Fig. 7(b) for the performance limiting fields. In the case of a 3 mm pupil, the lowest MTF value at the Nyquist frequency is about 50%. Thus the polychromatic MTF is plotted up to 35 cycles/mm. In this prototype, the microdisplay has a ~12 mm diagonal and contains 640x480 pixels. In visual space, the display provides 1.5 arcminutes resolution as limited by the pixel spacing on the microdisplay. The maximum distortion occurs at (x=-8°, y=-6°) in the field and was measured in simulation to be -1.83%. Fig.7(c) shows the appearance of a rectilinear grid as viewed through the magnifier. Fig.7(d) shows an interferogram of the mirror surface compared to a flat reference wavefront. We found the pupil size limit to be 12mm while having a 20° field of view and 15.5mm eye clearance.



(a) (b) (c) (d)

Fig. 7. 12mm pupil system with a 20° diagonal full FOV (a) Optical layout (b) Modulation transfer function plotted out to 35 cycles/mm (c) Distortion grid (d) Interferogram of the free-form mirror

4. APPLYING FULL-FIELD DISPLAYS TO ANALYZE THE DUAL-ELEMENT MAGNIFIER

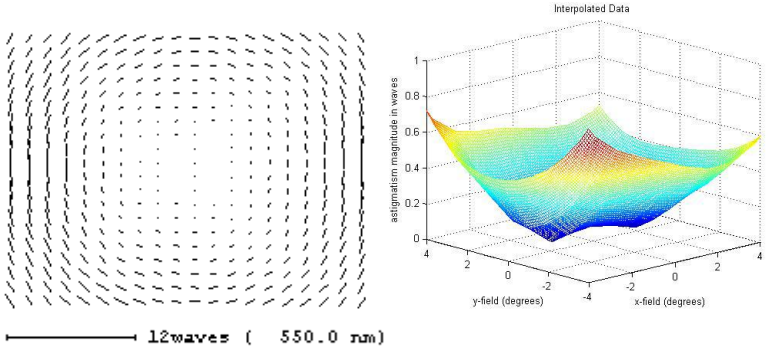
Full field displays from nodal aberration theory provide design insight into the aberration characteristics of tilted and decentered systems[28]. The full field displays are based on real raytrace data. The full field display plots can be used to provide magnitude and orientation information for individual aberration terms or for summary measures of image quality. Line length or circle size is proportional to the magnitude of the aberration of interest at the displayed grid point in the full field plot. To study the astigmatic aberration component in isolation, a generalized Coddington raytrace algorithm can be used in conjunction with the full-field display. In this case, the orientation of the displayed images aligns with the spot diagram. Zernike coefficient based full field displays are generated by tracing rays, calculating the OPD for each ray with respect to the chief ray, and then fitting Zernike polynomials to the OPD map. Thompson applied full field displays to the analysis of a triplet in Chapter 7 of his thesis[28]. Figoski applied nodal aberration theory to find an alternate solution to a two mirror telescope problem[29]. Rogers illustrated an application of nodal aberration theory to guide the design of a head-worn display[30].

In Figure 8, we show field maps prior to targeted Zernike optimization for the system shown in Fig. 6(a). Figure 8(a) displays FRINGE Zernike terms Z5 and Z6 representing astigmatism and Fig. 8(c) displays FRINGE Zernike polynomial terms Z7 and Z8 representing coma. We provide an interpolated surface in Fig.8(b) to the astigmatism magnitude given in Fig. 8(a) in order to clearly display the binodal behavior with the nodes being close to ±2 degrees along the *x*-direction with a 0 degree *y*-field component. In Figure 8(c), the minimum coma magnitude is 0.003 waves (λ=550nm) occurring at (-7.97°, -7.6°) in the field. The RMS wavefront error across the field is displayed in Fig. 8(d). The RMS wavefront error is quasi-constant with an average value of about 1.1 waves (λ=544.7nm) and a standard deviation of 0.6 waves across the entire field of view.

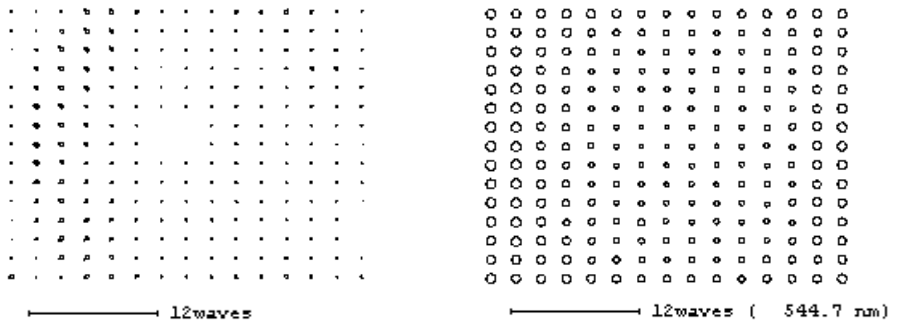
Figure 9 shows the field maps after Zernike targeted optimization. We further optimized the system by adding a constraint of targeting Z5, Z6, Z7 and Z8 to 0 at field 1 with the intention of collapsing the binodal astigmatism and coma nodes to the center of the field. Figure 9(a) displays Zernike terms Z5 and Z6 representing astigmatism. We can observe in Figure 9(b) that the binodal behavior is eliminated and we now have quadratic astigmatism at 0.5 degree in the x-field and 0.4 degree in the y-field. The RMS wavefront error across the field is displayed in Fig. 9(c). The RMS wavefront error across the field after Zernike targeted optimization and has an average of 1.1 wave (λ=544.7nm) across the field and 0.6 waves of standard deviation.

Figure 10 shows the fields maps after Zernike targeted optimization with the nodes at (±6°,0°). Figure 10(a) displays Zernike terms Z5 and Z6 representing astigmatism. We can observe the binodal behavior in Figure 10(b). The RMS wavefront error across the field is displayed in Fig. 10(c). The RMS wavefront error across the field after Zernike targeted optimization and has an average of 0.72 waves (λ=544.7nm) across the field and 0.15 waves of standard deviation.

Figure 11 summarizes the impact of moving the astigmatism nodes on the performance as measured with the MTF criteria. The MTF plots for the set of fields {(0°,0°),(0°,7.6°),(10.1°,0°),(-10.1°,0°),(9°,6.8°)} with binodal astigmatism nodes at (±2°,0°), (0.4°,0.5°), and (±6°,0°) are provided in Fig. 11(a), Fig. 11(b), and Fig. 11(c), respectively.

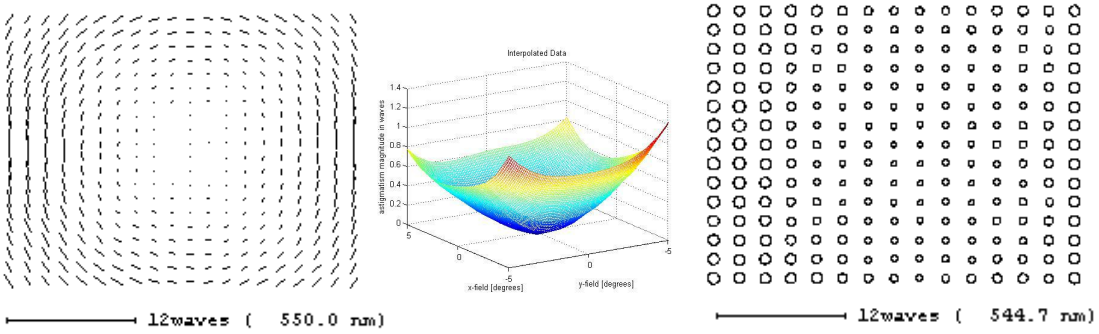


(a) (b)



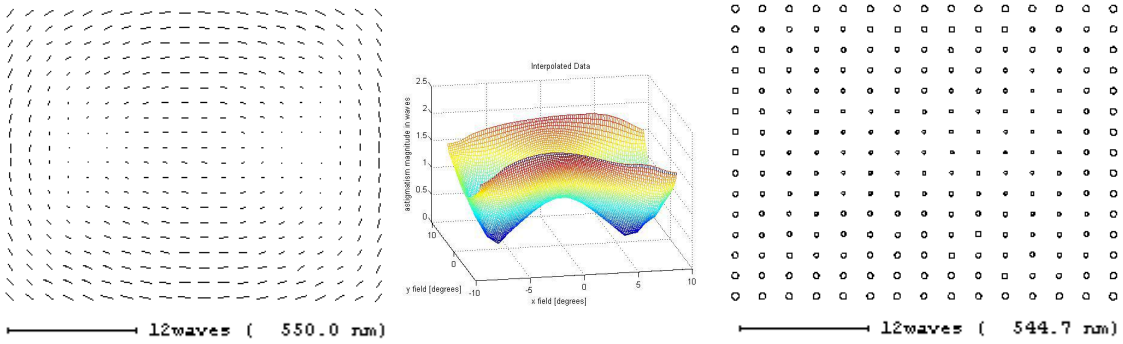
(c) (d)

Figure 8. Full field displays before targeting specific Zernike terms in the field (a) Full field display of the fringe Zernike Z5/Z6 components (astigmatism) in a 20x20 grid (b) Interpolated plot of the astigmatism magnitude to clearly visualize the binodal behavior (c) Full field display of the fringe Zernike Z7/Z8 (coma) (d) Full field display of the rms wavefront error (15x15 grid)



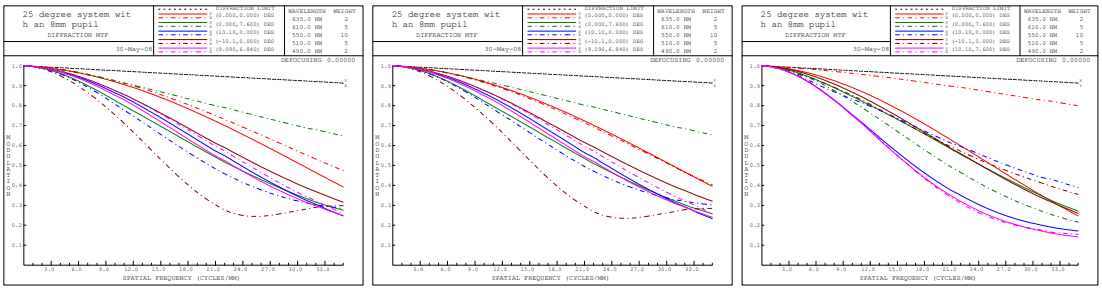
(a) (b) (c)

Figure 9. Full field displays with the astigmatism nodes collapsed at (0.5°, 0.4°) (a) Full field display of the fringe Zernike Z5/Z6 components (astigmatism) in a 20x20 grid (b) Interpolated plot of the astigmatism magnitude to clearly visualize the quadratic astigmatism behavior (c) Full field display of the rms wavefront error (15x15 grid)



(a) (b) (c)

Figure 10. Full field displays with the astigmatism nodes at (±6, 0.4°) (a) Full field display of the fringe Zernike Z5/Z6 components (astigmatism) in a 20x20 grid (b) Interpolated plot of the astigmatism magnitude to clearly visualize the binodal astigmatism behavior (c) Full field display of the rms wavefront error (15x15 grid)



(a) (b) (c)

Figure 11. Modulation transfer function plots for the set of fields {(0°,0°),(0°,7.6°), (10.1°,0°),(-10.1°,0°),(9°,6.8°)} in the cases of (a) nodes at (±2°,0°) (b) nodes at (0.4°,0.5°) and (c) nodes at (±6,0.4°).

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

5. CONCLUSIONS AND FUTURE WORK

Free-form surfaces provide additional degrees of freedom to optical designs having limited space requirements (in meter^3). We summarized our initial experiences in the design and fabrication of free-form optical surfaces in the design of head-worn displays. Specifically, 1) a dual-element magnifier (20 degree field of view, 8mm exit pupil, 15mm eye clearance, 1.5 arcminute resolution in visual space) using an x-y polynomial free-form surface was fabricated and 2) a single-element magnifier (24 degree field of view, 3mm exit pupil, 15mm eye clearance, 3.6 arcminute resolution) using an x-y polynomial free-form surface was fabricated. Additionally, we proposed the radial basis function framework for describing free-form optical surfaces. Comparisons of radial basis functions to standard free-form descriptions such as an anamorphic asphere, Zernike polynomial and an *x-y* polynomial were provided. The field of limit for an 8mm exit pupil dual-element magnifier with a Gaussian RBF free-form surface was estimated to be 25 degrees. The pupil size limit for a 20 degree field of view dual-element magnifier with a Gaussian RBF free-form surface was estimated to be 12mm. Full field displays were used to analyze the impact of astigmatic node placement.

The optical design examples given in section 3 used shape factors that were found through experimentation. The shape factor plays an important role in the accuracy of an RBF based approximation scheme. The theoretical tools for choosing a shape factor (e.g., cross-validation) take the data values into account. However, these data values are not known before our raytrace optimization procedure. This makes the application of techniques such as cross-validation challenging. Therefore, the link between a shape factor that yields maximal performance with respect to an optical performance criterion such as the RMS wavefront error or the modulation transfer function remains to be studied. Early experimentation suggests that it is possible to obtain a diffraction limited surface, in the case of different (not coincident) basis centers and data points, for the variances of 0.05, 0.1, 0.5, 1, 2, 3, 5, 10, 20, 30, 40 and 100. The high success rate for a relatively large range of variances could be either due to the simplicity of the paraboloid as a testcase or the differing centers and data points. Comparison of alternative radial basis functions such as a thin-plate spline and a multiquadric with the MTF criteria for the design problem shown in Fig. 5(a) remains to be studied as well.

ACKNOWLEDGMENTS

This work was supported by the Link Foundation and the National Science Foundation (NSF) IIS/HCI 03-07189 grant.

REFERENCES

1. G.D. Wasserman and E. Wolf, ``On the theory of aplanatic aspheric systems," Proc. Phys. Soc. B **62**, 2—8,1949.
2. D. Knapp, ``Conformal Optical Design," Ph.D. Thesis, University of Arizona, 2002.

3. G.E. Fasshauer. Meshfree Approximation Methods with MATLAB. Interdisciplinary Mathematical Sciences, Vol. 6. World Scientific, 2007. ISSN 1793-1355.
4. J.E. Stacy, ``Design of Unobscured Reflective Optical Systems with General Surfaces," Ph.D. Thesis, University of Arizona, 1983.
5. H. Chase. Optical design with rotationally symmetric NURBS. Proc. SPIE **4832**(10), 2002.
6. T. Davenport. ``Creation of a uniform circular illuminance distribution using faceted reflective NURBS", Ph.D. Thesis, University of Arizona, 2002.
7. G.W. Forbes. ``Shape specification for axially symmetric optical surfaces," Optics Express (**15)**8, pp.5218-5226, 2007.
8. S. Lerner, ``Optical Design using Novel Aspheric Surfaces," Ph.D. Thesis, University of Arizona, 2003.
9. L.A. Piegl and W. Tiller. The NURBS book. Springer, 1997.
10. J. M. Rodgers, ``Nonstandard representations of aspheric surfaces in optical design," Ph.D. Thesis, University of Arizona, 1984.
11. J. M. Rodgers and K. P. Thompson. Benefits of Freeform Mirror Surfaces in Optical Design. Proceedings of the American Society of Precision Engineering 2004 Winter Topical Meeting on Freeform Optics, **31**, pp.73-78, 2004.
12. A.K. Chan, C.K. Chui, and L.T. Guan. Radial Basis Function Approach to Interpolation of Large Reflecting Surfaces. SPIE Vol 1251 Curves and Surfaces in Computer Vision and Graphics, 1990.
13. A. Greynolds, "Propagation of general astigmatic Gaussian beams along skew ray paths", Diffractive Phenomena in Optical Engineering Applications, Proc. SPIE Vol. 560, pp. 33-50, 1985.
14. O. Cakmakci and J. P. Rolland. Head-Worn Displays: A Review. Invited paper. Journal of Display Technology **2**(3), pp.199-216, 2006.
15. Computerized Anthropometric Research & Design Laboratory, Warfighter Interface Division, Human Effectiveness Directorate, Air Force Research Labs. http://www.hec.afrl.af.mil/HECP/Card4.shtml. Last visited on July 22, 2006.
16. J. Rolland, Y. Ha, and C. Fidopiastis, "Albertian errors in head-mounted displays: I. Choice of eye-point location for a near- or far-field task visualization," J. Opt. Soc. Am. A 21, 901-912, 2004.
17. J. P. Rolland, C. Meyer, K. Arthur, E. Rinalducci. Method of Adjustments versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Head-Mounted Displays. Presence: Teleoperators & Virtual Environments, Vol. 11, No. 6: 610-625, December 2002.
18. Upton. H.W. Speech and Sound Display System. US3,463,885. Filed Oct. 22, 1965.
19. Upton. H.W. Eyeglass mounted visual display. US3,936,605. Filed Feb. 14, 1972.
20. Bettinger. Spectacle-mounted ocular display apparatus. US4,806,011. Filed Jul 6, 1987.
21. Spitzer M.B. “Eyeglass-Based Systems For Wearable Computing”. In Proc. First International Symposium on Wearable Computers (ISWC 1997), 13-14 October 1997, Cambridge, Massachusetts, USA. IEEE Computer Society. ISBN 0-8186-8192-6.
22. I. Kasai, Y. Tanijiri, T. Endo, and H. Ueda, “A Forgettable Near-Eye Display,” In Proc. International Symposium on Wearable Computing (2000), pp.115-118
23. H. Mukawa, K. Akutsu, I. Matsumura, S. Nakano, T. Yoshida, M. Kuwahara, K. Aiki, M. Ogawa. A full color eyewear display using holographic planar waveguides. In Proc. Society of Information Display 2008, Digest of Technical Papers, Volume XXXIX, pp.89-92.
24. O. Cakmakci and J. P. Rolland. Design and fabrication of a dual-element off-axis near-eye optical magnifier. Optics Letters (**32**)11, pp.1363-1365, 2007.
25. Ozan Cakmakci, Brendan Moore, Hassan Foroosh, and Jannick P. Rolland. Optimal local shape description for rotationally non-symmetric optical surface design and analysis. Optics Express **16**(3), pp.1583-1589, 2008.
26. O. Cakmakci, S. Vo, K.P. Thompson, and J.P. Rolland. Application of Radial Basis Functions to Shape Description in a Dual-Element Off-Axis Eyewear Display: Field of View Limit. Submitted to the special issue on mobile displays, Journal of Society of Information Display (JSID), June 2008.
27. O. Cakmakci, S. Vo, H. Foroosh and J. P. Rolland. Application of Radial Basis Functions to Shape Description in Dual-Element Off-Axis Magnifier. Optics Letters **33**(11), June 2008.
28. K. Thompson. Aberrations of Tilted and Decentered Systems. Ph.D. thesis, University of Arizona, 1980.
29. J. Figoski. Aberration characteristics of nonsymmetric systems. SPIE Vol. 554 International Lens Design Conference, 1985.
30. J. Rogers. Techniques and tools for obtaining symmetrical performance from tilted-component systems. Opt. Eng. **39**(7), July 2000, pp.1776-1787.