Analysis of Wavefront Coding Technology

Current and Future Potentialities for Optical Design

Future Work

Reporte Técnico

Sergio Vázquez y Montiel

Instituto Nacional de Astrofísica, Óptica y Electrónica

Tonantzintla Puebla, México

svazquez@inaoep.mx

July 2003

This work was funded by HOYA Corp. Tokyo Japan

INDICE

1. Introduction

In general optical system design consists of meeting specific requirement under some set of constraints. The constraints of a system vary from application to application. Number of optical elements (lenses, mirrors, prisms, DOEs, etc.), materials, cost, size and weight. Similarly, the performance requirements of a system are application dependent and often times include the minimization of aberrations. Some aberrations may or may not be important or they may be organized into a spectrum of tolerable to intolerable aberrations. Because meeting the system constraints can cause the aberrations to worsen and, conversely, minimizing the aberrations may require breaking the constraints, the challenge lies in making the tradeoff between the two. For example, the design of a digital color imaging system which uses only plastic optical elements is a nontrivial task. While plastic is a desirable material due to its low cost, ease of manufacturing, light weight and durability, it may greatly increase the difficulty in correcting chromatic aberration.

Control of lens aberrations has traditionally been a very difficult problem. The lens designer often works to determine a lens design where a number of aberrations, such as aberrations due to chromatic aberration, spherical aberration, astigmatism, coma, distortion, temperature dependent aberrations, fabrication tolerance, etc. are minimized over a broad spatial region in the image plane. The large number of constraints on the lens system invariably leads to designed systems with a large number of tightly specified optical elements.

If the number of constraints on the optical system can be decreased, or the number of lens aberrations that can be tolerated is increased, then it should be possible to develop simplified and/or less costly lens system.

One of the most prevalent problems in incoherent imaging systems is the limited range over which a system is able to produce an in-focus image. For example, in macroscopic systems (surveillance, endoscopy, etc.), depth of focus is important due to the three-dimensional properties of the objects being imaged. In microscopy depth of focus is equally as important for viewing small, detailed, three-dimensional objects. Imaging systems that image two-dimensional objects (i.e. flatbed scanners) have focus-related problems due to aberrations such as curvature of field or due to system misalignments caused by poor assembly, settling, temperature variations, etc. Consequently, these systems can also benefit from extended depth of focus.

Optical system designers have traditionally extended the depth of focus of optical systems by working with exit pupil definitions[1, 2, 3, 4, 5], usually stopping down the aperture until the desired focal depth has been reached. Three problems, however, quickly arise from stopping down exit pupils.

a) The first problem is the obvious reduction in the amount of light power allowed through the system. Reducing the exit pupil diameter of any system to 1/x times that

of the original system will require that x^2 times more power or x^2 times more exposure time be used in the system before an image of equal exposure is produced.

b) The second problem is that stopping down the aperture reduces the resolution of the system (by performing a type of low pass filtering).

c) The third problem is that stopping down the aperture will also increase the risk of object motion during an increased exposure time, resulting in a blurred image.

By applying the ambiguity function from the radar field to optical transfer functions, Thomas Cathey and Edward Dowski[6] have developed a special phase plate which can be placed in the aperture stop of an optical system. This causes the optical system to be focus invariant over a significantly larger range than the focal range of a standard optical system.

The cubic phase plate works by causing the optical transfer function (OTF) of any plane in the desired range to be invariant from the OTF of any other plane in that range. They call the effect of the phase plate on the OTF "**wavefront coding**".

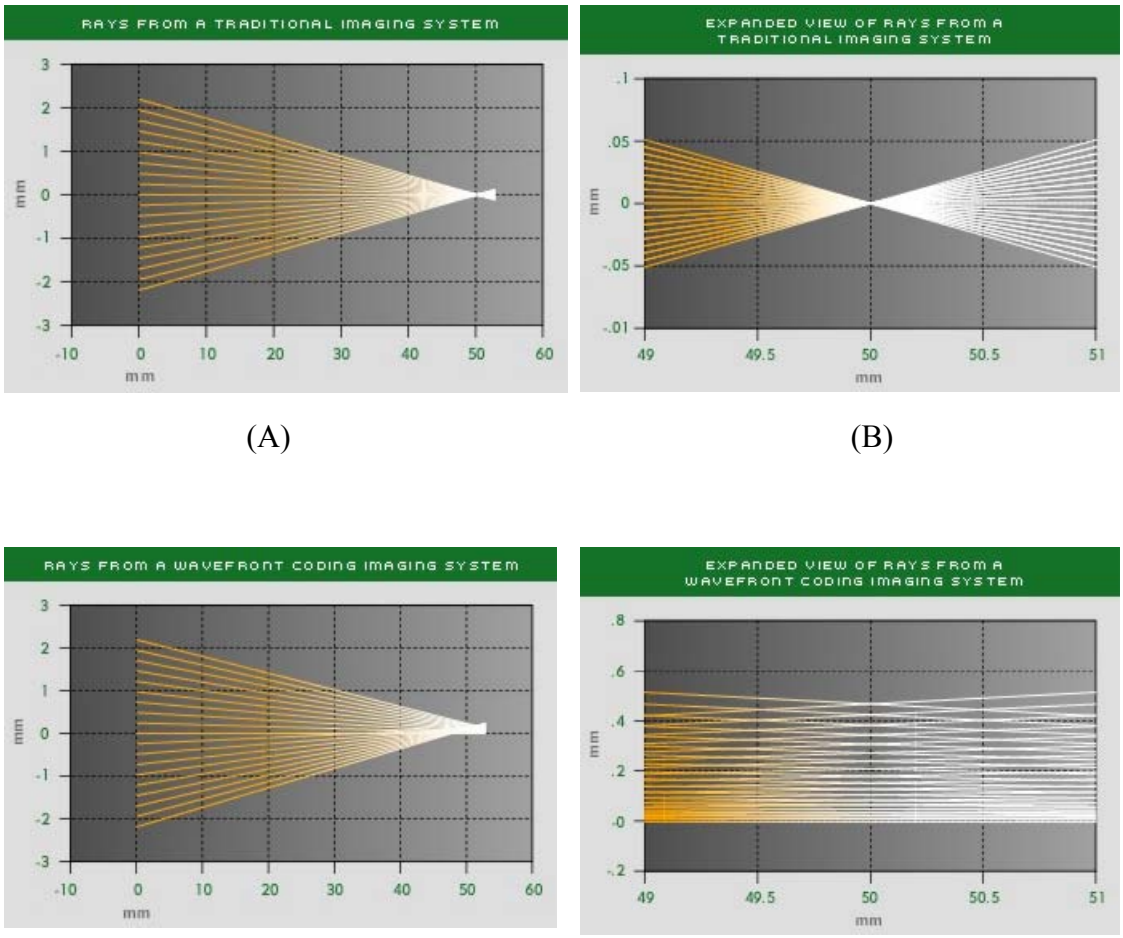
These coded optical transfer functions are not those of a standard in-focus system, but, nevertheless, they are invariant to defocus. This invariance allows us to perform simple signal processing on an entire image, thus allowing the whole range of the object to be in focus. Wavefront coding allows objects containing three dimensional properties beyond the standard depth of focus to be completely and clearly resolved without the loss of light that occurs when the aperture size is reduced.

A series of ray diagrams illustrates how wavefront coding technology is able to increase the depth of field of an imaging system by encoding images to make them invariant to defocus, see figure 1.1.

A thorough description of the development and mathematics of this special phase plate can be found in next sections.

In the section 2, we explain with detail the concepts and necessary tools to understand the wavefront coding technology. We begin with the concepts of wavefront aberrations and pupil function because they are necessary to understand the point spread function and the optical transfer function. it is also explained the ambiguity function because it is very useful to understand the variations of the optical transfer function with defocus and later we show too that ambiguity function is also useful as design tool for the wavefront coding technology. To determine which phase surface is necessary to code the wavefront a technique call the method of stationary phase is used, this method is explained with detail. Another topic that it is approached in section 2 it is the corresponding to images processing, in particular, the convolution concept and its use in the filters design.

In the section 3, we explain the technique of the wavefront coding technology with all detail. Finally, we give conclusions.



(A) (B)

(C) (D)

Figure 1.1. Rays of a point source focused by a one-dimensional lens (A) without and (C) with a cubic phase plate incorporated. The detail near the normal image plane is shown in (B) and (D).

2. Analysis Tools

One of the most useful tools for understanding optical systems is the Optical Transfer Function (OTF). The OTF of an optical system is the Fourier transform of the Point Spread Function (PSF) or impulse response of that system. For understanding these concepts, I first explain the aberration concept and the pupil function.

2.1. Wavefront aberrations

If P is any object point and P' is its Gaussian image, then all rays from P, i.e. all those transmitted by the aperture stop, should pass through P'; alternatively the wavefronts from P should, in image space, be portions of spheres centred on P'. Aberrations appear as a non-fulfillment of either condition. It is possible that the rays and wavefronts from P might all converge to a point P'' in image space which is not P' but which is near P'. This leads to a very useful distinction between point-imaging aberrations, in which the rays in image space are not concurrent, and aberrations of image shape, where each object point forms a true image point but there is not the correct similarity between object and image shapes. The latter aberrations are, of course, field curvature and distortion while the former are spherical aberration, coma, astigmatism and higher order compounds of complex form.

The finite ray tracing process give information about aberrations as ray deviations, and this is a relatively simple matter if a fast computer is available, it is also fairly simple to supplement the ray tracing procedure to obtain the wave front shape errors directly. Thus we have already solved the problem of determining the aberrations of a given optical system. However, the inverse problem, that of obtaining a system to have given aberrations, i.e. zero aberrations to within certain tolerances, is the basic problem of optical design.

In figure 2.1, let Σ be the wavefront of the pencil from P and let S be a reference sphere with center P'_0 and radius P'_0O. Let another ray **r** of the pencil from P meet S and Σ in Q_0 and Q respectively and let it meet the image plane at P'. The wavefront aberration is then defined as the optical path length from Q_0 to Q, i.e. it is n Q_0Q, where n is the refractive index of the space in question. It is a function of the position of Q_0, thus we write it as **W(x, y)** and expresses *the deformation of the wavefront from the ideal spherical shape*.

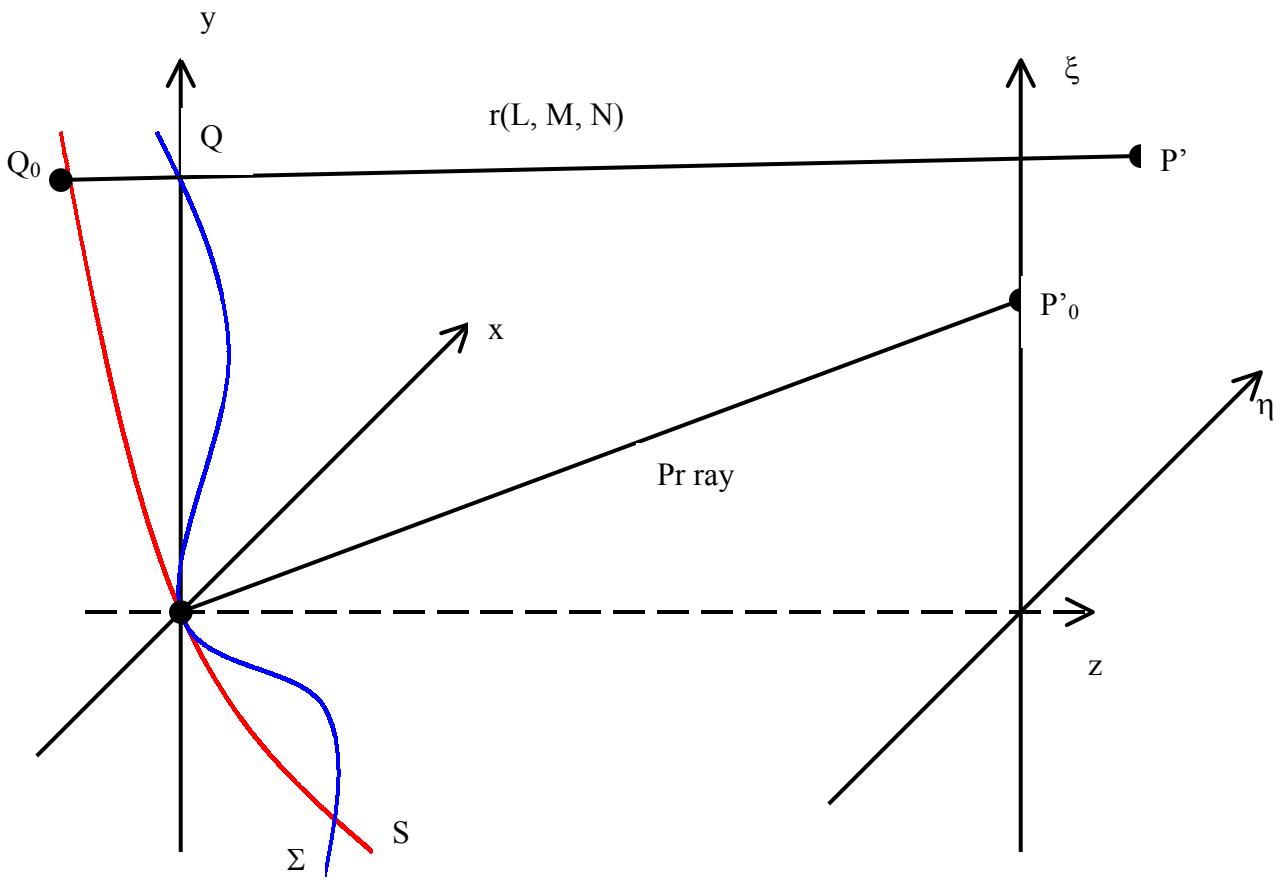


Figure 2.1. Wavefront aberrations

Sir William Hamilton developed a elegant theory of aberration types and he assume that W can be expanded as a power series and he write

$$W(y_0,\rho,\theta)=\sum_{j,m,n} W_{k,l,m} y_0^k \rho^l \cos^m\theta, \quad \text{with } k=2j+m,\ \ l=2n+m, \qquad (2.1)$$

or

$$W(y_0,\rho,\theta)=W_{200}y_0^2+W_{111}y_0\rho\cos\theta+W_{020}\rho^2+W_{040}\rho^4+W_{131}y_0\rho^3\cos\theta$$
$$+W_{222}y_0^2\rho^2\cos^2\theta+W_{220}y_0^2\rho^2+W_{311}y_0^3\rho\cos\theta+\ldots+\text{fifth and higher order terms} \qquad (2.2)$$

where y_0 is the height of image.

The names of the wavefront aberrations coefficients are given in table 1. Note that piston, tilt and focus are first-order properties of the wavefront.

Table 1

Wavefront aberration Coefficient	Functional form	Name
W_{200}	y_0^2	Piston
W_{111}	$y_0\rho\cos\theta$	Tilt
W_{020}	ρ^2	Focus
W_{040}	ρ^4	Spherical
W_{131}	$y_0\rho^3\cos\theta$	Coma
W_{222}	$y_0^2\rho^2\cos^2\theta$	Astigmatism
W_{220}	$y_0^2\rho^2$	Field curvature
W_{311}	$y_0^3\rho\cos\theta$	distortion

2.2. Pupil function

The complex amplitude of the electrical field in the exit pupil produced by a point object is proportional to $\exp\{(2\pi j/\lambda)W(x, y)\}$, thus we can define a **pupil function** as

$$F(x,y) = e^{j\frac{2\pi}{\lambda}W(x,y)}, \tag{2.3}$$

over the area of the pupil and is zero elsewhere.

Let the distance of the pupil from the image plane be R, i.e. this is the radius of the reference sphere. Then the complex amplitude in the point image at (ξ, η) is, apart from a constant factor,

$$A(\xi,\eta) = \int_{-\infty}^{\infty}\int F(x,y)e^{-j\frac{2\pi}{\lambda R}(\xi x+\eta y)}dxdy\,. \tag{2.4}$$

In this expression the integration is actually over the area of the pupil, but it is customary to define the pupil function *F(x, y)* formally as above in order to be able to put in infinite limits for the integration. This device is adopted so that the formalism of Fourier transform theory can be used. It can be seen that if we regard $\xi/\lambda R$ and $\eta/\lambda R$ as single variables eqn. (2.4) does express $A(\xi, \eta)$ as the two-dimensional Fourier transform of $F(x, y)$ except that the variables ξ and η are to be re-scaled. This is true for any shape of pupil and any aberrations; varying transmission over the pupil can be incorporated in *F(x, y)* and a shift of focus an be produced by adding a focal shift term to the aberration function *W(x, y)*.

2.3. Point Spread Function

If an optical system has no aberrations the image of a bright point object against a dark background, the so-called point spread function (PSF), is a rather complex diffraction figure of which the form depends on the shape of the exit pupil of the system. It is customary to make certain approximations in calculating the PSF; these are that polarization effects are ignored, that convergence angles are small and that the exit pupil is large in diameter compared to the wavelength; these are usually known as the scalar diffraction theory approximations and it is an experimental fact that the scalar theory is very accurate under the conditions which usually hold in optics.

The intensity point spread function, which is what is actually measured or seen, is the squared modulus of the amplitude spread function:

$$I(\xi,\eta) = |A(\xi,\eta)|^2 . \tag{2.5}$$

I have omitted constant factors which would give appropriate dimensions to the physical quantities involved, since these will disappear in the normalizing process I will use. The appropriate factors can be found in Born an Wolf[7].

The PSF for a circular pupil and zero aberrations is the Airy pattern, shown in fig. 2.2, of which the light intensity distribution is

$$I(\eta) = \left(\frac{2J_1(v)}{v} \right)^2 , \qquad v = \frac{2\pi n a}{\lambda R}\eta \tag{2.6}$$

in this equation η is the radial distance from the center of the pattern, *a* is the radius of the pupil, *R* is the radius of the reference sphere, *n* is the refractive index, and λ is the vacuum wavelength of the light. There are no physical units in this equation: it has been normalized so that the intensity of the center of the pattern is unity, as is customary in work of this kind.

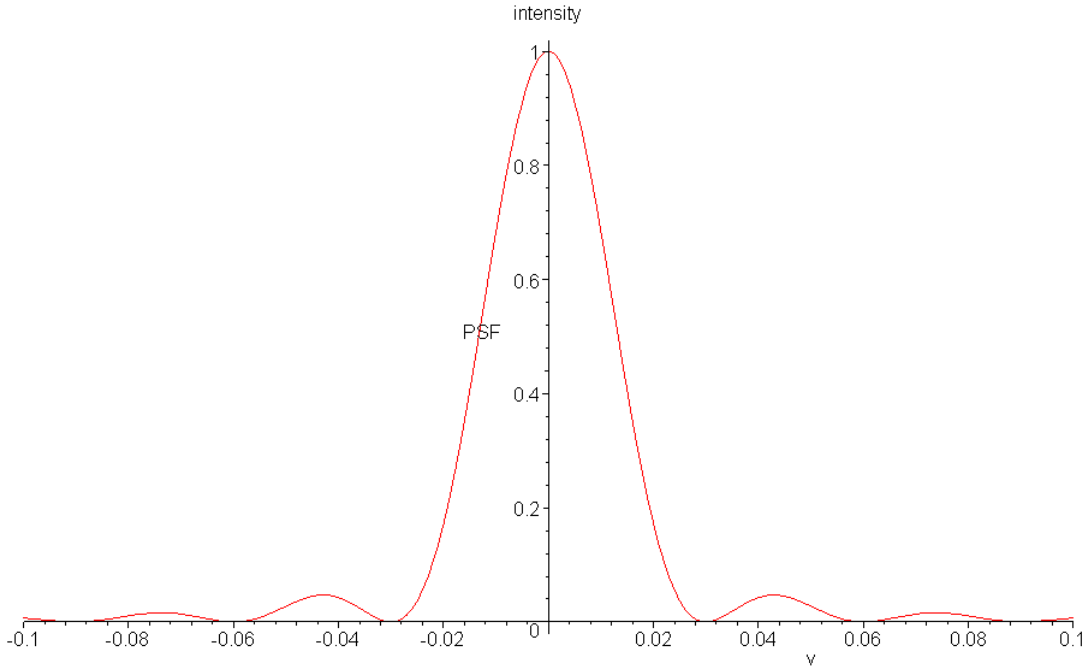


Figure 2.2. The Airy pattern or aberration-free PSF for a circular pupil, linear scale of intensity

2.4. The Optical transfer Function

The optical transfer function (OTF) concept originally developed by analogy with the electrical communications concept of a transfer function for a network or an amplifier, for example, an amplifier will amplify purely sinusoidal signals of different frequencies by different gain factors; the amplification as a function of frequency is the transfer function or response function. In optical system we introduce a notional sinusoidal test object, which can be thought of as a grating with a sinusoidal transmission profile across the rulings. If this is illuminated incoherently it can be shown that the image formed of it by an optical system with *any* aberrations will be of sinusoidal form but with reduced contrast, i.e. the dark parts of the image will not have zero intensity. The reduction in contrast will depend on the aberration of the optical system and on the spatial frequency of the grating; this reduction in contrast as a function of spatial frequency is given by the OTF.

The two-dimensional OTF is calculated as the normalized inverse Fourier transform of the PSF

$$H(s,t)=\frac{\int_{-\infty}^{\infty}\int I(\xi,\eta)e^{j2\pi(s\xi+t\eta)}d\xi d\eta}{\int_{-\infty}^{\infty}\int I(\xi,\eta)d\xi d\eta}, \qquad (2.7)$$

here s and t are spatial frequency.

An alternative expression for the OTF can be obtained in terms of the auto-correlation of the pupil function:

$$H(s,t)=\frac{\int_{-\infty}^{\infty}\int F(\lambda Rs+x,\lambda Rt+y)F^{*}(x,y)dxdy}{\int_{-\infty}^{\infty}\int |F(x,y)|^{2}dxdy}. \tag{2.8}$$

Fig. 2.3 show the OTF for an aberration-free circular pupil of radius a as a function of one variable and is given by

$$H(s)=\frac{2}{\pi}\cos^{-1}\left(\frac{\lambda Rs}{2a}\right)-\frac{\lambda Rs}{\pi a}\sqrt{1-\frac{\lambda^{2}R^{2}s^{2}}{4a^{2}}}. \tag{2.9}$$

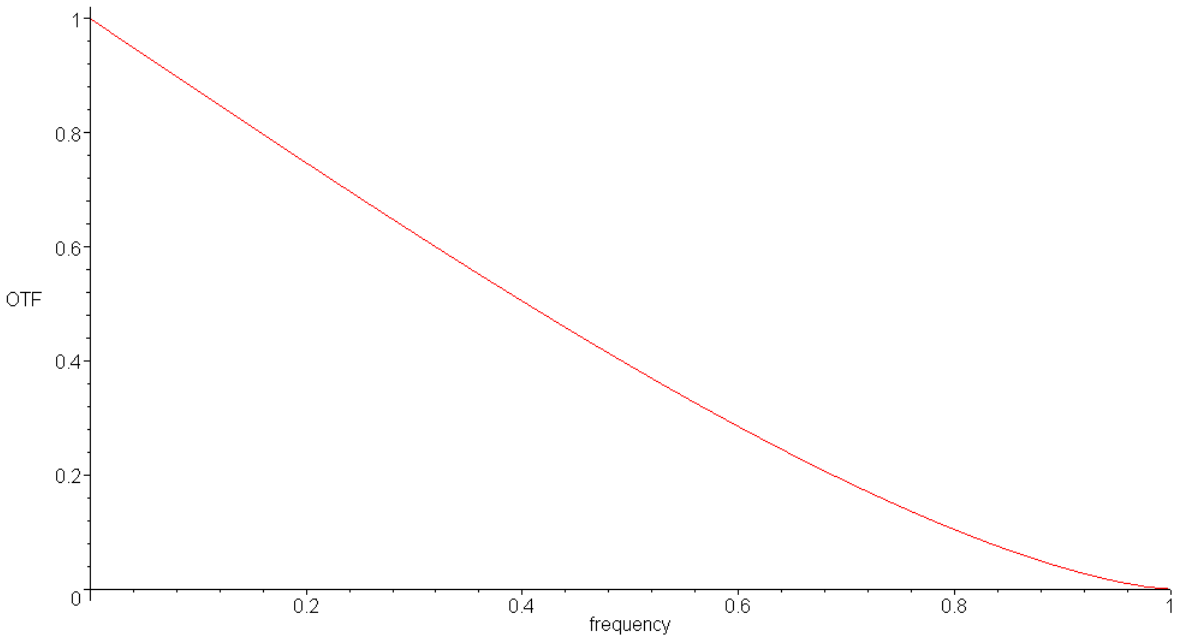


Fig. 2.3. The OTF for an aberration free circular pupil, the OTF is plotted against the normalized spatial frequency, $(\lambda R/2a)s$

2.5. The Ambiguity Function as a polar Display of the OTF

The Ambiguity Function (AF) was defined by Woodward[8] for applications in radar ranging. This definition was employed by Papoulis[9] for the theory of Fresnel diffraction

and of coherent image formation. These results were extended by Guigay[10] to describe Fresnel diffraction and isoplanatic imaging, under partially coherent illumination, and Brenner[11] indicated that the AF can be employed as a polar display of the OTF with variable focus errors.

Let us write the OTF as

$$H(u)=\frac{\int_{-\infty}^{\infty} Q\left(v+\frac{u}{2}\right)Q^{*}\left(v-\frac{u}{2}\right)dv}{\int_{-\infty}^{\infty}|Q(v)|^{2}dv}, \tag{2.10}$$

where u and v denote normalized spatial frequencies, and Q is the generalized pupil function given as

$$Q(u)=P(u)e^{jk\left(W(u)+W_{020}u^{2}\right)}=Q_{0}(u)e^{jkW_{020}u^{2}}, \tag{2.11}$$

where the pupil itself is

$$P(u)=\begin{cases}1 & \text{if } |u|<1,\\ 0 & \text{if } |u|>1,\end{cases} \tag{2.12}$$

and $W(u)$ is the 1-D aberration function. The defocus term is conveniently specified explicitly by $W_{020}u^2$.

Thus

$$H(u;W_{020})=\frac{1}{2}\int Q_{0}(v+u/2)Q_{0}^{*}(v-u/2)e^{j2\pi\left(\frac{2W_{020}}{\lambda}\right)uv}dv, \tag{2.13}$$

this function has been calculated for the 2-D case by Hopkins[12].

Now, from the standard definition of the AF[9] we have that

$$A(u,y)=\int_{-\infty}^{\infty}Q_{0}(v+u/2)Q_{0}^{*}(v-u/2)e^{j2\pi vy}dv, \tag{2.14}$$

where y denotes the normalized coordinate $y = Y/\lambda$.

Thus, from eqs. (2.13) and (2.14) we have that

$$H(u;W_{020}) = A\left(u, \frac{2uW_{020}}{\lambda}\right), \tag{2.15}$$

in other words, the AF associated with the base function Q_0 contains the OTF along the line

$$y = \frac{2uW_{020}}{\lambda}, \tag{2.16}$$

for a given amount W_{020} of defocusing. Consequently, the whole 2-D AF contains the OTFs for arbitrary amounts of defocusing, arranged in a polar fashion, see fig. 2.4.

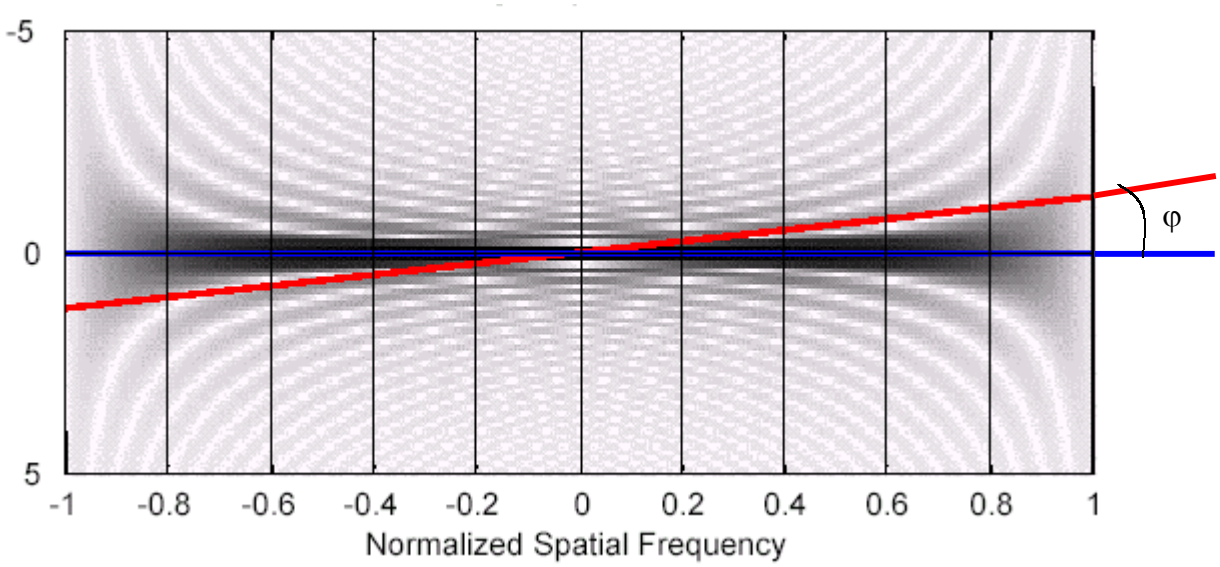


Figure 2.4. Ambiguity Function, AF, of a rectangular pupil. The pupil coordinated are r and φ.

In particular, for the in-focus case, $W_{020} = 0$, is clear that

$$H(u;0) = A(u,0), \tag{2.17}$$

if actual polar coordinates are employed for the AF; then eq.(2.15) becomes

$$H(u;W_{020}) = A(r,\varphi), \tag{2.18}$$

where

$$W_{020} = \frac{\lambda}{2}\tan\varphi\,, \tag{2.19}$$

and

$$u = \frac{r}{\sqrt{1+\tan^2\varphi}}\,. \tag{2.20}$$

This result is illustrated with the simple example of a diffraction limited 1-D pupil, i.e.

$$Q_0(u) = P(u) = rect(u). \tag{2.21}$$

The AF in this case is

$$A(u,y) = \left(1 - \frac{|u|}{2}\right)\mathrm{sinc}\left[\pi y\left(2-|u|\right)\right], \tag{2.22}$$

which is shown in fig. 2.4. The zero value loci are the symmetric hyperbolas

$$y = \frac{n}{2-|u|}\,, \tag{2.23}$$

where $n = 1, 2, \ldots$ The zeros of the OTF for different amounts of defocus can be found as the intersection of the above hyperbolas with the lines $y = u\tan\varphi$. That is

$$u = 1 \pm \sqrt{1 - n\tan\varphi}\,. \tag{2.24}$$

Thus, there are an odd number of zeros, $2n$-1, when

$$\tan\varphi = n \quad \text{or} \quad y_{\max} = 2n \quad \text{or} \quad W_{020} = n\lambda/2, \tag{2.25}$$

there are an even number of zeros, $2n$, when

$$n + 1 > \tan\varphi > n \quad \text{or} \quad 2(n+1) > y_{max} > 2n \quad \text{or} \quad (n+1)\lambda/2 > W_{020} > n\lambda/2. \tag{2.26}$$

These results are shown in fig. 2.5 for the case $n = 16$. The actual OTFs[12] are presented in fig. 2.6.

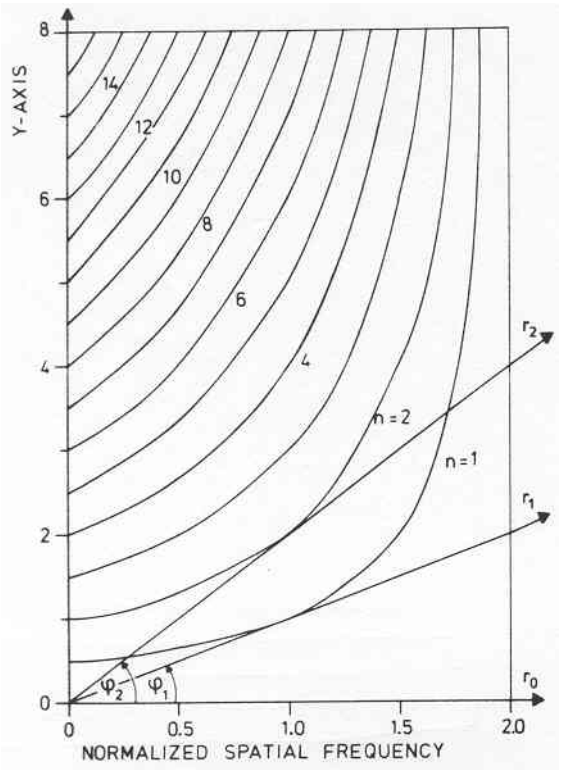


Fig. 2.5. Zero value loci of a quarter-plane of fig. 2.4

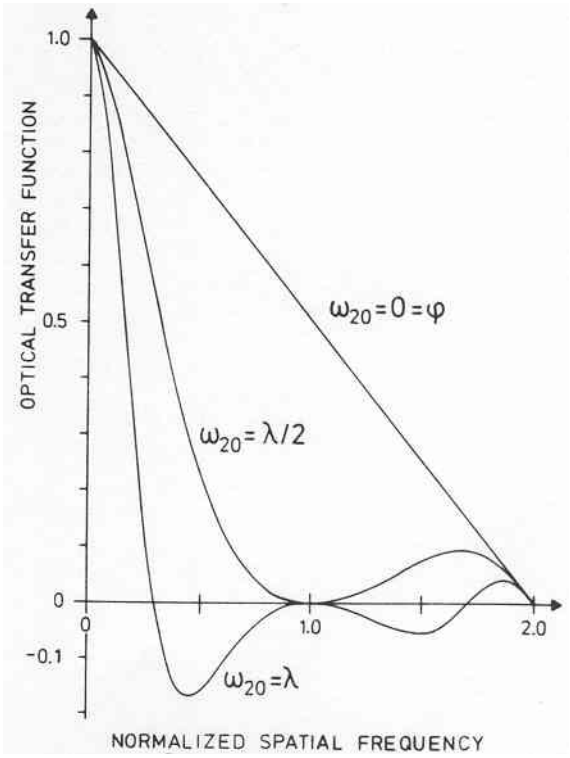


Fig. 2.6. Variation of the OTF with focus error

The treatment as presented so far for one-dimensional pupil functions can be extended easily to two-dimensional pupils. Writing down the equations corresponding to eqs. (2.11), (2.14), and (2.15)

$$Q(u_1,u_2)=Q_0(u_1,u_2)e^{jkW_{020}\left(u_1^2+u_2^2\right)}, \tag{2.27}$$

$$A(u_1,u_2,y_1,y_2)=\int\int Q_0(v_1+u_1/2,v_2+u_2/2)Q_0^*(v_1-u_1/2,v_2-u_2/2)e^{j2\pi(v_1y_1+v_2y_2)}dv_1dv_2 \tag{2.28}$$

$$H(u_1,u_2;W_{020})=A(u_1,u_2,2u_1W_{020}/\lambda,2u_2W_{020}/\lambda). \tag{2.29}$$

This result can be interpreted geometrically in the following way. The OTF H with coordinates (u_1,u_2) and parameter W_{020} occupies a two-dimensional plane in the four-dimensional ambiguity function A, associated with the pupil function $Q_0(u_1,u_2)$. The plane containing the 2-D OTF is defined by

$$y_1=\frac{2u_1W_{020}}{\lambda} \quad \text{and} \quad y_2=\frac{2u_2W_{020}}{\lambda}. \tag{2.30}$$

These planes always crosses the origin for all values of the defocus parameter W_{020}.

Hopkins's tolerance criterion establishes that the ratio between the aberrated and defocused OTF and the diffraction-limited OTF should be ≥ 0.8 at any specific spatial frequency; that is,

$$M(u;W_{020})=\left|\frac{H(u;W_{020})}{H(u;0)}\right|=\left|\frac{A(u;y)}{A(u;0)}\right|\geq 0.8. \tag{2.31}$$

This means that we can visualize, as a drop in the brightness of the AF along the line u = constant, the variations of $M(u;W_{020})$. Thus, by simply looking at the AF one is able to judge the tolerance of an optical system to varying focus errors.

2.6. The Method of Stationary Phase

A very useful procedure for obtaining approximations to various integrals that frequently occur in wave theory is based on the so-called method of stationary phase. It provides an asymptotic approximation to integrals for large values of an appropriate parameter.

Let us consider an integral of the form

$$F(k) = \int_a^b f(x) e^{jkg(x)} dx \,, \tag{2.32}$$

where *f(x)* and *g(x)* are real, well-behaved functions of a real variable *x* and *a* and *b* are real constants. The parameter *k* is also assumed to be real. Without loss of generality, we may assume *k* to be positive. In physical applications *F(k)* often represents the combined effect of waves of amplitudes *f(x)* and phases *g(x)*, all having the same wave number *k*.

to obtain some idea about the behavior of the integral (2.32) as a function of *k*, we consider a simple example. We take

$$f(x) \equiv 1 \,, \qquad g(x) = x^2 \,, \tag{2.33}$$

and consider the real part of the integral:

$$\mathrm{Re}\, F_1(k) = \int_a^b \cos\left(kx^2\right) dx \,, \qquad (a<0,\ b>0). \tag{2.34}$$

Let us compare the behavior of the integrand

$$G(x,k) = \cos\left(kx^2\right), \tag{2.35}$$

in eq. (2.34) for different values of *k*. with fixed *k*, *G(x, k)* will oscillate between the values +1 and –1, with the rate of the oscillations depending on the value of *k*. Since, with *k* fixed, the zeros of *G(x, k)* are given by

$$x = \pm\sqrt{\left(n + \frac{1}{2}\right)\frac{\pi}{k}} \,, \qquad n = 0,\ 1,\ 2,\ 3,\ \ldots., \tag{2.36}$$

it is clear that if one takes *k* larger and larger, *G(x, k)* will oscillate more and more rapidly, see fig. 2.7.

Suppose now that we drop the assumption that *f(x) = 1,* but still take $g(x) = x^2$. instead of eq. (2.34) we then have the integral

$$\mathrm{Re}\, F_2(k) = \int_a^b f(x) \cos\left(kx^2\right) dx \,. \tag{2.37}$$

The factor *f(x)* will give rise to amplitude modulation of the cosine term; but it is clear that irrespective of the exact form of *f(x)* , if only *k* is large enough, the integrand of the integral on the right-hand side of eq. (2.37) will again oscillate very rapidly and there will be a

tendency for the positive and negative contributions of the integrand to cancel out. Moreover this cancellation can be expected to take place irrespective of whether $g(x) = x^2$ or whether it has some other form. However, for sufficiently larger values of k, the cancellation will not be complete in the neighborhood of point where $g(x)$ is stationary within the interval of integration, i.e. where

$$\frac{dg(x)}{dx} = 0\,, \tag{2.38}$$

and also at the end points

$$x = a \quad \text{and} \quad x = b. \tag{2.39}$$

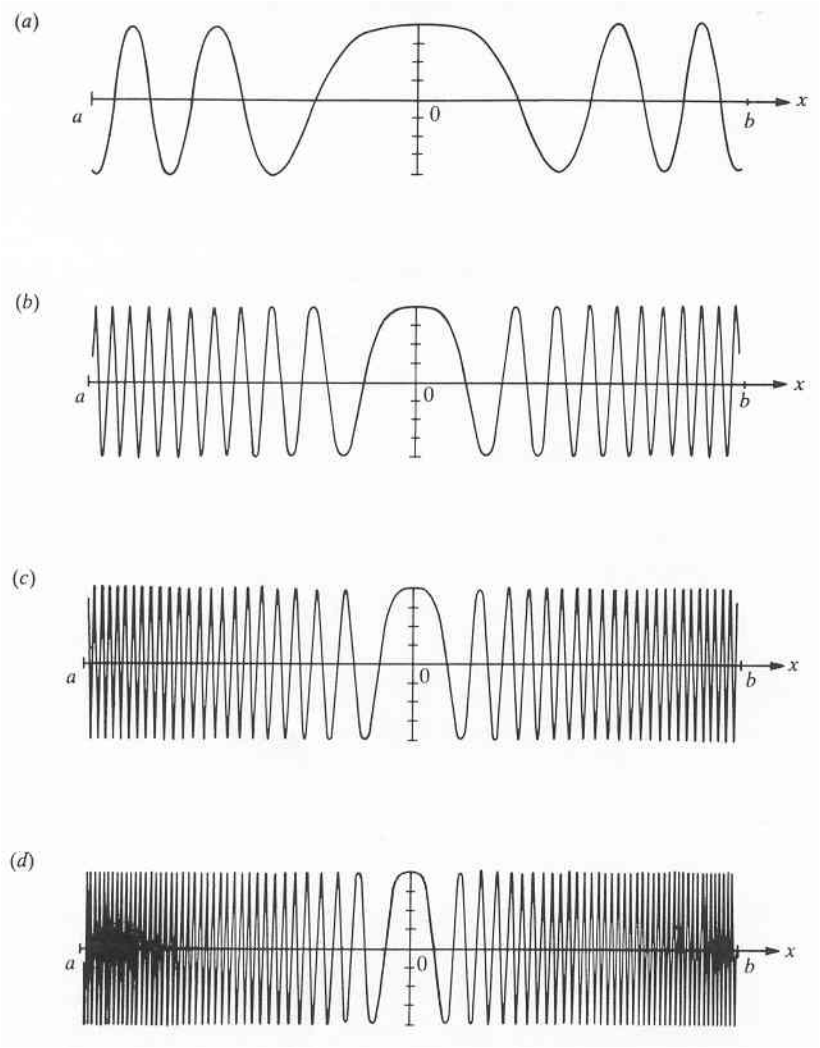


Figure 2.7. Illustrating the principle of stationary phase. Comparison of the behavior of the functions (a) $G(x,1) = \cos x^2$; (b) $G(x,4) = \cos 4x^2$; (c) $G(x,9) = \cos 9x^2$; (d) $G(x,16) = \cos 16x^2$.

These special points are called critical points of the integrand in eq. (2.32). Those that satisfy eq. (2.38) are said to be critical points of the first kind, and the end points (2.39) are said to be critical points of the second kind. Of course, in special cases several critical points coincide or one or both of the end points may themselves be stationary points of *g(x)*.

Other complications may arise, for example when *g(x)* or *f(x)* have singular behavior. However, excluding these more complicated situations one can show that the asymptotic behavior as $k \to \infty$ of an integral of the form (2.32) is determined entirely by the behavior of the integrand at the critical points, and, moreover, the leading term in the asymptotic expansion of *F(k)* often depends on the critical points of the first kind, i.e. on interior points in the range of integration where *g(x)* is stationary. This fact is the essence of the **principle of stationary phase**.

Let us assume that in the integral (2.32) *f(x)* is continuous and *g(x)* is twice continuously differentiable in the interval $a < x < b$. Suppose, to begin with, that there is one and only one critical point of the first kind, i.e. that there is one point x_1 and no other point in the interval at which

$$g'(x) = 0, \tag{2.40}$$

where the prime denotes differentiation with respect to x. We also assume that the second derivative of *g(x)* is not zero at $x = x_1$. then at points x in the immediate neighborhood of x_1,

$$g(x) \approx g(x_1) + \frac{1}{2}(x - x_1)^2 g''(x_1). \tag{2.41}$$

Since, according to the principle of stationary phase, the asymptotic approximation to the integral for large values of k comes from the immediate neighborhood of x_1, we have from eqs. (2.32) and (2.41)

$$F(k) \approx \int_a^b f(x_1) e^{jkg(x_1)} e^{jk(x-x_1)^2 g''(x_1)/2} dx. \tag{2.42}$$

For the same reason we may extend the range of integration from (a, b) to $(-\infty, +\infty)$ and we obtain the formula

$$F(k) \approx f(x_1) e^{jkg(x_1)} \int_{-\infty}^{\infty} e^{jk(x-x_1)^2 g''(x_1)/2} dx. \tag{2.43}$$

If we change the variable of integration from x to $u = x - x_1$ and use the symmetry of the integrand, the formula (2.43) becomes

$$F(k) \approx 2 f(x_1) e^{jkg(x_1)} \int_0^\infty e^{jkg''(x_1)u^2/2} du \,. \tag{2.44}$$

The integral on the right side is the well-known Fresnel integral, whose value is

$$\int_0^\infty e^{jau^2} du = \frac{1}{2}\left(\frac{\pi}{|a|}\right)^{1/2} e^{\pm j\pi/4}, \tag{2.45}$$

then

$$F^{(1)}(k) \approx \left[\frac{2\pi}{k|g''(x_1)|}\right]^{1/2} f(x_1) e^{jkg(x_1)} e^{\pm j\pi/4} \qquad \text{as } k \to \infty, \tag{2.46}$$

we have written $F^{(1)}(k)$ rather than $F(k)$ to stress that the expression on the right-hand side of eq. (2.32) is a contribution from a critical point of the first kind

Next we will determine the contributions from critical points of the second kind. We will assume that the end points are not stationary points of *g(x)*, i.e. that

$$g'(a) \neq 0, \quad g'(b) \neq 0. \tag{2.47}$$

We re-write the integral (2.32) in the form

$$\int_a^b f(x) e^{jkg(x)} dx = \frac{1}{jk} \int_a^b \left[\frac{d}{dx} e^{jkg(x)}\right] \frac{f(x)}{g'(x)} dx \,. \tag{2.48}$$

Integrating by parts on the right-hand side we obtain the formula

$$\int_a^b f(x) e^{jkg(x)} dx = \frac{1}{jk}\left[\frac{f(b)}{g'(b)} e^{jkg(b)} - \frac{f(a)}{g'(a)} e^{jkg(a)}\right] - \frac{1}{jk}\int_a^b \frac{d}{dx}\left[\frac{f(x)}{g'(x)}\right] e^{jkg(x)} dx \,. \tag{2.49}$$

Again integrating by parts, we find at once that the third term on the right-hand side of eq. (2.49) is of the order of $1/k^2$ and hence can be neglected in comparison with the two other terms. Thus we may conclude that the contribution from critical points of the second kind is

$$F^{(2)}(k) \approx \frac{1}{jk}\left[\frac{f(b)}{g'(b)} e^{jkg(b)} - \frac{f(a)}{g'(a)} e^{jkg(a)}\right]. \tag{2.50}$$

The preceding analysis indicates that when k is large enough

$$F(k) \approx F^{(1)}(k) + F^{(2)}(k). \tag{2.51}$$

2.7. Some things about image processing

2.7.1. Digital images

A digital image $a[m, n]$ described in a 2D discrete space is derived from an analog image $a(x, y)$ in a 2D continuous space through a *sampling* process that is frequently referred to as digitization.

The 2D continuous image $a(x, y)$ is divided into N *rows* and M *columns*. The intersection of a row and a column is termed a *pixel*. The value assigned to the integer coordinates $[m, n]$ with $\{m=0,1,2,\ldots,M-1\}$ and $\{n=0,1,2,\ldots,N-1\}$ is $a[m, n]$. In fact, in most cases $a(x, y)$—which we might consider to be the physical signal that impinges on the face of a 2D sensor—is actually a function of many variables including depth (z), color (l), and time (t).

2.7.2. Convolution

There are several possible notations to indicate the convolution of two (multidimensional) signals to produce an output signal. The most common are:

$$c = a \otimes b = a * b. \tag{2.52}$$

In 2D continuous space:

$$c(x, y) = a(x, y) \otimes b(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\chi, \zeta) b(x - \chi, y - \zeta) d\chi d\zeta. \tag{2.53}$$

In 2D discrete space:

$$c[m, n] = a[m, n] \otimes b[m, n] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a[j, k] b[m - j, n - k]. \tag{2.54}$$

There are a number of important mathematical properties associated with convolution.

• Convolution is *commutative*.

$$c = a \otimes b = b \otimes a. \tag{2.55}$$

• Convolution is *associative*.

$$c = a \otimes (b \otimes d) = (a \otimes b) \otimes d = a \otimes b \otimes d \,. \tag{2.56}$$

• Convolution is *distributive*.

$$c = a \otimes (b + d) = (a \otimes b) + (a \otimes d), \tag{2.57}$$

where a, b, c, and d are all images, either continuous or discrete.

2.7.3. Statistics

In image processing it is quite common to use simple statistical descriptions of images and sub–images. The notion of a statistic is intimately connected to the concept of a probability distribution, generally the distribution of signal amplitudes. For a given region—which could conceivably be an entire image—we can define the probability *distribution* function of the brightnesses in that region and the probability *density* function of the brightnesses in that region. We will assume in the discussion that follows that we are dealing with a digitized image $a[m, n]$.

The probability distribution function, $P(a)$, is the probability that a brightness chosen from the region is less than or equal to a given brightness value a. As a increases from $-\infty$ to $+\infty$, $P(a)$ increases from 0 to 1. $P(a)$ is monotonic, nondecreasing in a and thus $\mathrm{d}P/\mathrm{d}a \geq 0$.

The probability that a brightness in a region falls between a and $a+\Delta a$, given the probability distribution function $P(a)$, can be expressed as $p(a)\Delta a$ where $p(a)$ is the probability density function:

$$p(a)\Delta a = \left(\frac{dP(a)}{da} \right) \Delta a \,. \tag{2.58}$$

For an image with quantized (integer) brightness amplitudes, the interpretation of Δa is the width of a brightness interval. We assume constant width intervals. The brightness probability *density* function is frequently estimated by counting the number of times that each brightness occurs in the region to generate a *histogram,* $h[a]$. The histogram can then be normalized so that the total area under the histogram is 1. Said another way, the $p[a]$ for a region is the normalized count of the number of pixels, Λ, in a region that have quantized brightness a:

$$p[a] = \frac{1}{\Lambda} h[a] \quad \text{with} \quad \Lambda = \sum_a h[a]. \tag{2.59}$$

Both the distribution function and the histogram as measured from a region are a statistical description of that region. It must be emphasized that both $P[a]$ and $p[a]$ should be viewed as *estimates* of true distributions when they are computed from a specific region. That is,

we view an image and a specific region as one realization of the various random processes involved in the formation of that image and that region. In the same context, the statistics defined below must be viewed as estimates of the underlying parameters.

The average brightness of a region is defined as the *sample mean* of the pixel brightnesses within that region. The average, m_a, of the brightnesses over the Λ pixels within a region ($\Re$) is given by:

$$m_a = \frac{1}{\Lambda} \sum_{(m,n) \in R} a[m,n]. \tag{2.60}$$

Alternatively, we can use a formulation based upon the (unnormalized) brightness histogram, $h(a) = \Lambda \bullet p(a)$, with discrete brightness values a. This gives:

$$m_a = \frac{1}{\Lambda} \sum_{a} a \bullet h[a]. \tag{2.61}$$

The average brightness, m_a, is an estimate of the mean brightness, μa, of the underlying brightness probability distribution.

The *unbiased estimate* of the standard deviation, s_a, of the brightnesses within a region ($\Re$) with Λ pixels is called the *sample standard deviation* and is given by:

$$\begin{aligned} s_a &= \sqrt{\frac{1}{\Lambda - 1} \sum_{(m,n) \in R} (a[m,n] - m_a)^2} \\ &= \sqrt{\frac{\sum_{(m,n) \in R} a^2[m,n] - \Lambda m_a^2}{\Lambda - 1}} \end{aligned} . \tag{2.62}$$

Using the histogram formulation gives:

$$s_a = \sqrt{\frac{\left(\sum_{a} a^2 \bullet h[a] \right) - \Lambda \bullet m_a^2}{\Lambda - 1}} . \tag{2.63}$$

2.7.4. Signal-to-Noise ratio

The signal–to–noise ratio, *SNR*, can have several definitions. The noise is characterized by its standard deviation, s_n. The characterization of the signal can differ. If the signal is known to lie between two boundaries, $a_{min} \le a \le a_{max}$, then the *SNR* is defined as:

$$SNR = 20\log_{10}\left(\frac{a_{\max} - a_{\min}}{s_n}\right) dB. \tag{2.64}$$

If the signal is not bounded but has a statistical distribution then two other definitions are known:

S & N inter-dependent
$$SNR = 20\log_{10}\left(\frac{m_a}{s_n}\right) dB. \tag{2.65}$$

S & N independent
$$SNR = 20\log_{10}\left(\frac{s_a}{s_n}\right) dB. \tag{2.66}$$

2.7.5. Convolution-Based Operations

Convolution, the mathematical, *local* operation defined in Section 2.7.2 is central to modern image processing. The basic idea is that a window of some finite size and shape—the *support*—is scanned across the image. The output pixel value is the weighted sum of the input pixels within the window where the weights are the values of the filter assigned to every pixel of the window itself. The window with its weights is called the *convolution kernel*. This leads directly to the following variation on eq. (2.54). If the filter $h[j, k]$ is zero outside the (rectangular) window $\{j=0,1,\ldots,J-1;\ k=0,1,\ldots,K-1\}$, then, using eq. (2.55), the convolution can be written as the following finite sum:

$$c[m,n] = a[m,n] \otimes h[m,n] = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} h[j,k]a[m-j,n-k]. \tag{2.67}$$

This equation can be viewed as more than just a pragmatic mechanism for smoothing or sharpening an image.

In a variety of image-forming systems an appropriate model for the transformation of the physical signal $a(x, y)$ into an electronic signal $c(x, y)$ is the convolution of the input signal with the impulse response of the sensor system. This system might consist of both an optical as well as an electrical sub-system. If each of these systems can be treated as a linear, shift-invariant (*LSI*) system then the convolution model is appropriate. The definitions of these two, possible, system properties are given below:

Linearity - *If* $a_1 \to c_1$ *and* $a_2 \to c_2$ *Then* $w_1 \bullet a_1 + w_2 \bullet a_2 \to w_1 \bullet c_1 + w_2 \bullet c_2$ (2.68)

Shift Invariance - *If* $a(x, y) \rightarrow c(x, y)$ *Then* $a(x - x_0, y - y_0) \rightarrow c(x - x_0, y - y_0)$ (2.69)

where w_1 and w_2 are arbitrary complex constants and x_o and y_o are coordinates corresponding to arbitrary spatial translations.

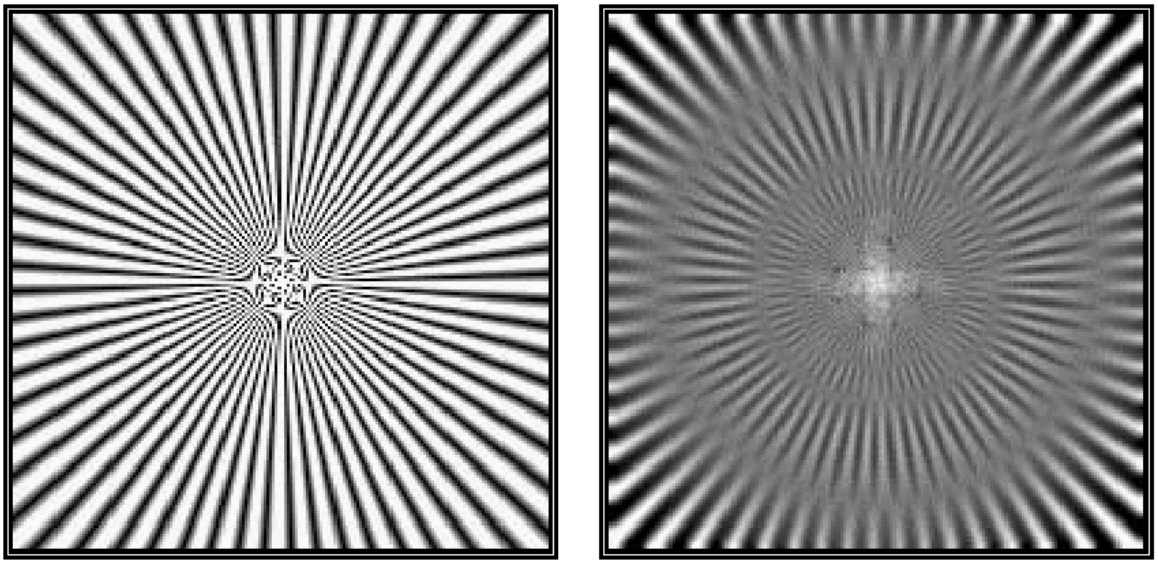
Two remarks are appropriate at this point. First, linearity implies (by choosing $w_1 = w_2 = 0$) that "zero in" gives "zero out". The offset means that such camera signals are not the output of a linear system and thus (strictly speaking) the convolution result is not applicable. Fortunately, it is straightforward to correct for this non-linear effect.

Second, optical lenses with a magnification, M, other than 1X are not shift invariant; a translation of 1 unit in the input image $a(x, y)$ produces a translation of M units in the output image $c(x, y)$. Due to the Fourier property this case can still be handled by linear system theory.

If an impulse point of light $d(x, y)$ is imaged through an LSI system then the impulse response of that system is called the *point spread function* (*PSF*). The output image then becomes the convolution of the input image with the *PSF*. The Fourier transform of the *PSF* is called the *optical transfer function* (*OTF*). For optical systems that are circularly-symmetric, aberration-free, and diffraction-limited the *PSF* is given by the Airy disk.

If the convolution window is not the diffraction-limited PSF of the lens but rather the effect of defocusing a lens then an appropriate model for $h(x, y)$ is a pill box of radius a. The effect on a test pattern is illustrated in Figure 2.8.

The effect of the defocusing is more than just simple blurring or smoothing. The almost periodic negative lobes in the transfer function produce a 180° phase shift in which black turns to white and vice-versa. The phase shift is clearly visible in Figure 2.8b.



a) Test pattern **b)** Defocused image

Figure 2.8: Convolution of test pattern with a pill box of radius a=4.5 pixels.

In describing filters based on convolution we will use the following convention. Given a filter $h[j, k]$ of dimensions J x K, we will consider the coordinate $[j=0,k=0]$ to be in the center of the filter matrix, **h**. This is illustrated in Figure 2.9. The "center" is well-defined when J and K are odd; for the case where they are even, we will use the approximations $(J/2, K/2)$ for the "center" of the matrix.

$$\mathbf{h} = \begin{bmatrix} h\left[-\left(J-\tfrac{1}{2}\right),-\left(K-\tfrac{1}{2}\right)\right] & \cdots & \cdots & h\left[0,-\left(K-\tfrac{1}{2}\right)\right] & \cdots & \cdots & h\left[\left(J-\tfrac{1}{2}\right),-\left(K-\tfrac{1}{2}\right)\right] \\ \vdots & \ddots & \vdots & \vdots & \vdots & \iddots & \vdots \\ \vdots & \cdots & h[-1,-1] & h[0,-1] & h[1,-1] & \cdots & \vdots \\ h\left[-\left(J-\tfrac{1}{2}\right),0\right] & \cdots & h[-1,0] & h[0,0] & h[1,0] & \cdots & h\left[\left(J-\tfrac{1}{2}\right),0\right] \\ \vdots & \cdots & h[-1,1] & h[0,1] & h[1,+1] & \cdots & \vdots \\ \vdots & \iddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h\left[-\left(J-\tfrac{1}{2}\right),\left(K-\tfrac{1}{2}\right)\right] & \cdots & \cdots & h\left[0,\left(K-\tfrac{1}{2}\right)\right] & \cdots & \cdots & h\left[\left(J-\tfrac{1}{2}\right),\left(K-\tfrac{1}{2}\right)\right] \end{bmatrix}$$

Figure 2.9. Coordinate system for describing $h[j, k]$

When we examine the convolution sum eq. (2.67) closely, several issues become evident.

• Evaluation of formula (2.67) for $m = n = 0$ while rewriting the limits of the convolution sum based on the "centering" of $h[j, k]$ shows that values of $a[j, k]$ can be required that are outside the image boundaries:

$$c[0,0] = \sum_{j=-J_0}^{J_0} \sum_{k=-K_0}^{K_0} h[j,k]a[-j,-k] \qquad J_0 = \frac{(J-1)}{2}, \quad K_0 = \frac{(K-1)}{2}. \tag{2.70}$$

The question arises – what values should we assign to the image $a[m, n]$ for $m < 0$, $m \geq M$, $n < 0$, and $n \geq N$? There is no "answer" to this question. There are only alternatives among which we are free to choose assuming we understand the possible consequences of our choice. The standard alternatives are a) extend the images with a constant (possibly zero) brightness value, b) extend the image periodically, c) extend the image by mirroring it at its boundaries, or d) extend the values at the boundaries indefinitely. These alternatives are illustrated in figure 2.10.

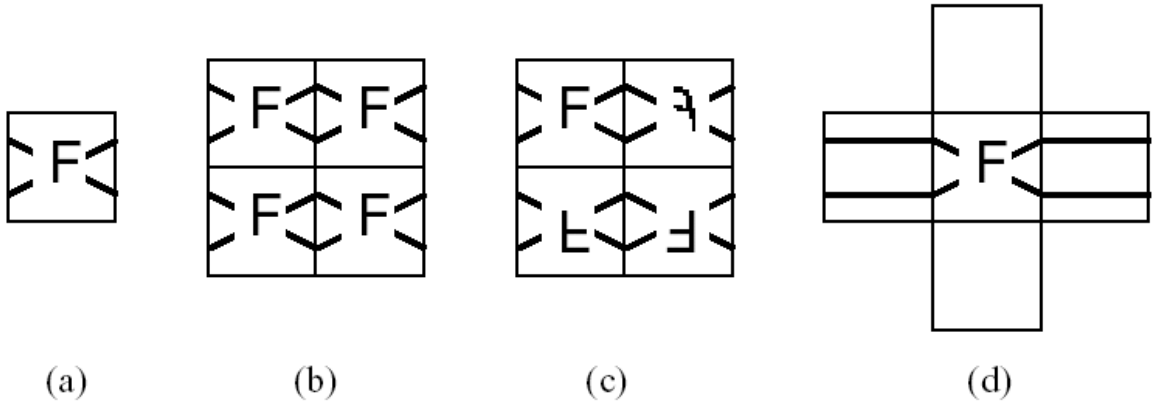


Figure 2.10. Examples of various alternatives to extend an image outside its formal boundaries.

• When the convolution sum is written in the standard form eq. (2.54) for an image $a[m, n]$ of size M x N:

$$c[m,n] = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} a[j,k]h[m-j,n-k]. \tag{2.71}$$

we see that the convolution kernel $h[j, k]$ is mirrored around $j = k = 0$ to produce $h[–j, –k]$ before it is translated by $[m, n]$ as indicated in eq. (2.71). While some convolution kernels in common use are symmetric in this respect, $h[j, k] = h[–j, –k]$, many are not. Care must therefore be taken in the implementation of filters with respect to the mirroring requirements.

• The computational complexity for a K x K convolution kernel implemented in the spatial domain on an image of N x N is $O(K^2)$ where the complexity is measured *per pixel* on the basis of the number of multiplies-and-adds.

• The value computed by a convolution that begins with integer brightnesses for $a[m, n]$ may produce a rational number or a floating point number in the result $c[m, n]$. Working exclusively with integer brightness values will, therefore, cause round off errors.

• Inspection of eq. (2.67) reveals another possibility for efficient implementation of convolution. If the convolution kernel $h[j, k]$ is *separable*, that is, if the kernel can be written as:

$$h[j,k] = h_{row}[k] \bullet h_{col}[j]. \tag{2.72}$$

then the filtering can be performed as follows:

$$c[m,n] = \sum_{j=0}^{J-1} \left\{ \sum_{k=0}^{K-1} h_{row}[k] a[m-j, n-k] \right\} h_{col}[j]. \tag{2.73}$$

This means that instead of applying one, two-dimensional filter it is possible to apply two, one-dimensional filters, the first one in the k direction and the second one in the j direction. For an N x N image this, in general, reduces the computational complexity per pixel from $O(J \bullet K)$ to $O(J+K)$.

An alternative way of writing separability is to note that the convolution kernel (figure 2.9) is a matrix **h** and, if separable, **h** can be written as:

$$[\mathbf{h}] = [\mathbf{h_{col}}] \bullet [\mathbf{h_{row}}]^t, \tag{2.74}$$

where "t" denotes the matrix transpose operation. In other words, **h** can be expressed as the *outer product* of a column vector [$\mathbf{h_{col}}$] and a row vector [$\mathbf{h_{row}}$].

• For certain filters it is possible to find an *incremental implementation* for a convolution. As the convolution window moves over the image, see eq. (2.71), the leftmost column of image data under the window is shifted out as a new column of image data is shifted in from the right. Efficient algorithms can take advantage of this and, when combined with separable filters as described above, this can lead to algorithms where the computational complexity per pixel is $O(constant)$.

There is an alternative method to implement the filtering of images through convolution, based in Fourier transforms, $\boldsymbol{F}$, by the following sequence of operations:

i) Compute $A(\Omega,\Psi) = \boldsymbol{F}\{a[m, n]\}$

ii) Multiply $A(\Omega,\Psi)$ by the *precomputed* $H(\Omega,\Psi) = \boldsymbol{F}\{h[m, n]\}$ (2.75)

iii) Compute the result $c[m, n] = \boldsymbol{F}^{-1}\{A(\Omega,\Psi) \bullet H(\Omega,\Psi)\}$

• While it might seem that the "recipe" given above in eq. (2.75) circumvents the problems associated with direct convolution in the spatial domain—specifically, determining values for the image outside the boundaries of the image—the Fourier domain approach, in fact, simply "assumes" that the image is repeated periodically outside its boundaries as illustrated in figure 2.10b. This phenomenon is referred to as *circular convolution*.

If circular convolution is not acceptable then the other possibilities illustrated in figure 2.10 can be realized by embedding the image $a[m, n]$ and the filter $H(\Omega, \Psi)$ in larger matrices with the desired image extension mechanism for $a[m, n]$ being explicitly implemented.

• The computational complexity per pixel of the Fourier approach for an image of N x N and for a convolution kernel of K x K is $O(\log N)$ *complex independent of* K. Here we assume that $N > K$ and that N is a highly composite number such as a power of two. This latter assumption permits use of the computationally-efficient Fast Fourier Transform (*FFT*) algorithm. Surprisingly then, the indirect route described by eq. (2.75) can be faster than the direct route given in eq. (2.67). This requires, in general, that $K^2 >> \log N$. The range of K and N for which this holds depends on the specifics of the implementation. For the machine on which this manuscript is being written and the specific image processing package that is being used, for an image of $N = 256$ the Fourier approach is faster than the convolution approach when $K \geq 15$. (It should be noted that in this comparison the direct convolution involves only integer arithmetic while the Fourier domain approach requires complex floating point arithmetic.)

3. Extended Depth of Field through Wavefront Coding

3.1. Introduction

The depth of field of an imaging system is the distance in the object space in which objects are considered to be in focus. As such, extending the depth of field of an imaging system will extend the distance over which objects are in focus. Extending the depth of field of incoherent optical systems has been an active research topic for many years. The majority of the literature on this topic has concerned methods of employing an optical power absorbing apodizer, with possible $\pm\pi$ phase variations, on a standard incoherent optical system as a means to increase the depth of field[1, 2, 3, 4, 5] . These methods have all suffered from two significant deficiencies; a decrease of optical power at the image plane, and a decrease of image resolution. For example, stopping down a lens one F-stop reduces the amount of light passing through the lens by a factor of two. A typical high speed imaging application may require the aperture to be stopped down to the F/30 range, necessitating added illumination on the order of kilowatts to obtain useful data. A unique method of achieving extended depth of field without an apodizer[13] was described in 1972. The major shortcoming of this method is that focus must be varied during exposure.

Thomas Cathey and Edward Dowski[6, 14, 15] describe a novel method for extending the depth of field of incoherent optical systems that does not suffer from the significant deficiencies of earlier methods. Their method employs a **phase mask** to modify the incoherent optical system in such a way that the PSF is insensitive to defocus, while forming an OTF that has no regions of zero values within its pass band. The PSF of the modified optical system is not directly comparable to that produced from a diffraction limited PSF. However, because the OTF has no regions of zeros, digital processing can be used to "restore" the sampled intermediate image. Further, because the OTF is insensitive to defocus, the same digital processing restores the image for all values of defocus. This combined optical/digital system produces a PSF that is comparable to that of the diffraction limited PSF, but over a far larger region of focus. They term the general process of modifying the incoherent optical system and received incoherent wavefront, by means of a phase mask, ***wavefront coding.***

By modifying only the phase of the received wavefront, general wavefront coding techniques maximize optical power at the image plane. Optics Wavefront Coding technology enables digital imaging systems to extend the depth of field by ten or more times without stopping down the aperture or increasing illumination. To achieve extended depth of field, Wavefront Coding technology uses generalized aspheric optics to encode images to be invariant to defocus, and digital signal processing to decode the images. Using this technology, an F/2 system can produce a depth of field equivalent to an F/20 system while retaining the light gathering ability and spatial resolution of the F/2 system.

When designing extended depth of field systems, Cathey and Dowski make two main assumptions:

a) The first is that the incoherent optical system is being modified by a rectangularly separable phase mask. This leads to a rectangularly separable PSF and OTF.

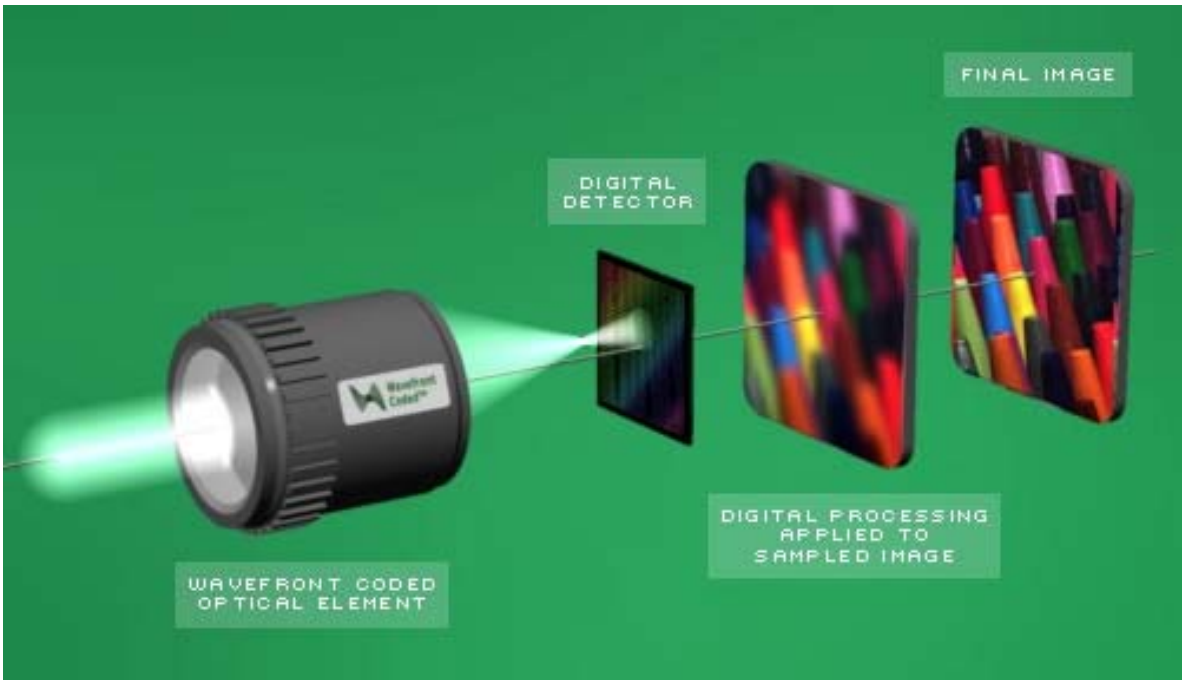
b) Secondly, they assume that any resulting image will be an "intermediate image". This intermediate image will require digital processing. This second assumption follows from their belief that best performance is obtained by optimum preprocessing through optics, followed by optimum digital postprocessing[16]. These pre- and post-processing stages are optimum in the sense that each is "matched" to the other in order to solve an interesting problem.

A Wavefront Coded system differs from a classical digital imaging system in two fundamental ways.

a) First, the light traveling through a Wavefront Coded lens system does not focus on a specific focal plane. Because of a special surface that is placed in the lens system at the aperture stop, no points of the object are imaged as points on the focal plane. Rather, these points are uniformly blurred over an extended range about the focal plane. This situation is referred to as "encoding" the light passing through the lens system. Another way to describe this effect is to say that the special Wavefront Coded surface in the lens system changes the ray paths such that each ray (except the axial ray) is deviated differently from the path that it would take in a classical, unaltered lens system and therefore they do not converge at the focal plane, see figure 1.1.

b) The second difference found in a Wavefront Coded system is that the image detected at the detector is not sharp and clear, as discussed above, and thus must be "decoded" by a subsequent digital filtering operation. The image from the digital detector is filtered to produce an image that is sharp and clear, but has non-classical properties such as a depth of field (or depth of focus) that is much greater then that produced by an unaltered classical lens system of the same f number.

The illustration below graphically depicts a Wavefront Coded imaging system.



Cathey and Dowski solution to extended depth of field systems relies on the theory of the ambiguity function, and the method of stationary phase. The ambiguity function can be used as a polar display of the OTFs of a rectangularly separable incoherent optical system as a function of defocus, see section 2.5. Extended depth of field systems can be noticed almost by inspection of their corresponding ambiguity functions. The method of stationary phase, see section 2.6, allows the design of phase masks whose corresponding ambiguity functions have desired extended depth of field qualities.

3.2. System Design

The basic physical function of the cubic phase plate (CPP) in the optical digital system is to encode the incoherent wavefronts that are propagated through an optical system and then recorded by a CCD camera. A system diagram, Figure 3.1, shows the path that information about the object takes in such an optical/digital system. Wavefronts are transmitted from an object and encoded by the cubic phase plate. The optical system then acts as an information channel which low pass filters, aberrates, and sometimes defocuses the wavefront. A CCD camera acts as a receiver which samples the received wavefront, or "intermediate image". The signal processing portion of the system then decodes the digital information so that the final images can be displayed.

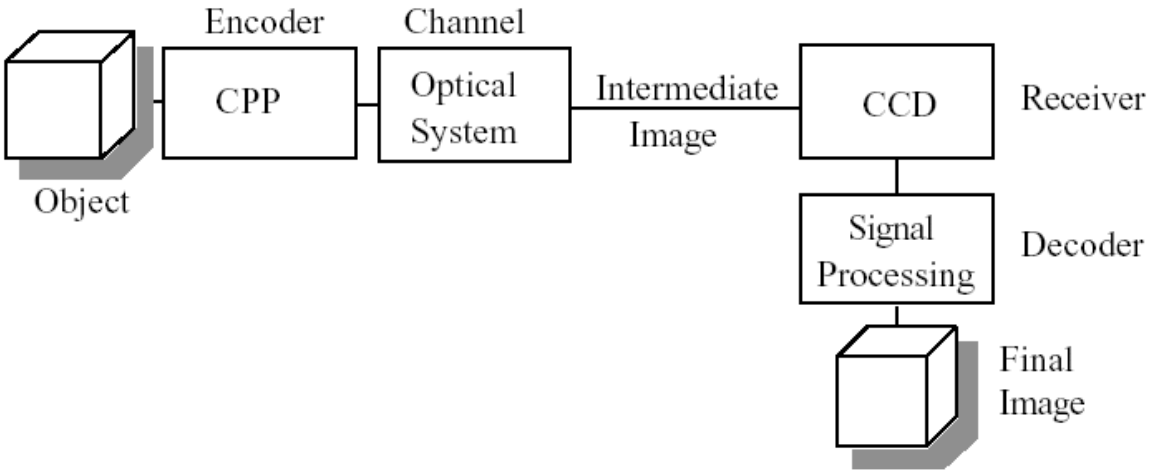


Figure 3.1. Information system view of Cubic Phase System

The general block diagram of wavefront coding imaging system is shown in figure 3.2. In this figure can be thought to "code" the image wavefront to make the sampled image insensitive to a defocus. The sample or coded imagery is blurred and requires digital processing to restore image clarity and sharpness. The digital processing component performs the removal of the image blur from the sampled image through linear and non-linear processing. Any type of digital processing to remove an image blur will amplify some spatial frequency information of the image as well as any additive noise. This noise amplification can affect the final image quality if the final noise value becomes too large. Non-linear processing can be used to remove the effect of some of this amplified additive noise. Algorithms such as block-Wiener filtering, median filtering, and similar can be applied to greatly reduce the effects of additive noise while maintaining the non-random spatial frequency details.

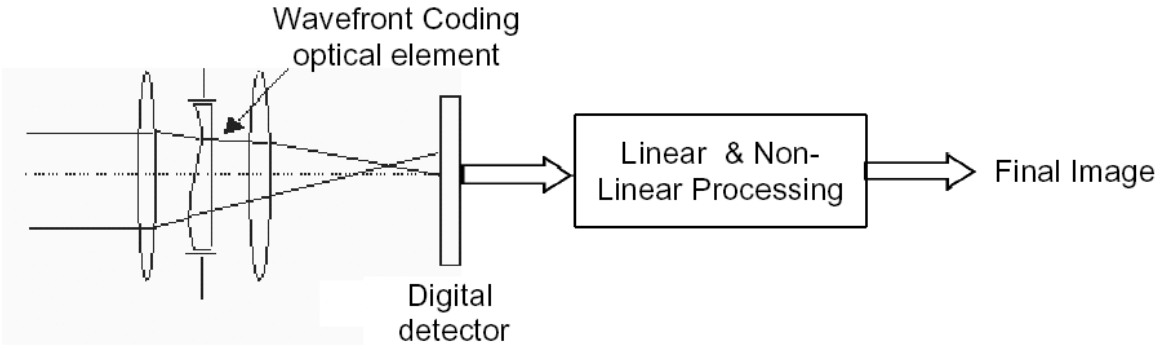


Figure 3.2. Block diagram of general wavefront coded imaging system.

3.3. Design of extended depth of field systems

Through the use of the ambiguity function and the method of stationary phase, phase masks for an extended depth of field incoherent optical system are readily found. The ambiguity function is an analytical tool that allows us to observe and to design OTFs for all values of

defocus at the same time, see section 2.5. The method of stationary phase provides the analytical flexibility needed to consider only phase masks in this design process.

Consider a one-dimensional unit power phase mask or phase function, in normalized coordinates, such as

$$P(x)=\begin{cases}\dfrac{1}{\sqrt{2}}e^{j\theta(x)} & \text{for } |x|\le 1\\ 0 & \text{otherwise}\end{cases}, \tag{3.1}$$

where θ(x) is some unspecified non-linear function. Knowledge of this phase function determines the PSF and OTF of the incoherent optical system for all values of defocus[17, 18]. We have assumed that a two-dimensional rectangularly separable phase mask will be used in practice. The one-dimensional OTF, as a function of defocus, is given by

$$H(u,\psi)=\int\left[P(x+u/2)e^{j\psi(x+u/2)^2}\right]\left[P^*(x-u/2)e^{-j\psi(x-u/2)^2}\right]dx\ , \tag{3.2}$$

with spatial frequency u and defocus parameter ψ. The symbol * denotes complex conjugate. The defocus parameter ψ is dependent on the physical lens size as well as the focus state.

$$\psi=\frac{\pi L^2}{4\lambda}\left(\frac{1}{f}-\frac{1}{d_o}-\frac{1}{d_i}\right)=\frac{2\pi}{\lambda}W_{020}=kW_{020}, \tag{3.3}$$

where L is the one-dimensional length of the lens aperture, and λ is the wavelength of the light. The distance d_o is measured between the object and the first principal plane of the lens, while d_i is the distance between the second principal plane and the image plane. The quantity f is the focal length of the lens. The wave number is given by k while the traditional defocus aberration constant is given by W_{020}. The traditional or Hopkins criteria for defocus[19] is equivalent to $\psi \approx 1$.

As we see in section 2.5, the ambiguity function related to this general mask can be used as a polar display of the OTF for all values of defocus. The ambiguity function of the mask $P(x)$ is given by eq. (2.14)

$$A(u,v)=\int P(x+u/2)P^*(x-u/2)e^{j2\pi vx}dx\,. \tag{3.4}$$

From (3.2) and (3.4) the ambiguity function can be shown to be related to the OTF of the system generated by $P(x)$ as

$$H(u,\psi)=A(u,u\psi/\pi). \tag{3.5}$$

Or, the projection of the point (u, $u\psi/\pi$) of the ambiguity function onto the horizontal u-axis yields the OTF for spatial frequency u and defocus ψ. In this way the two-dimensional ambiguity function can be used to determine the one-dimensional OTF for all values of defocus.

As an example of the utility of the ambiguity function approach to visualizing defocus OTFs, consider the standard rectangularly separable incoherent optical system. Such a system is formed with a rectangular pupil or mask function. Calculation of the magnitude of the ambiguity function of this one-dimensional rectangular function leads to the image shown in figure 3.3. In this image, regions of large power are given by dark shades. Notice that the majority of power in the ambiguity function of the rectangular aperture is concentrated along the $v = 0$ axis, which corresponds to the in focus OTF. The radial line in this figure has a slope of $\pi/2$.

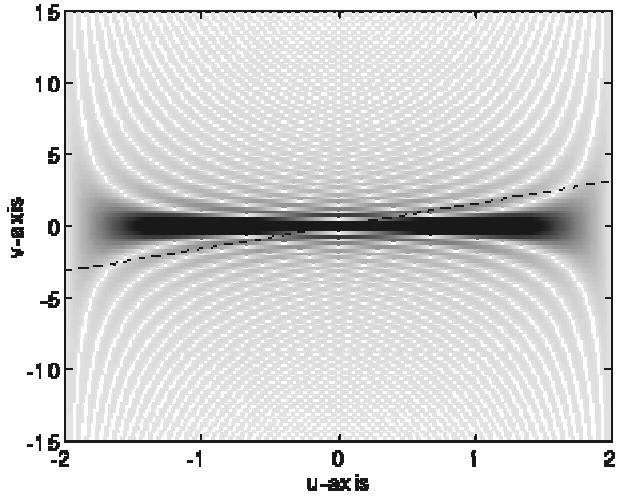


Figure 3.3. Ambiguity function of rectangular aperture. The radial line has a slope of $\pi/2$ corresponding to an OTF with defocus of $\psi = \pi^2/2$

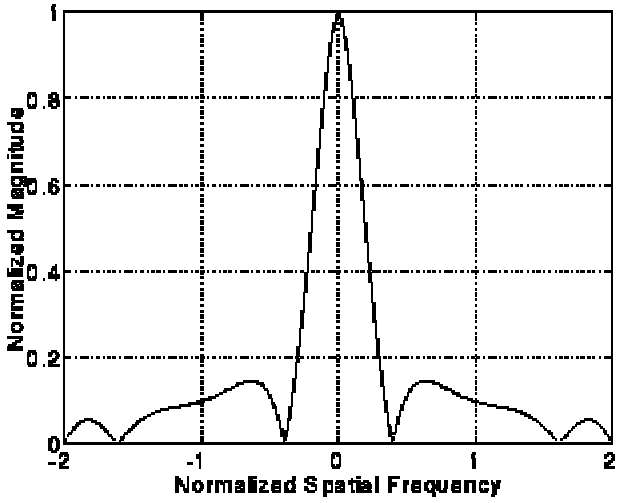


Figure 3.4. Defocus OTF of the standard optical system with defocus parameter of $\psi = \pi^2/2$

Figure 3.4 shows a defocused OTF related to the rectangular aperture, or standard optical system. The defocus parameter for this OTF is $\psi = \pi^2/2$. From (3.5), the ambiguity function of figure 3.3 along the radial line with a slope of $\pi/2$ describes this OTF. By inspection of these two figures we can confirm this relationship between the OTF and ambiguity function.

Extended depth of field systems, or systems that are insensitive to changes of focus, have ambiguity functions that are not a function of the second parameter, here given as *v*. From (3.5), ambiguity functions that are independent of the second parameter *v* lead to OTFs that are invariant to defocus ψ. In practice, extended depth of field systems are those with ambiguity functions approximately independent of *v* over a relatively wide angular region about the *u*-axis. From the ambiguity function of figure 3.3, we can immediately notice that a rectangular pupil function does not describe an extended depth of field system.

By careful selection of the non-linear function $\theta(x)$ of the general mask given in (3.1), a phase function that produces an ambiguity function with the desired extended depth of field characteristics can be found. We term this mask the *cubic-pm*, or *cubic phase modulation mask*. The next section describes this mask.

3.4. Stationary phase derivation of the cubic-pm OTF

Through the method of stationary phase applied to the ambiguity function we can find an ambiguity function, and associated phase function, that is independent of the second parameter, here called *v*. Such ambiguity functions define incoherent optical systems insensitive to defocus.

The ambiguity function of the general phase mask or function, given in (3.1), is

$$A(u,v) = \int_{-(1-|u|/2)}^{(1-|u|/2)} e^{j\theta(x+u/2)} e^{-j\theta(x-u/2)} e^{j2\pi vx} dx, \qquad |u| \le 2. \tag{3.6}$$

Let us assume that the non-linear function $\theta(x)$ is some single-term polynomial

$$\theta(x) = \alpha x^{\gamma}, \quad \gamma \neq \{0,1\}, \quad \alpha \neq 0. \tag{3.7}$$

This form of $\theta(x)$ will result in a mathematically tractable solution. We can then re-write (3.6) as

$$A(u,v) = \frac{1}{2} \int_{-(1-|u|/2)}^{(1-|u|/2)} e^{j\alpha(x+u/2)^{\gamma}} e^{-j\alpha(x-u/2)^{\gamma}} e^{j2\pi vx} dx, \qquad |u| \le 2, \tag{3.8}$$

or

$$A(u,v)=\frac{1}{2}\int_{-(1-|u|/2)}^{(1-|u|/2)} e^{j\vartheta(x)}e^{j2\pi vx}dx, \qquad |u|\le 2, \tag{3.9}$$

where

$$\vartheta(x)=\alpha\left[(x+u/2)^{\gamma}-(x-u/2)^{\gamma}\right]. \tag{3.10}$$

If the phase term $\vartheta(x)$ varies "fast enough", the above integral can be approximated through the stationary point of ($\vartheta(x)$+2π*vx*). The stationary phase approximation for $A(u,v)$ is given by

$$A(u,v)\approx\frac{1}{2}\sqrt{\frac{2\pi}{|\vartheta''(x_i)|}}e^{j\phi(v)}=\frac{1}{2}\sqrt{\left|\frac{\partial x_i}{\partial v}\right|}e^{j\phi(v)}, \tag{3.11}$$

where x_i is the stationary point and

$$\phi(v)=2\pi vx_i+\vartheta(x_i). \tag{3.12}$$

From (3.11) the *magnitude* of the ambiguity function will be independent of its second parameter *v* when the second derivative of $\vartheta(x_i)$ with respect to x_i is independent of *v*, or equivalently when the stationary point x_i is linear in *v*. In order to find the stationary point x_i, we can begin by taking the derivative of (3.12) and setting the result equal to zero. We obtain

$$\frac{\partial}{\partial x_i}\left(2\pi vx_i+\vartheta(x_i)\right)=0, \tag{3.13}$$

$$2\pi v+\gamma\alpha\left(x_i+u/2\right)^{\gamma-1}-\gamma\alpha\left(x_i-u/2\right)^{\gamma-1}=0. \tag{3.14}$$

We can show that the solution for x_i above, as a function of γ, will be linear in *v* as required *if and only if* γ = 3. The needed mask will then have a cubic phase profile. We term this "**cubic phase modulation**", or a **cubic-pm mask**. This cubic-pm function has a stationary point of

$$x_i=\frac{-\pi v}{3\alpha u}, \quad u\neq 0. \tag{3.15}$$

The stationary phase approximation to the magnitude of the ambiguity function of the cubic-pm system is then

$$\left|A(u,v)\right| \approx \frac{1}{2}\sqrt{\left|\frac{\partial x_i}{\partial v}\right|} = \sqrt{\frac{\pi}{12|\alpha u|}}, \qquad u \neq 0 \ . \tag{3.16}$$

Using (3.10), (3.12), (3.15), we can find that the phase term of this ambiguity function, $\phi(v)$ from (A6), is given by

$$\phi(v) \approx \frac{\alpha u^3}{4} - \frac{\pi^2 v^2}{3\alpha u}, \qquad u \neq 0 \ . \tag{3.17}$$

Combining both the magnitude and phase approximations we have

$$A(u,v) \approx \sqrt{\frac{\pi}{12|\alpha u|}} e^{j\frac{\alpha u^3}{4}} e^{-j\frac{\pi^2 v^2}{3\alpha u}}, \qquad u \neq 0 \, . \tag{3.18}$$

From (3.5), the resulting approximation to the OTF of the cubic-pm system is then given by

$$H(u,\psi) \approx \sqrt{\frac{\pi}{12|\alpha u|}} e^{j\frac{\alpha u^3}{4}} e^{-j\frac{\psi^2 u}{3\alpha}}, \qquad u \neq 0 \ . \tag{3.19}$$

The magnitude of the approximate OTF above is independent of the defocus parameter ψ. The phase approximation contains two terms, however. One term is independent of defocus, the other not. Specifically, the second of the phase terms, $e^{-j\frac{\psi^2 u}{3\alpha}}$, is a function of defocus ψ. Notice that this term is a linear phase term in u. Such a term has the effect of merely shifting the location of the resulting point spread function (PSF) with large defocus. Fortunately, this term can be controlled through the constant α. Large values of α, from (3.7), minimize the sensitivity of the cubic-pm system to movement of the PSF with defocus. In practice, this defocus dependent term can be effectively controlled so as to be negligible. The final approximation for the OTF is then

$$H(u,\psi) \approx \sqrt{\frac{\pi}{12|\alpha u|}} e^{j\frac{\alpha u^3}{4}}, \qquad \text{for large } |\alpha|, \quad u \neq 0 \, . \tag{3.20}$$

It is easy to show from (A1) that $H(0,\psi) = 1$. The stationary phase approximations are valid for large space-bandwidth product (SBP) functions[20, 21]. The definition of a "large" SBP is usually accepted to be greater than 100. With the general mask of (3.1), the spatial extent is 2. The bandwidth of this general mask is given by its maximum instantaneous frequency. Since instantaneous frequency is the derivative of phase, the bandwidth of the general mask is

$$BW = \max_x \frac{\partial}{\partial x}\theta(x) = \max_x \frac{\partial}{\partial x}\alpha x^3 = 3\alpha . \tag{3.21}$$

The SBP of the cubic-pm mask must then satisfy

$$SBP = 2(3\alpha) = 6\alpha >> 100 , \tag{3.22}$$

or approximately

$$\alpha >> 20 . \tag{3.23}$$

3.5. The cubic-pm phase mask

Modification of a standard incoherent optical system by a cubic-pm phase mask produces intermediate images that are insensitive to defocus. Conceptually simple filtering techniques applied to these intermediate images form a complete system that images with high resolution and large depth of field. The cubic-pm mask, in normalized coordinates, is given by

$$P(x) = \begin{cases} \frac{1}{\sqrt{2}} e^{j\alpha x^3} & \text{for } |x| \le 1, \quad |\alpha| >> 20 \\ 0 & \text{otherwise} \end{cases} , \tag{3.24}$$

where the constant α controls the phase deviation. The OTF of the incoherent system related to this function can be approximated as

$$H(u,\psi) \approx \begin{cases} \sqrt{\frac{\pi}{12|\alpha u|}} e^{j\frac{\alpha u^3}{4}}, & |\alpha| >> 20, \quad u \neq 0 \\ 1 & u = 0 \end{cases} . \tag{3.25}$$

The approximation of the OTF is independent of defocus. This can be inferred from the ambiguity function related to the cubic-pm mask, with α = 90, shown in figure 3.5. The cubic-pm ambiguity function has uniform non-zero values distributed about the u-axis. Radial lines through the origin of this ambiguity function have nearly the same values as a function of angle, for a broad range of angles. Hence, the cubic-pm mask should form an extended depth of field incoherent optical system. Figure 3.6 shows a comparison between the stationary phase approximation and the actual calculated OTF using (3.2). The smooth curve in this figure is the approximation of the magnitude of the OTF; the other is the calculated magnitude of the OTF. For this figure, the constant α of (3.24) was also selected as 90, while the defocus parameter ψ was set to 15. The approximation holds for other values of defocus as well as for the phase of the OTF. See figure 3.7. This figure is a plot of the magnitude of three defocused OTFs related to the cubic-pm mask, with α = 90. The

defocus values of the three OTFs are ψ = 0, 15, and 30. These OTFs are nearly constant with defocus and have no zeros. This is what makes it possible to use one focus independent digital filter to restore the intermediate image. Note the dramatic variation of the OTFs of the standard optical system, with the same defocus values, shown in figure 3.8. Also notice that the vertical scale in figure 3.8 is different from that of figure 3.7.

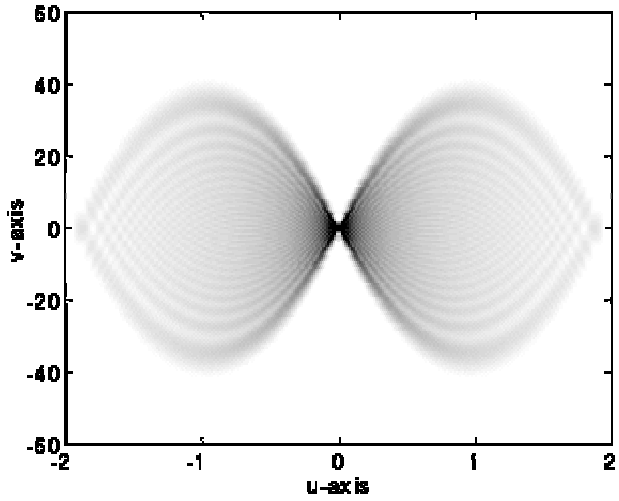


Figure 3.5. Magnitude of the ambiguity function of the cubic-pm function with α = 90. Notice that radial lines through this function are insensitive to angle, for a broad range of angles.

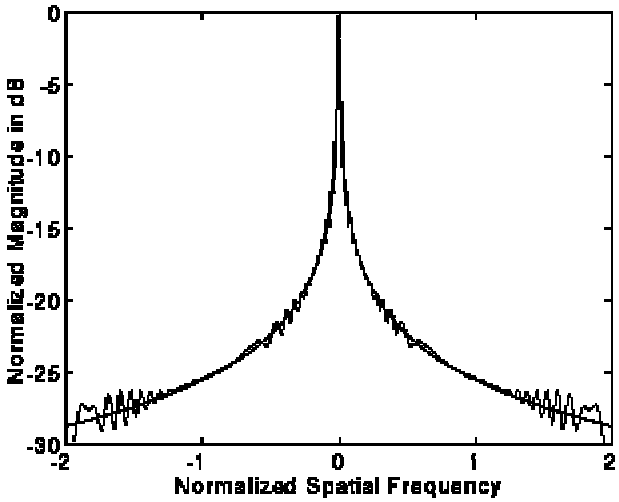


Figure 3.6. Magnitude of the OTF of the cubic-pm system with α = 90 and defocus of ψ = 15. The smooth curve is the stationary phase approximation of the OTF. The other curve is the calculate OTF.

3.6. Cubic phase plate

Although the defocus transfer function was originally used in optics as a tool for analyzing optical systems, it has also been used as a design tool. To extend the depth of focus, it

would be desirable to have an Defocus Transfer Function, DTF, that has projected traces, MTFs, that are invariant to defocus. A family of cubic phase pupil functions was found (3.24) that would produce such a defocus transfer function. In order to implement this pupil function, a specific phase plate was developed by Cathey and Dowski that, when correctly placed into an imaging system, causes the imaging system to have the desired invariant DTF, and therefore to have an extended depth of focus.

The family of cubic phase plates has a general two-dimensional phase delay as a function of spatial coordinates according to:

$$P(x, y) = \alpha\left(x^3 + y^3\right). \tag{3.26}$$

This gives a radian phase for a rectangularly separable phase plate with a unit amplitude square aperture. The spatial coordinates x and y are representative of the normalized distance from the center (origin) of the plate. The total phase deviation is controlled by the constant α and is designed for a particular wavelength using:

$$\alpha = \frac{2\pi\xi}{\lambda}, \tag{3.27}$$

where ξ is the total optical path difference introduced by the plate. A three-dimensional diagram of a general phase mask, with $\alpha = 1$, x, y normalized, is shown below in Figure 3.

The cubic phase plate is a transparent optical element with a cubic phase profile. It can be fabricated by grinding the surface on glass or plastic directly or, for high volume, an ultra precision mold insert can be machined. The cubic phase plates can then be mass produced with an injection molding process. However, the physical phase plate used by Cathey and Dowski in their experiments is a modulo 2π cubic function implemented in photo resist. It has a modulo 2π thickness corresponding to a two-dimensional function of spatial position as given in Eq. 3.26. The phase deviation for this phase plate was $\alpha = 20\pi$ for a design wavelength of 632.8 nm. The dimensions of the plate are 12 mm by 12 mm with a substrate thickness of 2 mm.

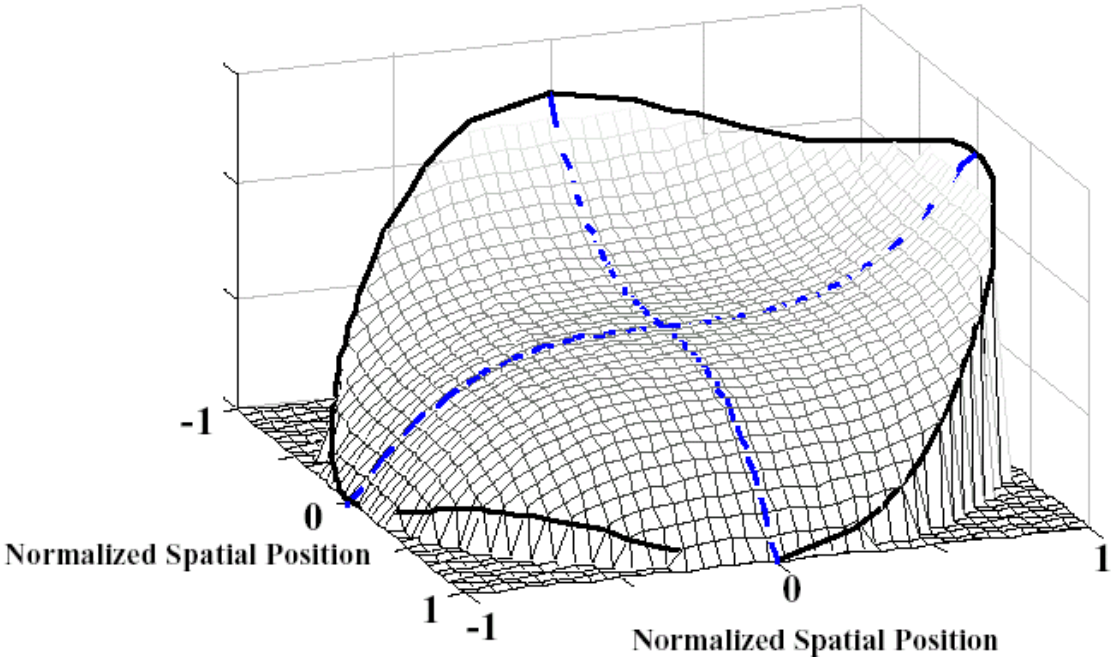


Figure 3.7. Three-dimensional diagram of general phase mask.

3.7. Evaluation of the optical/digital cubic-pm system

In order to illustrate the performance of the optical/digital cubic-pm system for extended depth of field imaging, Cathey and Dowski present two methods of comparison. These are the simulated measurement of the full width at half maximum amplitude (FWHM) of the PSF as a function of defocus, and simulated imaging of a spoke target at different defocus values. Comparison is made to the standard optical system in both cases.

Figure 3.8 illustrates the FWHM criterion applied to the standard optical system and cubic-pm optical/digital system. The width of the standard system, with no defocus, has been normalized to unity. The width of the PSF from the cubic-pm system, after focus independent digital filtering, is essentially constant out to the normalized defocus value of $\psi = 30$. From (3.3) we can show that the normalized defocus of $\psi = 30$ is nearly 29 times that of the Hopkins criterion for defocus where $W_{020} = \lambda/6$. As expected, the width of the PSF of the standard system greatly increases with defocus. The width of the unfiltered or intermediate PSF of the cubic-pm system would be much wider than that of the in focus PSF from the standard system.

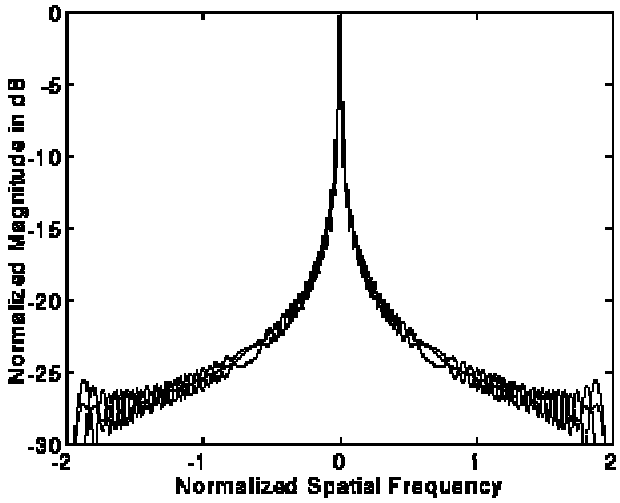


Figure 3.7. Magnitude of OTFs from the cubic-pm system with $\alpha = 90$ and defocus $\psi = 0$, 15, and 30

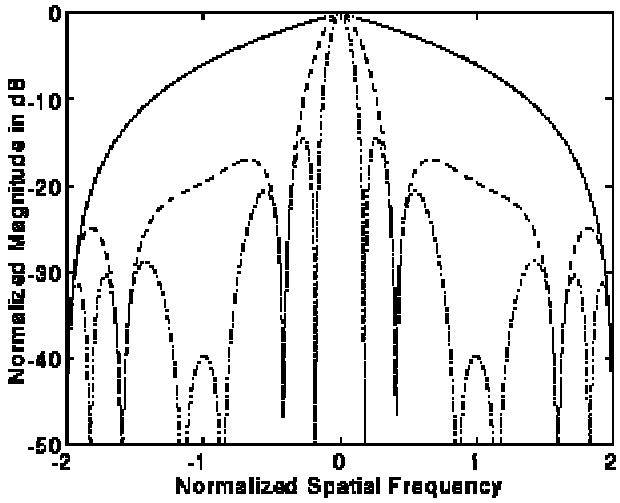


Figure 3.8. Magnitudes of OTFs from the standard optical system. The solid line denotes the OTF a defocus $\psi = 0$, the dashed lines is for $\psi = 15$, and the dashed-dotted line is for $\psi = 30$. Notice that the vertical scale is different than that of figure 13.

Figure 3.9 illustrates simulated imaging of a spoke target with the cubic-pm optical/digital system, along with a comparison of images from the standard optical system. The cubic-pm optical/digital system includes both the formation of the incoherent intermediate image as well as focus independent digital filtering of this image. Without digital filtering the intermediate images would be unrecognizable. The digital filter used for this example was a simple inverse filter that, when combined with the intermediate OTF of (3.25), resulted in a

triangular system OTF, in a least squares sense. The left column of this figure simulates imaging a spoke target with a standard optical system under varying defocus. The right column shows a simulation of the same imaging conditions using the cubic-pm optical/digital system. the cubic-pm system. The term "mild defocus" corresponds to $\psi = 5$, or about 5 times the Hopkins criterion for defocus. The term "extreme defocus" corresponds to $\psi = 30$ or about 29 times the Hopkins limit. The image of the spoke target from the standard system is severely degraded for even mild defocus. The images from the cubic-pm system are essentially constant with defocus while the image quality is nearly the same as that from the standard system with no defocus. Only a single digital filter is used for all values of defocus with the cubic-pm system. No single filter can be applied to the defocused images from the standard system to correct for the effects of defocus.

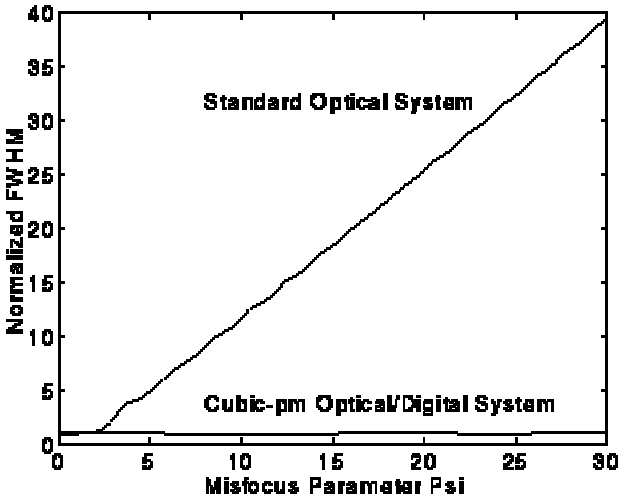


Figure 3.8. Normalized full width at half maximum amplitude (FWHM) of the PSF of the cubic-pm optical/digital system with comparison to that of the standard optical system

These simulations assumed a noise free optical/digital system. In practice, "restoration" of the intermediate image through digital filtering will alter the noise properties of the final image. As in other restorative schemes, a signal-to-noise ratio (SNR) or dynamic range premium is required at the image. Different filtering schemes require different SNR premiums. The simple inverse filtering used here requires the largest premium. Other more complex filtering schemes would require less. The least squares inverse filter used for the simulations of figures 3.8 and 3.9 has a transfer function which is given in figure 3.10. From (3.25) the phase of this filter is approximately cubic. The zero spatial frequency component of this filter is normalized to unity. With this filter, the maximum magnification of any spatial frequency component is approximately 20 dB. An exaggerated estimate of the required SNR premium for this simple filter is then approximately 20dB, or the required extra dynamic range would be approximately 3.5 bits.

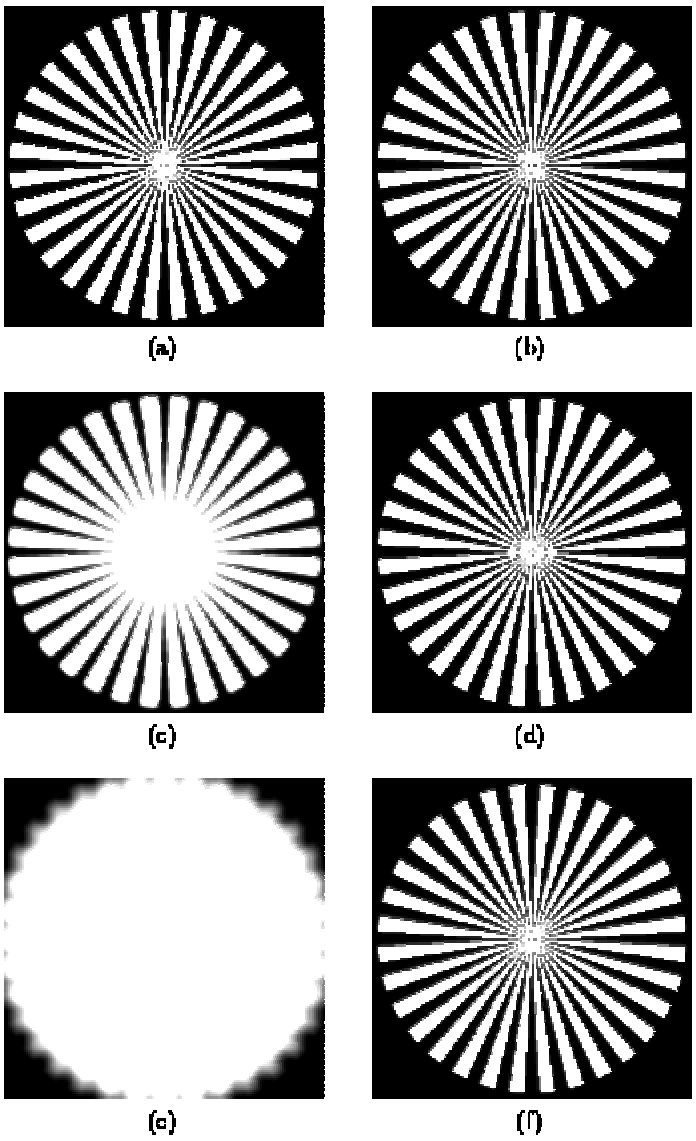


Figure 3.9. Simulated images of a spoke target from standard optical system (a, c, & e) and cubic-pm optical/digital system (b, d, & f). (a, b) Geometrically in focus, (c, d) mild defocus, and (e, f) extreme defocus.

An algorithm independent measure of the increase in performance of the cubic-pm optical/digital system over the standard system can be found from the Fisher[22] information of defocus. Fisher information is a measure used to describe the information content of a given signal pertaining to a certain parameter. For an ideal focus-invariant system the Fisher information of defocus would be zero. In other words, the ideal focus-invariant system would produce an image that contains no information pertaining to the focus state. Such an image would not be a function of defocus. A system whose OTF has a large

variation with defocus cannot employ a single focus-independent digital filter to correct for defocus. A focus-dependent digital filter can be used if the focus state is known *a priori*. Assume that a general incoherent system is imaging a point object, or one with a flat spatial frequency spectrum. We can show that the Fisher information of defocus from this assumed scenario is

$$J(\psi) = \int \left| \frac{d}{d\psi} H(u,\psi) \right|^2 du, \tag{3.28}$$

where $J(\psi)$ is the traditional notation for the Fisher information of the defocus parameter ψ, and $H(u, \psi)$ is the OTF.

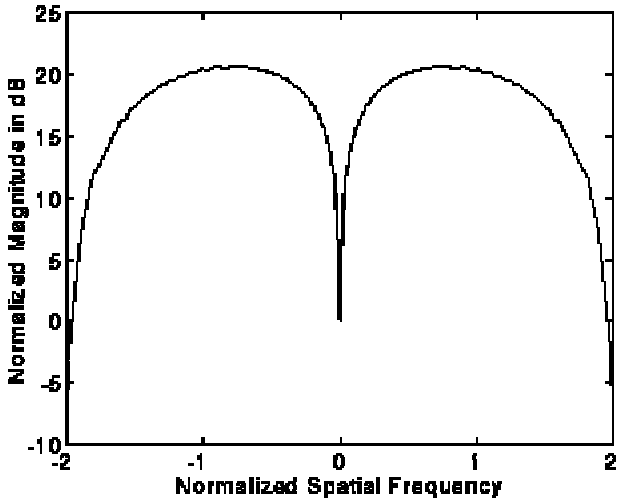


Figure 3.10. Magnitude of the digital filter transfer function used in simulations of the cubic-pm optical/digital system.

A ratio of the Fisher information related to the standard system over the Fisher information related to the cubic-pm system can be used as a measure of performance of the cubic-pm system. When this ratio is greater than unity, or 0 dB, the theoretical variation with defocus of the standard system exceeds that of the cubic-pm system. Again, the cubic-pm system was chosen with the constant α from (3.24) as equal to 90. This ratio of the Fisher information is given in figure 3.11. For example, the variation of the OTF of the standard system at defocus of $\psi = 10$ will, on average, be 20 dB larger than the variation of the OTF for the cubic-pm optical/digital system. Increasing the constant α of the cubic-pm system increases this difference in the variation of the OTF; decreasing α decreases the difference. The defocus value where the Fisher information of defocus is equal for the standard and cubic-pm system is monotonically related to the parameter α. Other methods of

characterizing the performance of the cubic-pm optical/digital system are currently under investigation.

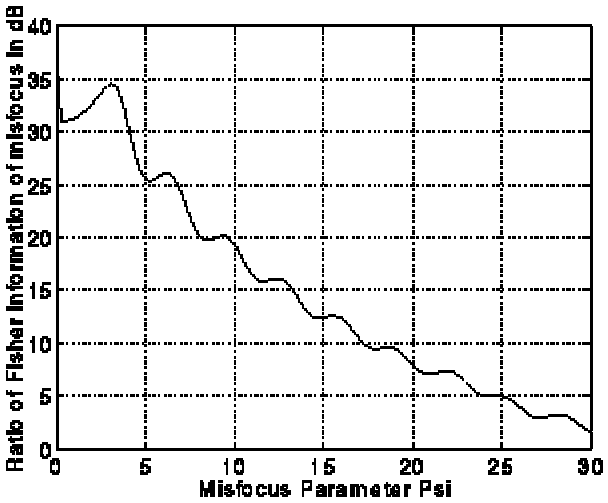


Figure 3.11. Ratio of the Fisher information of defocus, assuming a point object, of the standard optical system over the Fisher information of defocus for the cubic-pm optical/digital system.

3.8. Signal Processing

The objective of the signal processing of the focus invariant imaging system is to decode the image, thus removing some of the effects of the cubic phase plate and producing a normal resolution, in-focus image.

Invariance and lack of zeros in the modulation transfer functions allow us to use a single, simple, inverse filter on an entire image to decode the effects of the phase plate. The filter will cancel the phase effects of the mask by multiplying the OTF of the image by a conjugate phase. In all the reports about experiments with this technique, the signal processing step was performed digitally in software. Two-dimensional frequency domain multiplication filtering was used.

The mathematics for the filter are simple variations on the least-mean-squares filtering equations which now we explain.

Least mean squared error filter derivation starts with the following question: If we had the best possible filter, f, applied to our data vector, h, how closely does the resulting vector resemble the ideal diffraction limited data vector, g[23, 24]?

Applying the filter, f, to the data, h, requires a convolution procedure which can be algebraically complicated. The data, h, can be restructured, however, into a convolution matrix H, so that a simple matrix multiplication, Hf, can be employed to find the filtered data matrix. The difference between this filtered data matrix, Hf, and the diffraction-limited

data matrix, g, is an error signal. In order to evaluate this error signal it is best to find its total power. Since the square of the operator norm of a signal vector gives the total power of that signal, we can use it to evaluate the total power in the error signal.

$$\|e\|^2 = \|Hf - g\|^2. \tag{3.29}$$

For vector signals, this norm can be rewritten as the product of the conjugate of the vector times the vector itself:

$$\|e\|^2 = (Hf - g)^*(Hf - g). \tag{3.30}$$

Distributing the conjugate and multiplying out the vectors gives:

$$\|e\|^2 = \left((Hf)^* - g^*\right)(Hf - g) = f^*H^*Hf - g^*Hf - f^*H^*g + g^*g. \tag{3.31}$$

The next step in least mean square filter derivation is to minimize the power of the error signal. Vector differentiation of eq. (3.31) with respect to the vector of filter values is expressed mathematically by

$$\frac{d\|e\|^2}{f^*} = H^*Hf + \left(f^*H^*H\right)^* - \left(g^*H\right)^* - H^*g. \tag{3.32}$$

Next we set this derivative to zero to find the minimum value and rearrange to give

$$H^*Hf + H^*Hf - H^*g - H^*g = 0, \tag{3.33}$$

$$2H^*Hf - 2H^*g = 0. \tag{3.34}$$

The final step is to rearrange and solve for the filter value vector, f. This yields

$$f = \left(H^*H\right)^{-1}H^*g. \tag{3.35}$$

There are three ways of using these equations.

a) To find a one-dimensional spatial domain filter, f, H must be a convolution matrix of an average of several sets of one-dimensional coded PSF data, and g would be an ideal, diffraction limited 1-D PSF, i.e. a spike.

b) To find a one-dimensional frequency domain filter, one must find the one-dimensional Fourier transform (OTF) of each of the one-dimensional PSFs. Samples of each OTF at each frequency are then used to form a convolution matrix, H, (i.e. one convolution matrix containing the spatial frequency value, v, of each

PSF for all possible values of *v*). Then the value at the frequency, *v*, of the ideal, diffraction limited, one-dimensional OTF is used as the scalar value *g*. The output is the value of the frequency domain filter at frequency, *v*.

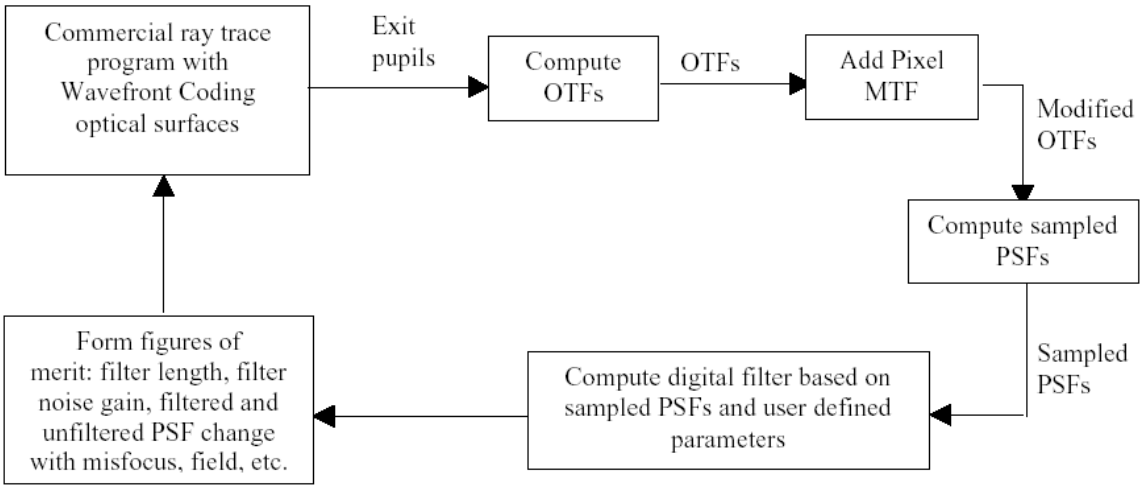
c) To find a two-dimensional spatial domain filter, the columns of the average of several sets of two-dimensional coded PSF data must be concatenated to form a very long column. The same must also be done with the columns of the 2-D ideal PSF data. Finally, to find a two-dimensional frequency domain filter, the coded PSF data sets are Fourier transformed. Then the value of each transformed data set at a particular frequency is put into a row vector, *h*. The desired value (in frequency domain) of *g* is also used in the equation to find the value of the filter at that particular frequency. This process is performed for each frequency value until all the points in the frequency domain filter data set are found. The entire process, however, can be quickly performed in matrix form with very simple code using a matrix oriented programming language like MATHLAB software.

The filter must include a frequency cutoff point above which all the frequency values of the filter are set to zero to compensate for the problem of noise. Additional filtering methods such as wavelet decomposition or Wiener filter denoising can also be used on the images both before or after decoding.

3.9. Design of wavefront coding systems.

Wavefront Coded imaging systems consist of non-traditional optical designs and digital signal processing of the resulting images. The signal processing used is dependent on the specific optical system. The Wavefront Coded optics are dependent on the type and amount of signal processing to be used. Since the optics and signal processing are closely coupled, it is natural to expect best performance from systems where the optical and digital components of the system are jointly optimized during design. Such a design is not directly possible with current commercial optical design tools. CDM Optics has developed custom software design tools that work in conjunction with commercial ray tracing software to provide optimized optical/digital Wavefront Coded system designs. The optical components are designed to minimize the changes or sensitivity of the optics to defocus effects as well enable efficient signal processing. The digital components are designed to minimize algorithm complexity, processing time, and effects of digital processing on image noise.

The base routine for Wavefront Coding design is a commercial ray-trace program that traces rays through typical spherical and aspherical surfaces as well as general Wavefront Coding surface forms. The commercial ray-trace program is used as the main user interface and to calculate exit pupils and optimize a given set of optical and digital merit functions or operands. A block diagram of general Wavefront Coding design process is shown below. The output of this design consists of 1) traditional optical surfaces, materials, thickness, and air spaces, 2) parameters of wavefront coding surfaces, and 3) digital filter coefficients.



Block diagram of CDM Optics design process for Wavefront Coded imaging systems. A commercial ray trace program is used as the main user interface, to calculate exit pupil OPDs, and to optimize a given set of operands. Special surface forms are used to model Wavefront Coded optical surfaces. CDM Optics' software forms OTFs and PSFs related to pixel sampled signals, digital filters, and digital processing figures of merit or operands. Digital processing figures of merit include digital filter noise gain (RMS value of digital filter), unfiltered and filtered PSF change with defocus, change with field, etc.

The general optical/digital design loop as shown in block diagram can be described as follows:

a. Start with given optical surfaces, thickness, and operating conditions (wavelengths, field of view, temperature range, sample object images, etc.).

b. Generate exit pupil optical path differences (OPDs) via traditional ray tracing.

c. Calculate OTFs

d. Include pixel OTF related to detector geometry

e. Calculate sampled OTFs and PSFs

f. Calculate digital filter coefficients for chosen processing algorithm based on sampled PSF set.

g. Form figures of merit (or Wavefront Coding operands) that are based on minimizing:

I. changes of the sampled PSF and MTF through focus, with field angle, with color, due to temperature changes, due to aliasing, etc.

II. digital processing parameters such as amount of processing, form of the processing, processing related image noise, digital filter noise gain etc.

h. Combine Wavefront Coding operands with traditional optical operands (Seidel Wavefront aberrations, RMS Wavefront errors, etc.) into optimization routines and modify optical surfaces.

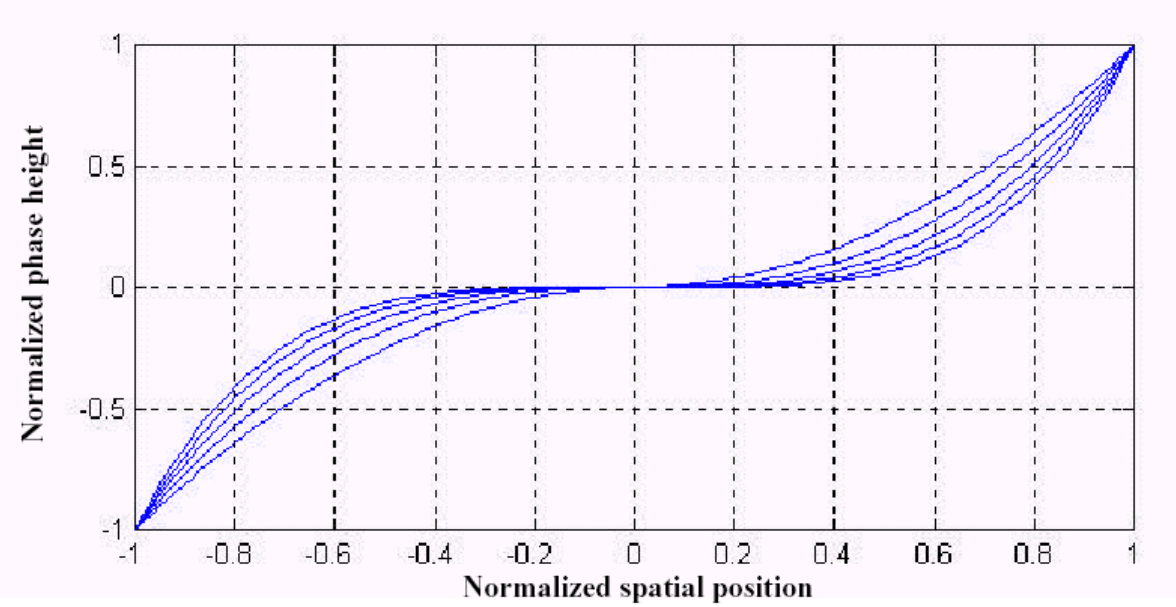
i. Return to step 2.

Theoretically calculated Wavefront Coding surface forms are used as starting points for the optical optimization. One general family of rectangularly separable surface forms is given in normalized coordinates as:

$$S(x) = |\beta| sign(x) |x|^{\alpha}$$

where $sign(x) = +1$ for $x > 0$, and $sign(x) = -1$ for $x \le 0$

The exponential parameter a controls the height of the MTF over a range of defocus, and the parameter β controls the sensitivity to defocus. In general, increasing the parameter β decreases the sensitivity to defocus while decreasing the height of the MTF and increasing the length of the resulting PSF. Next graph gives a two dimensional representation of this general family of rectangularly separable Wavefront Coding surface forms.



3.10. Realizations of focus invariance in optical/digital systems with wavefront coding

Sara Bradburn[25,] Cathey and Dowski report experimental verifications of an extended depth of focus system with near diffraction limited performance capabilities.

The physical phase plate used by Sara Bradburn , Cathey and Dowski in their experiments is a modulo 2π cubic function implemented in photo resist. It has a modulo 2π thickness corresponding to a two-dimensional function of spatial position as given in Eq. 3.26. The phase deviation for this phase plate was $\alpha = 20\pi$ for a design wavelength of 632.8 nm. The dimensions of the plate are 12 mm by 12 mm with a substrate thickness of 2 mm.

The new MTFs produced with this cubic phase plate (CPP) focus-invariant system do not resemble an MTF for an in-focus standard optical system. It is important, however, to examine a few of these MTFs for different defocus values and compare them to those of a standard system with same values of defocus. See Figure 3.12 which shows experimentally measured in-focus and defocused MTFs for both the standard and cubic phase plate focus-invariant systems. The MTFs in Figure 3.12 correspond to zero and 1.6 waves of defocus for each system.

At the higher spatial frequencies, the invariance of the MTFs decreases slightly. However, the signal processing portion will attenuate these differences. Plots of the standard in-focus and defocused MTFs and focus-invariant MTFs after signal processing are shown in Figure 3.13.

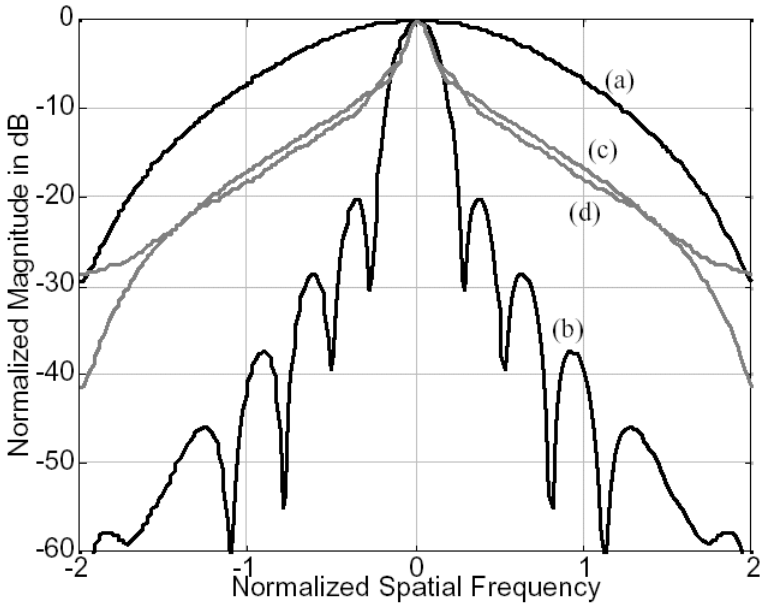


Figure 3.12. Measured MTFs of standard system (a) in-focus and (b) with 1.6 waves of defocus and MTFs of focus invariant system (c) in-focus and (d) with 1.6 waves of defocus. The difference between the MTFs of the focus invariant systems at higher frequencies is reduced in the signal processing step as shown in Figure 3.13.

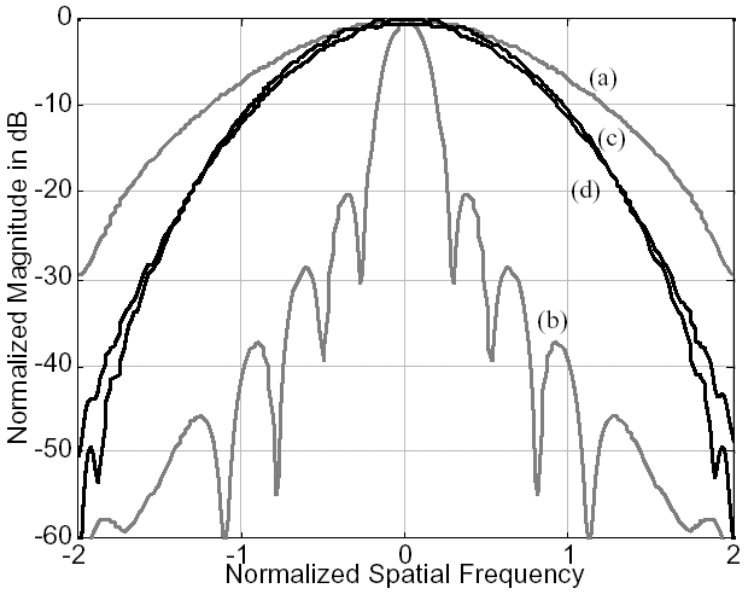


Figure 3.13. Measured in-focus and defocused MTFs of standard, (a) and (b), and focus invariant system, (c) and (d), after signal processing. Note the attenuation at higher frequencies in the filtered focus invariant MTFs. This helps to eliminate noise magnification in the inverse filtering process.

Figure 3.14 shows the PSFs corresponding to the MTFs in Figures 3.12 and 3.13 to demonstrate the increased amount of focus depth of the CPP system over the standard system. Figures 3.14(a) and 3.14(b) show the experimentally acquired PSFs for a standard system in focus (20 ms exposure time) and at 1.6 waves of defocus. Figure 3.14(b) was taken with longer exposure (115 ms) in order to get a brighter image of the defocused PSF. Figures 3.14(c) and 3.14(d) show the PSFs of the focus invariant system in focus and at 1.6 waves of defocus (both taken with 20 ms exposure time). Recall that the phase portion of the one-dimensional OTFs of the above coded PSFs, 3.14(c) and 3.14(d), would be cubic functions.

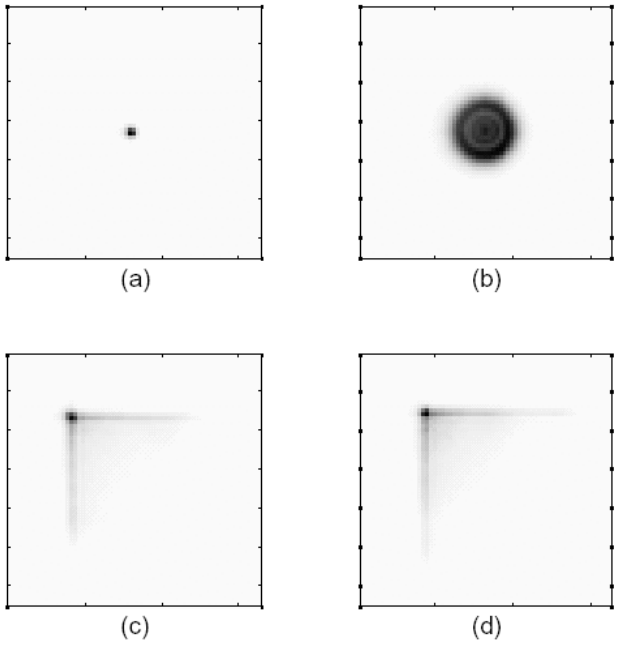


Figure 3.14. Measured PSFs of standard imaging system, (a) in focus, (b) with 1.6 waves of defocus (and increased exposure for visibility); and of a focus-invariant imaging system (c) in focus and (d) with 1.6 waves of

3.10.1. CCD: Signal to noise ratio and sampling effects

As shown in Figures 3.12 and 3.13, the MTF for the extended depth of focus (EDF) system is reduced at the higher spatial frequencies (from the standard in-focus MTF). It is therefore quite important to look at the resulting signal to noise ratio changes.

Consider a frequency where the MTF of the cubic phase plate system is reduced by 10X from the MTF of the standard in-focus system. The signal to noise ratio (SNR) is therefore also down by at most 10X . Because the digital camera used in these experiments has a fairly good dynamic range, (Kodak Megaplus Camera, Model 1.6, 10 bit - 72 dB dynamic range), the 10X loss in signal at a certain frequency does not represent an insurmountable problem. The systems used in these experiments are shot noise limited. The SNR of the system can be calculated by[26],

$$\frac{S}{N} = \frac{P_s \eta}{2h\nu\Delta\nu}\ , \tag{3.29}$$

where P_s refers to the incident optical signal power, η is the quantum efficiency, ν is the optical frequency, and h is Planck's constant.

Because $\frac{P_s}{h\nu}$ is in photons/second, $\frac{P_s\eta}{h\nu}$ is therefore in electrons/second. Realizing that $\Delta\nu$ is in units of $\sec^{-1}$ (bandwidth), and is inversely related to the integration time, we find that the SNR is directly related to the number of electrons in the CCD well divided by two, or:

$$\frac{S}{N} = \frac{\#\,\text{electrons}}{2}. \tag{3.30}$$

Therefore we find that if we have taken a factor of 10X reduction in SNR ratio, then increasing the exposure time by 10, or the aperture diameter by $\sqrt{10}$, and allowing 10X more electrons through will restore the original SNR value for that particular frequency. In practice, however, increasing the exposure time does not usually prove to be a feasible solution for restoring the SNR if the exposure is already at a level that fills the wells of the CCD. In the case of an object consisting of an isolated point of light, when the cubic phase plate codes this point, the energy of the PSF will be spread out over a larger area, thus decreasing the maximum value. The exposure time can then be increased somewhat, in order to increase the number of photons allowed through the system and therefore increase the SNR. In the case of a more uniform image, however, when the points on the object are coded by the cubic phase plate, the coded individual points overlap and add up. Therefore, there can still be several areas in an image which will fill the CCD wells with the maximum number of electrons. This prevents the possibility of increasing the exposure time to restore the SNR. In the experiments that have been performed, the exposure time has never been increased. It is very important to realize, however, that the change in SNR is only a loss when compared to the in-focus, standard MTF. When one compares the defocused standard MTFs to the defocused EDF MTFs, the increase in SNR when using the EDF system becomes apparent. In fact, there are nulls in the standard defocused system MTFs which translate into a SNR of zero. At these frequencies, the EDF system has a better SNR due to the fact that there are no nulls in its MTF. The benefits of the cubic phase plate for defocused systems therefore outweigh the loss of SNR in the in-focus MTF.

The sampling effects of the CCD are also important to understand. Assuming100% fill factor, the pixel sizes of the CCD will determine the sample rate of the image capturing process. Given the size of the pixels in the CCD, and low magnification values used in these experiments, the image is undersampled in the sense of Shannon's sampling theorem. When used with diffraction-limited optical systems, the CCD camera determines the spatial frequency cutoff of the entire system. This spatial frequency cutoff will be the limiting factor in the resolution of the output images. In the future, additional experiments implementing optical systems of higher magnification will be performed. The spatial frequency cutoff will then be set by either diffraction, noise, or digital post-processing procedures instead of CCD pixel size.

When the signal processing is performed on the coded image, simple inverse filtering can magnify the system noise, especially in areas of reduced SNR, and produce a very noisy image. The solution for this problem lies in a modification of the filter used in the signal

processing step. If a spatial frequency cutoff is chosen above which the filter frequencies are set to zero, then the spatial frequencies of the filtered image which are greater than the spatial frequency cutoff, including noise, will be set to zero[28, 29].Another solution is to use a window (i.e. a Kaiser or a Gaussian window) with a similar spatial frequency cutoff. When such a window is placed over the frequency data, the higher frequency values will be highly attenuated instead of eliminated. If this spatial frequency cutoff is chosen carefully, images are cleaned up significantly without drastically affecting the image quality. Many algorithms and filters (such a Wiener filters) have been designed that will help find the optimum tradeoff between bandwidth (spatial frequency cutoff value) and noise amplification[23, 24, 27]. It is also important to remember that the dynamic range of many modern CCD cameras will allow for some loss of SNR at higher frequencies without severely impairing the inverse filtering process. The better dynamic ranges allow movement of the cutoff frequency to a higher value to produce images with better resolution.

3.10.2. Experimental Procedure and setup

The experiments performed using this cubic phase plate optical/digital system were generally set up as shown in Figure 3.15.

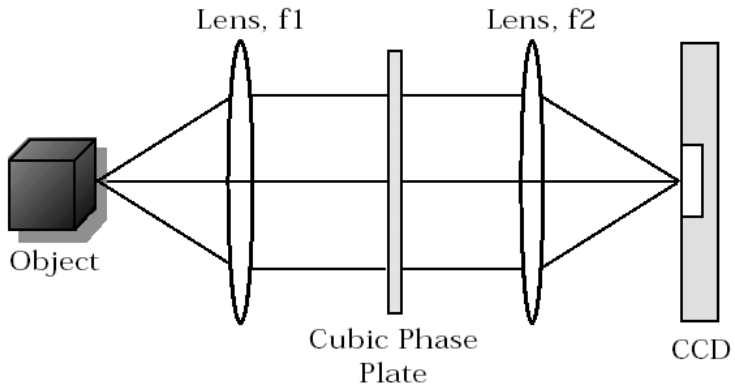


Figure 3.15. General experimental setup of Focus Invariant Optical/Digital System. Cubic phase plate at aperture stop of two lens system will code the wavefront.

Objects were illuminated with red light of 100 nm. spectral bandwidth. This bandwidth is close to the maximum bandwidth for which the modulo 2p cubic phase plate will perform well in a system using achromatic lenses. A newer continuous cubic phase plate has also been developed which can perform extended depth of focus imaging using the entire visible spectrum. Two achromatic lenses were used and modeled as a two-lens system. We then placed the cubic phase plate at the aperture stop of the two-lens system. The cubic phase plate can also be placed at the entrance or exit pupils of the optical system without changing its coding effect. We chose the aperture stop position, however, to simplify setup and alignment procedures.

3.10.3. Image collection procedure

The first part of these experiments, after setup and alignment of the optical system, was to collect images of the PSFs of the system at different values of defocus (including in focus). The standard system is the previously designed optical system with a clear aperture the same size and shape as the cubic phase plate aperture. The defocused PSFs were obtained by moving the CCD camera either in front of or behind the back focal plane. The clear aperture was then replaced by the cubic phase plate and the coded PSFs taken at the same in-focus and defocused positions of the camera. These coded PSFs are invariant over the necessary range as can be observed in both the spatial domain and spatial frequency domain. The coded PSFs were then used to design the filter to be used in the signal processing step. An important test of the subsequent filter is to filter a coded image of a point object. This should give a clean near-diffraction-limited PSF image. Next, a three dimensional object was imaged in a standard imaging system. The resulting image has the expected small in-focus properties in the focal range and the rest of the image will appear out of focus. This image is useful for qualitative comparison purposes and for determining the quantity of focal depth extension. A second reference image was obtained with a stopped down aperture but with increased exposure time in order to obtain an image with the same exposure as the original system image. When the cubic phase plate was used, the intermediate image (before signal processing) appeared evenly "defocused" for all in-focus and defocused areas of the object. This is a result of the coding effects of the phase plate. The intermediate image was then digitally decoded using the previously designed filter to provide the final image.

3.10.4. Focus invariant imaging system: results

A square wave target was used as the object to obtain an initial set of data for analyzing the extended depth of focus system. The square wave target with one line pair per millimeter was tilted at an angle of approximately 60° from the vertical axis. The image on the left in Fig. 3.16(a) shows the resulting image when using a traditional optical F/6.3 system and 100 ms exposure. The image on the right, 11(c), is the final decoded image obtained using the extended depth of focus system with the same F# and exposure time. The plot in Fig 3.16(b) shows traces down the center of the two images, demonstrating the effects of defocus and the improvement when using the extended depth of focus (EDF) system . Note that in the trace through the standard system image (darker line) one can observe the phase change that takes place due to defocus. In contrast, this phase change does not occur in the trace through the EDF image (lighter line).

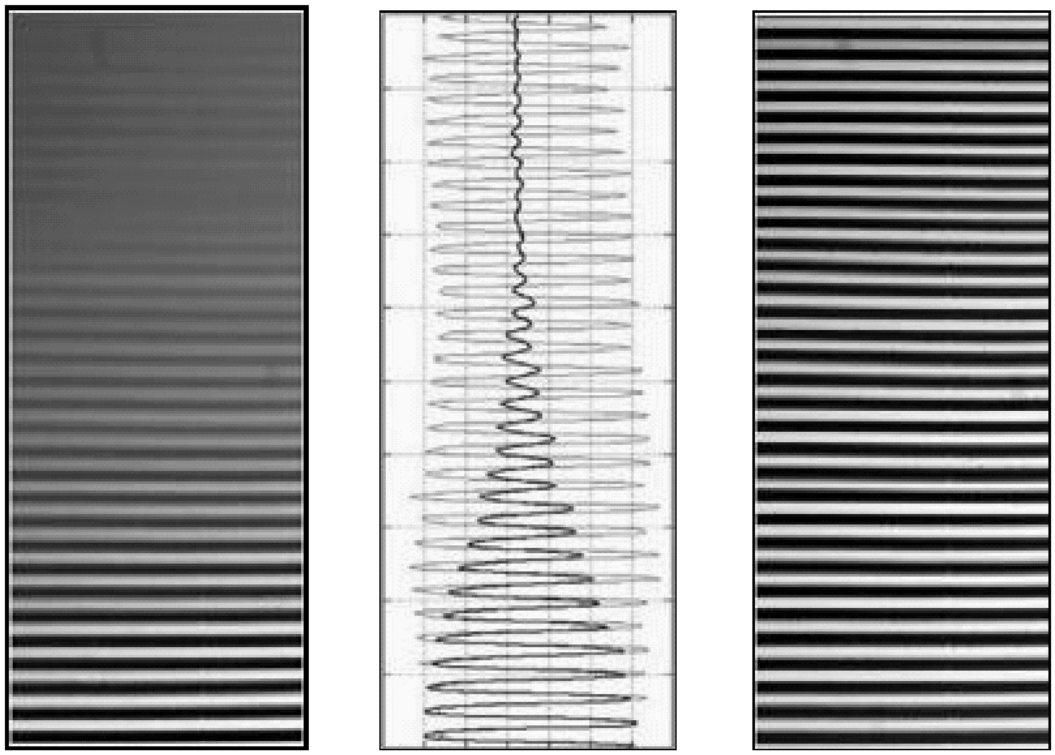


Figure 3.16. Square Wave Imaging with Standard (a) and Extended Depth of Focus (c) Systems. The traces of image brightness for the two cases are shown in (b). The phase shift due to defocus in the standard image trace does not occur in the Extended Depth of Focus image trace.

Images of gray scale objects are shown in Fig. 3.17. Small shells, each approximately 5 mm in diameter, were placed on a tilted surface to produce a highly three-dimensional object. The image in 3.17(a) was taken with a standard imaging system with F/10.5 and 200 ms exposure time. The focal plane of the this system lies at the front of this image so that the defocus can become more severe as the object slopes out of the focal range. The image in 3.17(b) was taken with an EDF imaging system, also with F/10.5 and 200 ms exposure time. Note that the defocus is invariant throughout the image. Image 3.17(c) is from a standard system with a reduced aperture (F/52.5) and approximately 5000 ms exposure time. The EDF imaging system output 12(d) has the same depth of focus as the small aperture system but the F# is still F/10.5 and exposure time was 200 ms. The expected resolution improvement of the F/10.5 EDF system over the F/52.5 small aperture system is limited due to factors such as undersampling, noise, system aberrations, and inverse filtering. The resolution improvement that does occur, however, is difficult to observe in these images due to the large number of pixels in each image. A black and white object, such as a plane of text, also helps demonstrate the power of the extended depth of focus cubic phase plate system. For this experiment a sheet of reinforced paper containing text

was placed in the focal plane and tilted approximately 60° to achieve a three dimensional effect. Figure 3.18 shows the standard and EDF system images for such an object. Once again, the image in 3.18(a) was taken with a standard system. The image in 3.18(b) is the intermediate image from the CPP system. The image in 3.18(c) is from the reduced aperture, standard system and the image in 3.18(d) is from the filtered CPP system. Some of the rough grainy look in the filtered CPP image 3.18(d) is due to the roughness of the paper upon which the text was printed.

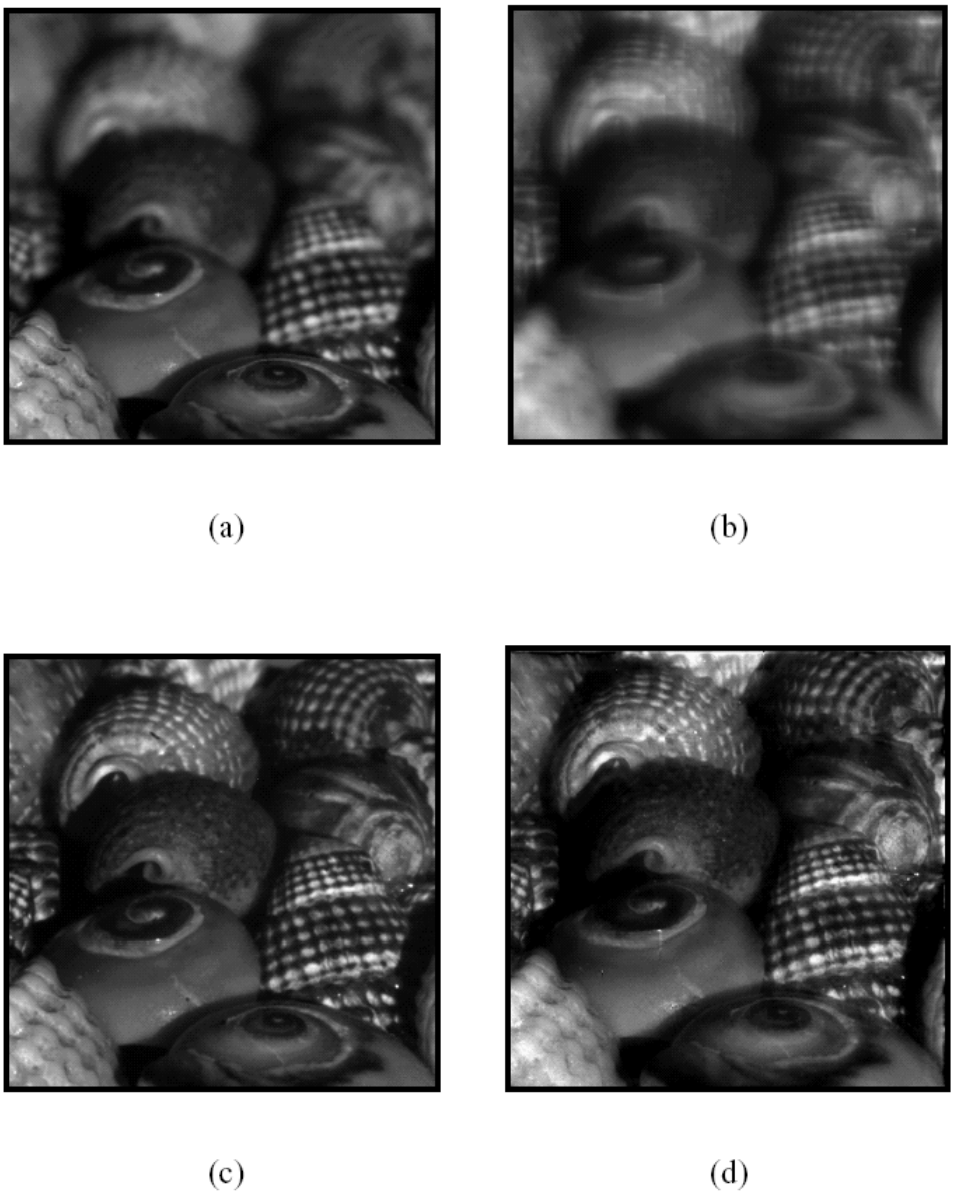


Figure 3.17. Images of Small Shells. (a) Standard imaging system with F/10.5, (b) Intermediate image from EDF imaging system with F/10.5, (c) Reduced aperture standard imaging system with F/52.5, and (d) EDF imaging system (F/10.5) image after signal processing.

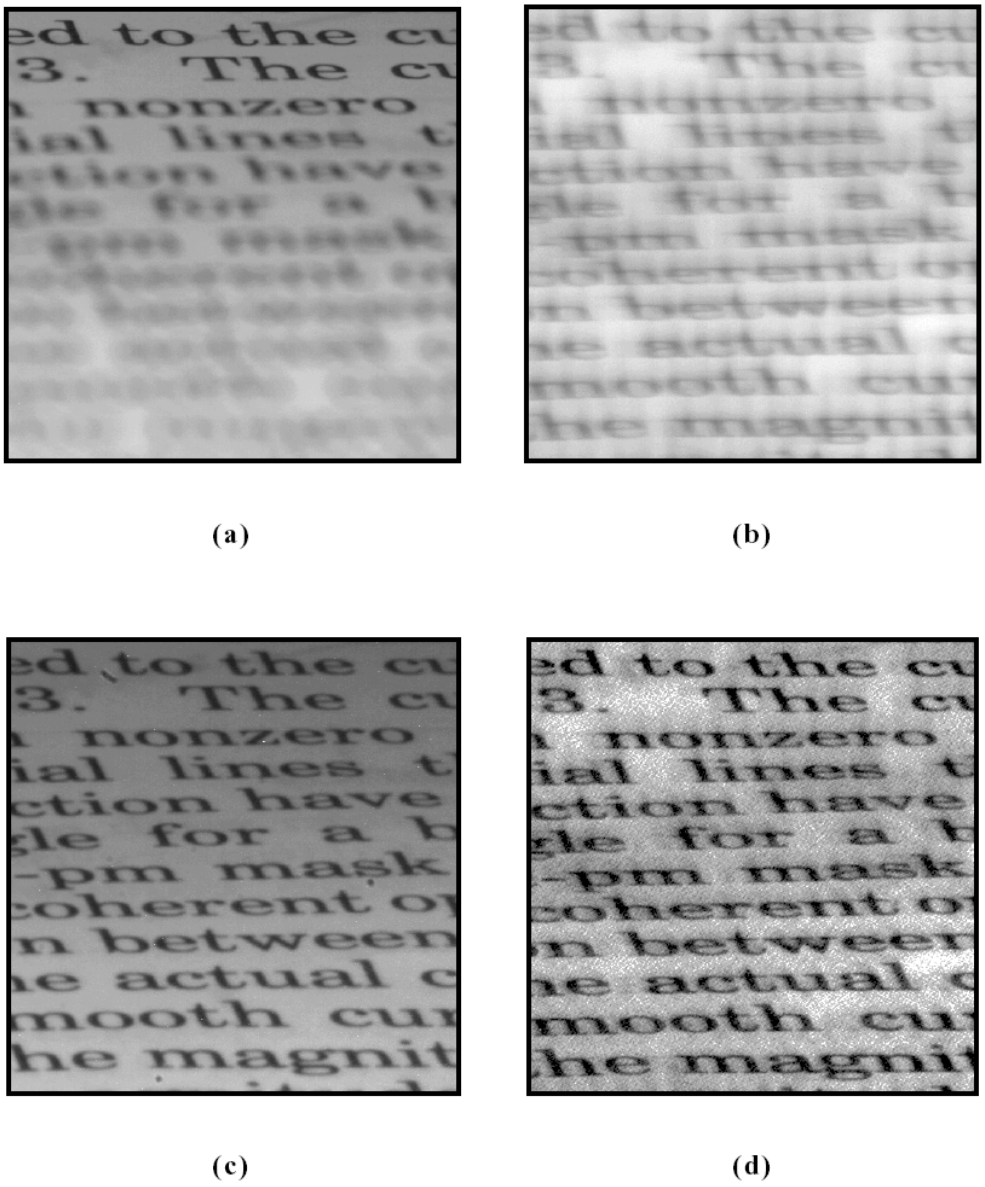


Figure 3.18. Images of Inclined Text. (a) Standard imaging system with F/6.3, (b) Intermediate image from EDF imaging system with F/6.3, (c) Reduced aperture standard imaging system with F/37.5, and (d) EDF imaging system image after signal processing.

3.11. Example applications

Wavefront Coding for extending the depth of field can add value to imaging applications where traditional methodologies (i.e. stopping down the aperture) are generally unacceptable. Constraints on illumination levels, exposure times, or spatial resolution often limit the application of previous optical methods. By using Wavefront Coding, applications such as machine vision or inspection, biometric analysis, and medical or microscopic imaging can enjoy fewer defocus-related problems, without sacrificing exposure times or requiring vast quantities of illumination. In a confined environment such as medical devices

and microscopes operate in, illumination is of great concern both in terms of delivery and safe power levels. In applications of machine vision or quality control inspection, exposure times must keep pace with the production line. In biometric analysis (often used for identification and security) reliable image quality must be maintained to a high degree. Wavefront Coding can supply all these applications with a viable solution to defocus related problems, without increasing the overall cost of the system.

Below we provide examples of Wavefront Coding applied to i) machine vision, and ii) biometrics. These two examples have been chosen since they can be seen as representing two extremes, respectively: imaging machine-generated information (i.e. labels, optical character recognition, barcodes, etc.) and the imaging of highly uncontrolled objects such as greasy fingerprints or retinal scans.

Fig. 3.19 gives six images comparing the performance of a traditional imaging system (a-c) to a system using Wavefront Coding (d-f), for a generic binary input image. The horizontal, vertical, and diagonal lines simply represent various constructs as may be seen in labels, barcodes, or machined parts. The center images for both systems (b, e) show the in-focus or best-focus image. The images to the left of center (a, d) give images for an object that is 35mm from the best-focus position, having been moved away from the camera. The images to the right of center (c, f) show results for the object at 35mm from the best-focus position, this time having been moved toward the camera. The Wavefront Coded images (d-f) clearly show an increase in depth of field over the traditional system, providing an overall nominal depth of field of greater than +/- 35mm. For this pictures the CDM Optics CPM 127-20 phase mask was used.

Fig. 3.20 gives six images comparing the performance of a traditional imaging system (a-c) to a system using Wavefront Coding (d-f) for a biometrics application. Here the image of a fingerprint may be used in a biometric identification scheme for security. Again, the center images for both systems (b, e) show the best-focus image. The images to the left of center (a, d) give images for the finger which has been translated 40mm from the best-focus position (away from the camera) and the images to the right of center (c, f) show results for the finger translated 40mm from best-focus position (toward the camera). Again, the Wavefront Coded images (d-f) clearly show an increase in performance over the traditional system, providing depth of field increase of 4 to 5 X over the traditional system. While with Fig 3.19a or 3.19c one may argue that the traditional imaging system still provides some resemblance to the object (albeit quite poor), in the biometric case the traditional imaging system barely produces recognizable images.

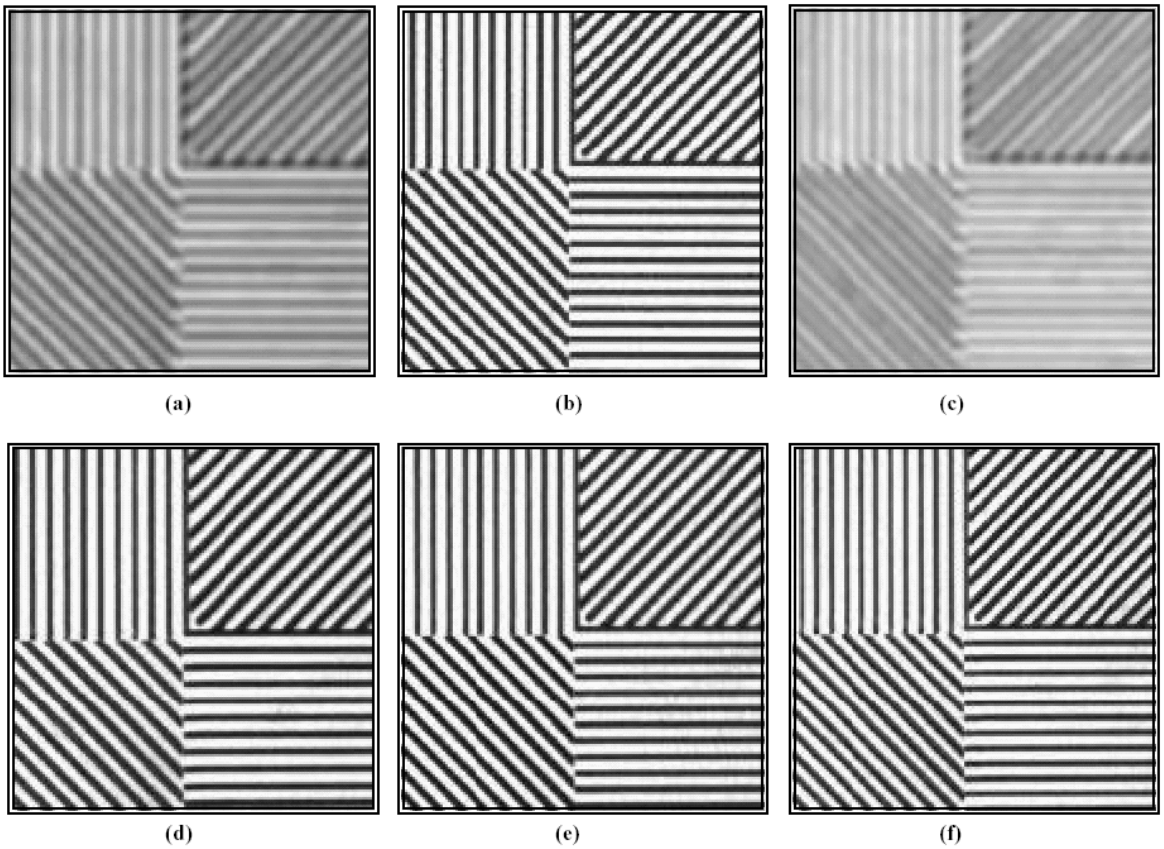


Fig. 3.19. Example images from machine vision/label reading application for a generic binary input. The horizontal, vertical, and diagonal lines simply represent various constructs as may be seen in labels, barcodes, or machined parts. Images from a traditional imaging system are given in a, b, and c. Images from a Wavefront Coded system are given in d, e, and f. Best focus images are b and e. Movement of the target 35mm away from best focus, away from the lens in images a and towards the lens in c, results in badly defocused images. Movement of the target in with the Wavefront Coded system, d and f, results in little noticeable change in the images.

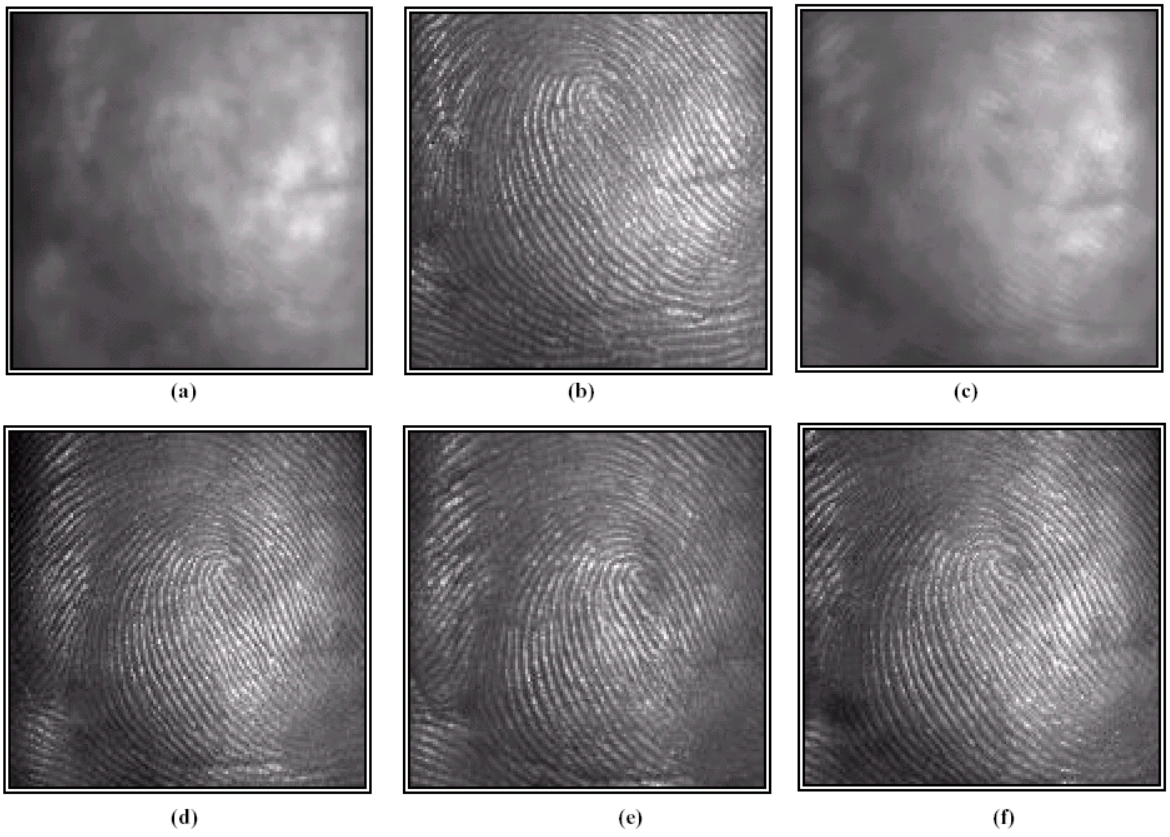


Fig. 3.20. Example images of biometric imaging of a fingerprint. Images with a traditional imaging are shown in a, b, and c. Images with a Wavefront Coded imaging system are shown in d, e, and f. The best focus images are given in b and e. Movement of the finger 40mm from best focus result in the images a and d (away from the lens) and c and f (towards the lens). The out of focus traditional images have very little spatial resolution. The Wavefront Coding images, both in and out of focus, are all sharp and clear.

3.12. **Effect of Extended Depth of Field on Aberrations**

Extending the depth of field of an optical system with a cubic-pm mask also decreases the sensitivity of the system to certain aberrations. For example, consider astigmatism. Astigmatism is present if the focus position for a vertical slice through the lens differs from that for a horizontal slice. But, if the depth of field is increased for both axes, then there is a region in the image plane where both are in focus. The effects of astigmatism are therefore reduced.

Effects of more complicated aberrations can be efficiently described through the ambiguity function. Below we specifically consider the spherical aberration.

The effects of aberrations on the ambiguity function can be compactly described through the multiplication/convolution property of ambiguity functions. Assume that the unaberrated pupil function is given by $P(x)$ and the contributing aberrations can be

described by a function $Q(x)$. The resulting aberrated pupil is then $P_{aberr}(x) = P(x)$ x $Q(x)$. Through the multiplication/convolution property of ambiguity functions, the ambiguity function corresponding to the aberrated pupil is

$$A(u,y) = \int A(u,\tau)P(x)A(u,y-\tau)Q(x)d\tau\,. \tag{3.31}$$

Or, the resulting ambiguity function is given by the convolution over the second variable, here given by y, of the corresponding component ambiguity functions. If the variable y is plotted as the vertical component, then the aberrated pupil ambiguity function is formed by filtering each vertical strip of the ambiguity functions corresponding to the unaberrated pupil and the pupil aberration respectively.

Convolution is often easy to visualize given simple component functions such as impulse and rectangular functions. If the main components of the ambiguity function corresponding to the standard system are shifted, or slightly blurred, due to aberrations then we would expect that the effect of this aberration on an extended depth of field system would be a minor alteration of the broad cubic-pm ambiguity function. Hence, the system would exhibit a low sensitivity to this aberration.

The effects of spherical aberration (SA) on the rectangular pupil is shown in figure 3.21. Spherical aberration on this pupil is defined in terms of W_{040} where

$$Q(x) = e^{jkW_{040}x^4}, \qquad |x| \le 1, \tag{3.32}$$

and where k is the wave number.

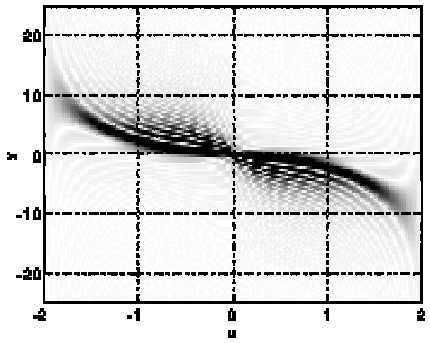


Figure 3.21. Magnitude of the ambiguity function of a rectangular pupil with spherical aberration of $W_{040} = 1.5\lambda$.

From figure 3.21, the effect of SA, with $W_{040} = 1.5\lambda$, on the rectangular pupil is to mainly shear the ambiguity function. Focus correction can be applied to minimize the effects of SA. Previous authors have suggested choosing the focus parameter W_{020} equal to minus W_{040} for best correction[30, 31] . Based on a graphical comparison of corrected ambiguity

functions, we feel that better focus correction for SA is given when W_{020} is set to $-3/5W_{040}$. This fraction is found by equating the sum total phase error across the pupil, with defocus and SA, to zero. The focus-corrected ambiguity function is shown in figure 3.22.

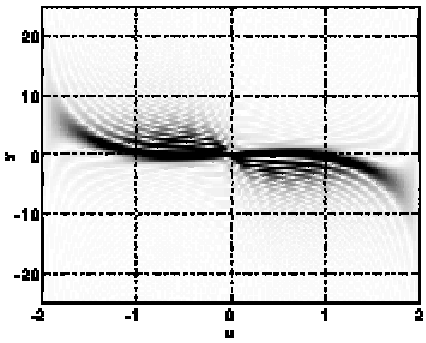


Figure 3.22. Magnitude of the ambiguity function of a rectangular pupil with spherical aberration corrected with defocus. Spherical aberration is W_{040} = 1.5λ, while defocus is $W_{020} = -\frac{3}{5}W_{040}$.

The aberrated cubic-pm ambiguity function is given in figure 3.23. The phase deviation constant α of is equal to 40. The value of W_{040} is again equal to 1.5λ. By comparing this ambiguity function with the unaberrated version with the same value of α, we see that the effect of SA on the cubic-pm extended depth of field system is mainly a shear. As in the rectangular pupil case, we would expect that focus correction could be used to minimize these shear effects.

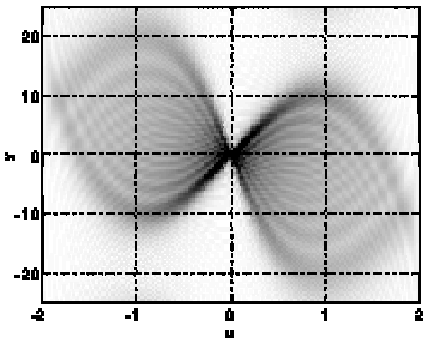


Figure 3.23. Magnitude of the ambiguity function of cubic-pm system with spherical aberration. Spherical aberration is W_{040} = 1.5λ.

Figure 3.24 shows a focus-corrected cubic-pm ambiguity function. The value of $W_{020} = -\frac{3}{5}W_{040}$ was found graphically. This focus-corrected ambiguity function is nearly

the same as the unaberrated ambiguity function. Unlike the focus-corrected ambiguity function of the rectangular system, the power of the corrected and unaberrated OTFs for each defocus value is nearly identical. Therefore, we can conclude that the cubic-pm extended depth of field system is much less sensitive to spherical aberration than is an equivalent rectangular pupil system.

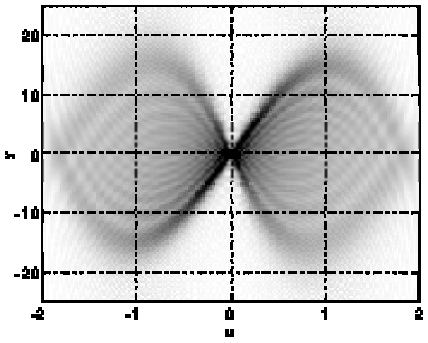


Figure 3.24. Magnitude of the ambiguity function of cubic-pm system with spherical aberration and focus correction.

If the optical system is of high quality, then the entire focus budget can be used to increase the depth of field. A portion, or all, of the focus budget can be used to make the imaging system invariant to focus-related aberrations. For example, with a lens that has significant amounts of Petzval curvature or curvature of field, the image focuses on a curved surface, not on a plane. If the depth of focus is extended, then instead of a thin curved region of focus, there is a thick curved region of focus that allows a flat detector array to be used. Consequently, the imaging system is insensitive to Petzval curvature.

In a similar manner, the imaging system can be made to be insensitive to astigmatism, and chromatic aberration.

4. With wavefront coding technology is possible to reduce the number of optical components in an optical system?

In an optic system, some optical components have as function to provide the optical power of the total system. However, most of the optical components have as objective to correct the mono-chromatic aberrations, like spherical aberration, coma, astigmatism, and the chromatic aberrations in order to obtain the wanted optical quality for the image. If with the wavefront coding technology the effects of the aberrations can be eliminated, then, when using this technology, it will be possible to reduce the optical number of components inside of any instrument.

Until the moment, the wavefront coding technology had demonstrated capacity to extend the field depth, this means that it is possible to design optical systems invariable to the defocus. This simple fact allows to relax the optical design of many systems when allowing large tolerances to production errors and assembles errors. **Thus we need less optical components.**

A color imaging system is susceptible to chromatic aberration due to the wavelength-dependent index of refraction of lenses. This results in three channels of data that are defocused relative to each other, see figure 4.1, thereby creating unwanted chromatic effects in the final color images. Existing methods for achromatism have limitations and may not meet the desired constraints.

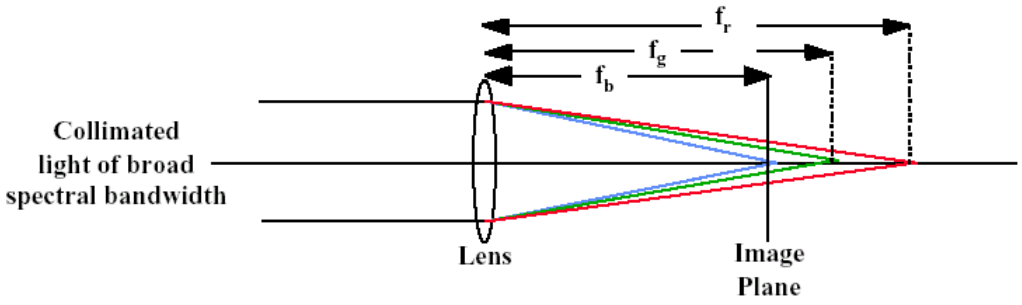


Figure 4.1. Demonstrating axial chromatic aberration with a single lens. The distances f_r, f_g and f_b are the focal lengths for red, green and blue light.

Chromatic aberration is often avoided by the use of achromatic doublets which are made by placing positive and negative lenses of different refractive indices in contact. This fact implies to duplicate the number of lenses and with this the cost, size, the time of production and weight of the optical system.

Thus, with this technology it is possible to reduce the components number in an optical system, because we do not need use achromatic doublets, only simple lenses.

Many optical system have a field curvature or Petzval curvature, the images focuses on a curved surface, not on a plane. The effect of this aberration is a variable defocus with the field. When the detector is flat, like a CCD, it is necessary to correct this aberration for obtain good quality images. The traditional method to eliminated this problem is to add a lens near to the detector.

If with wavefront coding technology it is possible to extend the depth of focus and large depth of focus is available, the detector plane can fit within the extended depth of focus volume. Thus all parts of the image within this volume will image clearly. **Therefore, we can to eliminated at least one lens in the system.**

The chromatic aberration and the field curvature are the only aberration that depend of defocus in a direct way. Therefore, with the wavefront coding technology these aberrations are the only aberration that we can to control, but *if the optical system is insensitive to this aberration, then we will be able to reduce the number of lenses about of 30%.*

5. Work for the future

The *challenges* for future work are:

a) **Suppress effects of other aberration**, such as

Spherical aberration

Coma

Astigmatism

Distortion

For this *we need phase surfaces with multiple design parameters* instead of single design parameter.

Cathey and Dowski use a cubic phase form described as $P(x,y)=\alpha\left(x^3+y^3\right)$, where α is the design parameter. My proposal is to use a phase function with multiple parameters expressed with Zernike polynomials as

$$P(\rho,\theta)=e^{j\sum_{5}^{10}a_i Z_i(\rho,\theta)}\,e^{ja_4 Z_4(\rho,\theta)}, \tag{5.1}$$

where Z_i are Zernike polynomials, and a_i are the design parameters.

The first Zernike polynomials are

$Z_1(\rho,\theta)=1$	aperture averaged wave front phase
$Z_2(\rho,\theta)=2\rho\cos\theta$	wave front tilt
$Z_3(\rho,\theta)=2\rho\sin\theta$	wave front tilt
$Z_4(\rho,\theta)=3.464\rho^2-1.732$	**pure defocus**
$Z_5(\rho,\theta)=2.449\rho^2\sin2\theta$	astigmatism
$Z_6(\rho,\theta)=2.449\rho^2\cos2\theta$	astigmatism
$Z_7(\rho,\theta)=(8.485\rho^3-5.657\rho)\sin\theta$	coma
$Z_8(\rho,\theta)=(8.485\rho^3-5.657\rho)\cos\theta$	coma
$Z_9(\rho,\theta)=2.828\rho^3\sin3\theta$	field curvature
$Z_{10}(\rho,\theta)=2.828\rho^3\cos3\theta$	field curvature
$Z_{11}(\rho,\theta)=3.416\rho^4-13.416\rho^2+2.236$	spherical aberration
.	.
.	.
.	.

To find the values of the design parameters, my proposal is to solve the problem as a optimization problem, where the Strehl ratio would be maximized and the variation of the PSF with defocus would be minimized. With the Strehl ratio we control the aberration, thus the Strehl ratio will be an image quality metric. At the same time if the PSF is insensitive to defocus we maintain all the properties of the previously analyzed wavefront coding technology.

The merit function of the optimization process will have several local minima, then we need a global optimization method, *my proposal is to use Genetic Algorithms with continuous parameters.*

The general procedure that I propose is:

- Expand pupil phase in basis of Zernike polynomials, truncated at some degree **p**, obtaining N terms with coefficients $\boldsymbol{a}_i$.
- Add quadratic defocus phase term, $\delta\rho^2$, to the expansion.
- Express Strehl ratio (SR) as a function of δ end the $\boldsymbol{a}_i$.
- Add constraints, maximal SR, minimal variance of PSF (require PSF to be insensitive to defocus).
- Solve optimization problem for basic coefficients $\boldsymbol{a}_i$.

b) Fabrication of phase surface

The phase plate is difficult to fabricate, and the phase plate calculated with multiples parameters will be more complicated. In figure 5.1 you can see a comparison between phase surface with one parameter and the phase surface with 6 parameters, for this I used the following values: $a_5 = -0.0018$, $a_6 = 0.0327$, $a_7 = -0.8975$, $a_8 = 0.8503$, $a_9 = 0.1428$, and $a_{10} = 0.1916$.

In figures 5.2 and 5.3, you can see the MTF for single and multiple parameters.

In figure 5.4 is shown a slices of MTF, and you can see the OTF do not have zeros, remember that this is important for the image processing stage.

My proposal for manufacture the phase plate are:

- Use a deformable mirror with a continuous membrane controlled electronically, like to the mirrors used in adaptive optics. With this kind of mirrors it is possible to control the shape of the surface with a high precision, and even more, we can have a surface that changes dynamically, this is important for the initial tests at laboratory level.

- Use a LCD to generate the phase distributions corresponding to the phase plate. The problem with this is that we would have a discreet distribution of phase due to the pixels of the LCD, although the ideal thing is a continuous distribution of the phase, the LCD could be useful in the laboratory tests.

Surface plot for cubic phase mask **Surface plot for Zernike phase mask**

Figure 5.1. surfaces plot for classical phase mask with one parameter and new proposal phase mask with 6 parameters.

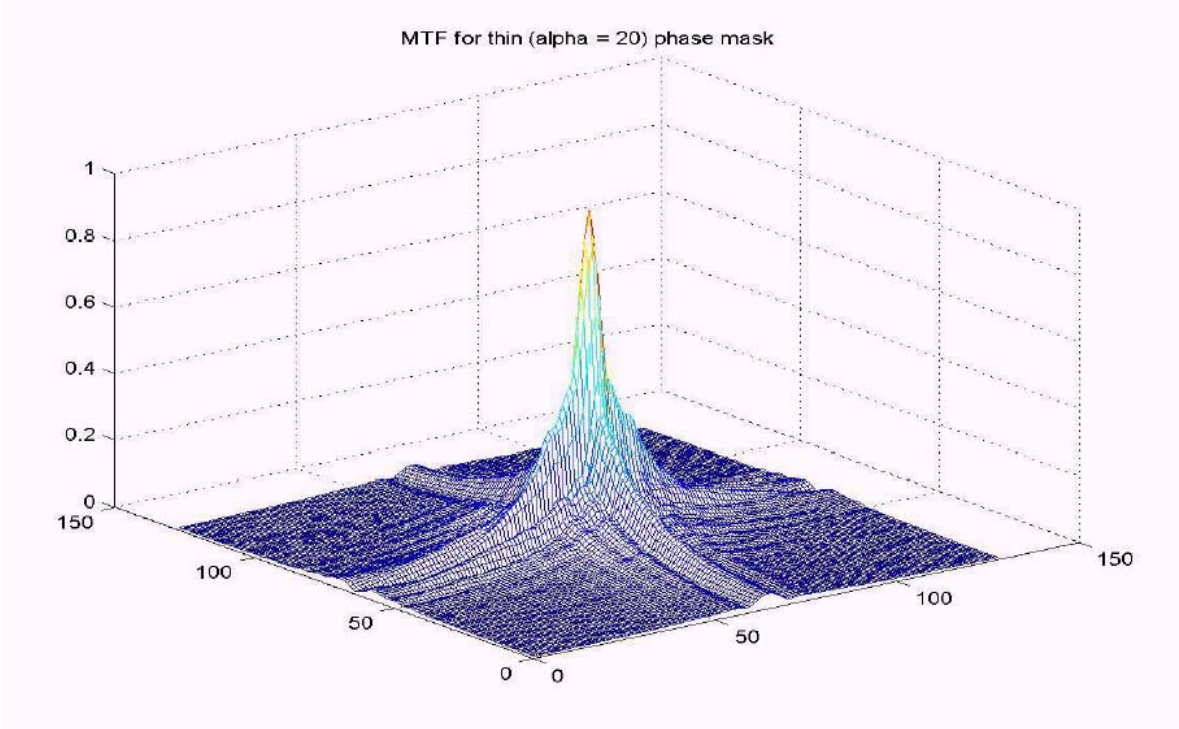


Figure 5.2. MTF of a cubic phase mask as determined with one parameter.

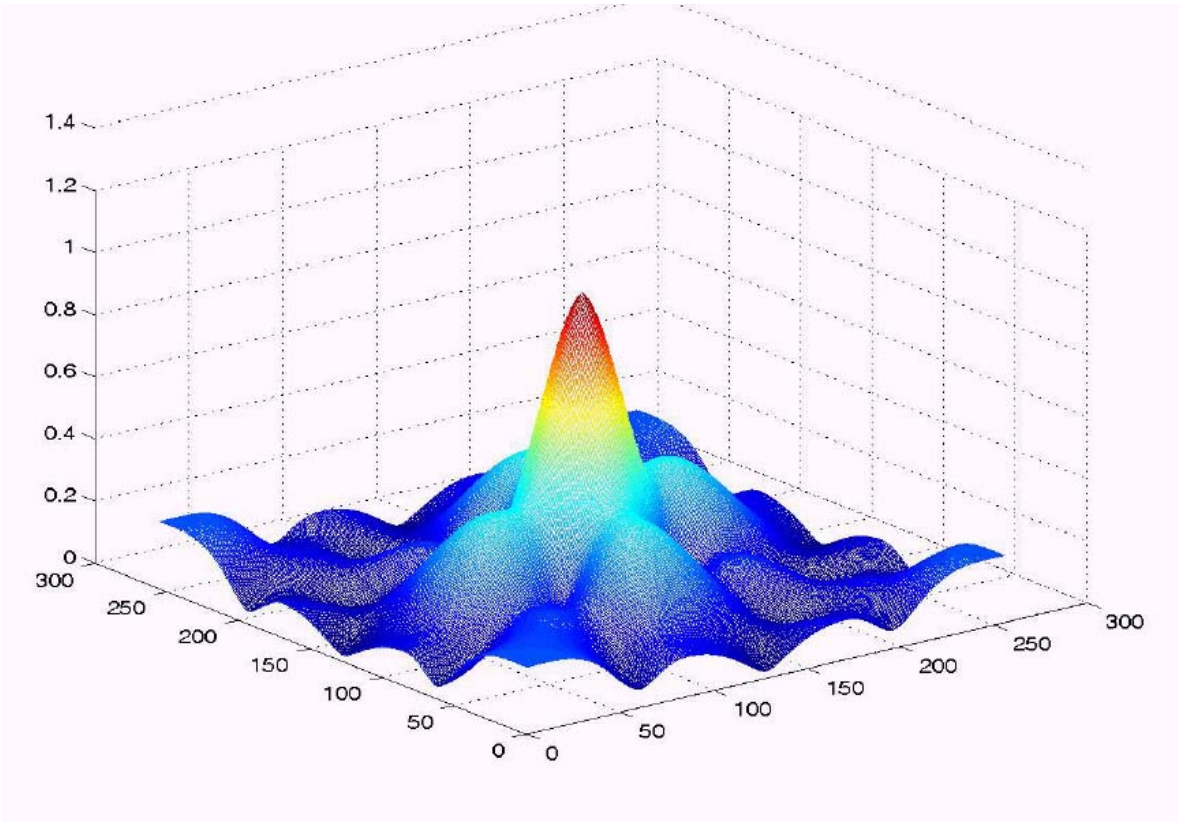


Figure 5.3. MTF of phase mask obtained from 6 design parameters

c) New methods for 2-D and 3-D image restoration

To decode the intermediate image, fast algorithms are needed, with less propagation of errors and an appropriate handling of the amplification of the noise.

It is very well-known that appropriate choice of restoration algorithm depend on many factor including: nature of blur, nature of object, image interpretation task, detection condition (varying light level), appropriateness of linearity and space invariance assumptions, etc.

For this point I have not a specific proposal, I think that we need to review the current methods, and if the current methods are not the solution, then we will need to develop a new method

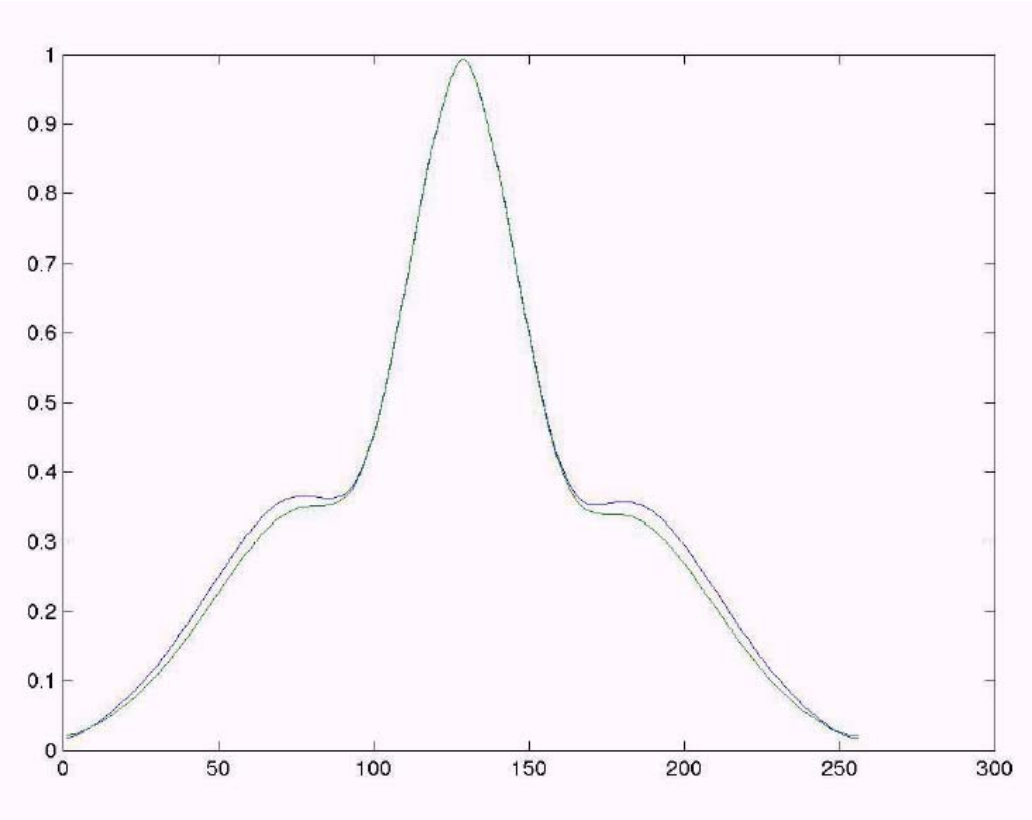


Figure 5.4. Slices of MTF along X and Y axes

My final comment is that the wavefront coding technology, in its current state, optimizes the combined work of the optics and the electronics, even just by extending the depth of focus it is an useful technique because it relaxes the restrictions of the optical design and it reduces, in some measure, the number of optical components. On the other hand, there is a lot of work to carry out in this direction because if we can control more and more aberrations, if we can develop faster and better image processing algorithms, and if we can develop cheap and fast techniques to manufacture the phase plates, then we will be in a situation in which the optical systems would reduce their components to the minimum, only the necessary components to provide the required optical power, but we should always include a optical component to code the wavefront and we should include a system digital decoder, with all the mentioned conditions the instruments of the future would reduce their size, weight and cost in important way.

References

1. M. Mino and Y. Okano, "Improvement in the OTF of a Defocused Optical System Through the Use of Shaded Apertures", Applied Optics **10,** 2219-2225 (1971).

2. J. Ojeda-Castañeda, P. Andres, and A. Diaz, "Annular Apodizers for Low Sensitivity to Defocus and to Spherical Aberration", Optics Letters **11,** 487-489 (1986).

3. J. Ojeda-Castañeda, R. Ramos, and A. Noyola-Isgleas, "High Focal Depth by Apodization and Digital Restoration", Applied Optics **27,** 2583-2586 (1988).

4. J. Ojeda-Castañeda, E. Tepichin, and A. Diaz, "Arbitrary high focal depth with a quasioptimum real and positive transmittance apodizer", Applied Optics **28,** 2666-2670 (1989).

5. J. Ojeda-Castañeda and Berriel-Valdos, ``Zone Plate for Arbitrarily High Focal Depth," Applied Optics **29,** 994-997 (1990).

6. W. Thomas Cathey and Edward R. Dowski, "New paradigm for imaging systems", Applied Optics, **41**, 6080-6092 (2002).

7. Max Born and Emil Wolf, *Principles of Optics*, Pergamon Press, New York, 1989.

8. P. M. Woodward, *Probability and information theory with applications to radar*, Pergamon, New York, 1953.

9. A. Papoulis, "Ambiguity function in Fourier optics", J. Opt. Soc. Am., **64**, 779-788, (1974).

10. J. P. Guigay, Optics Comm., **26**, 136, (1978).

11. K. H. Brenner, A. W. Lohmann, and J. Ojeda-Castañeda, "The ambiguity function as a polar display of the OTF", Optics Comm., **44**, 323-326, (1983).

12. J. W. Goodman, *Introduction to Fourier optics*, Mc Graw-Hill, New York, (1968).

13. G. Häusler, "A Method to Increase the Depth of Focus by Two Step Image Processing", Optics Communications **6,** 38-42 (1972).

14. E. R. Dowski, Jr., and W. T. Cathey, "Extended depth of field through wave-front coding," Appl. Opt. **34**, 1859-1866 (1995).

15. J. van der Gracht, E. R. Dowski, Jr., W. T. Cathey, and J. P. Bowen, "Aspheric optical elements for extended depth of field imaging", *SPIE Proceedings on Novel Optical Systems Design and Optimization*, San Diego, Vol. 2537, 279-288 (1995).

16. W. T. Cathey, B. R. Frieden, W. T. Rhodes, and C. K. Rushforth, "Image Gathering and Processing for Enhanced Resolution", Optical Society of America A **1,** 241-250 (1984).

17. J. W. Goodman, *Introduction to Fourier optics*, Mc Graw-Hill, New York, (1968).

18. W. T. Cathey, *Optical Information Processing and Holography* (John Wiley & Sons, New York, 1974) .

19. H. Bartelt, J. O. Casteñeda, and E. S. Enrique, ``Misfocus tolerance seen by simple inspection of the ambiguity function," Applied Optics **23,** 2693-2696 (1984).

20. A. W. Rihaczek, *Principles of High Resolution Radar* (McGraw-Hill, New York, (1969) .

21. C. E. Cook and M. Bernfeld, *Radar Signals* (Academic Press, New York, 1967) .

22. E. R. Dowski, "An information theory approach to incoherent information processing systems", in *Signal recovery and synthesis*, Vol. 11 of 1995 OSA technical Digest Series (OSA, Washington, D.C., 1995), 106-108.

23. B.R. Frieden, "Image Enhancement and Restoration," *Topics In Applied Physics*, Vol. 6, Picture Processing and Digital Filtering, Editor: T.S. Huang, (Springer Verlag, New York, 1979), 177-248.

24. Andrews, H. C. and B. R. Hunt, *Digital Image Restoration*, (Prentice-Hall, New Jersey, 1977), Chap. 8, 147-152.

25. S. Bradburn, E. R. Dowski, Jr., and W. Thomas Cathey, "Realizations of Focus Invariance in Optical-Digital Systems with Wavefront Coding", Applied Optics, vol. 36, 1997, pp. 9157-9166.

26. Amnon Yariv, Optical Electronics in Modern Communications, Fifth Edition, (Oxford University Press, Oxford, 1997), 431.

27. C. W. Helstrom, "Image Restoration by the Method of Least Squares," J. Opt. Soc. Am., **57**, 297-303 (1967).

28. J. van der Gracht, E. R. Dowski, Jr., M. G. Taylor, and D. M. Deaver, "Broadband behavior of an optical-digital focus-invariant system," Optics Letters **21**, 919-921 (1996).

29. J. van der Gracht, E. R. Dowski, Jr., W. T. Cathey, and J. P. Bowen, "Aspheric optical elements for extended depth of field imaging," *SPIE Proceedings on Novel Optical Systems Design and Optimization*, San Diego, Vol. 2537, 279-288 (1995).

30. J. Ojeda-Castañeda, P. Andres, and A. Diaz, "Annular Apodizers for Low Sensitivity to Defocus and to Spherical Aberration", *Optics Letters*, vol. 11, pp. 487-489, August 1986.

31. J. Ojeda-Castañeda, P. Andres, and A. Diaz, "Strehl Ratio with Low Sensitivity to Spherical Aberration", *J. Opt. Soc. Am. A*, vol. 5, pp. 1233-1236, August 1988.