Phase-Only Optical Information Processing

D. J. Potter

Index Chapter 1 2 3 4 5 6 7 8 9

Chapter 4
Optical Correlation

Chapters one, two and three of this thesis are concerned with elementary phase filtering operations on the spatial frequency spectra of a class of objects exhibiting spatial variations in phase only. Within each type of phase filtering operation (Schlieren, Dark Ground, Phase Contrast etc.) the filtering operation is independent of the particular phase object characteristics. Commonly, the purpose of the filter is to render the phase structure of the object visible as an intensity variation as described in the last chapter.

The second half of this thesis concentrates on a class of phase filters which allow much more intricate operations to be performed on the object spectra, and are subsequently more complicated in form. Specifically, attention is turned to the subject of spatial filters which allow certain pattern recognition operations to be performed. The object is usually an *amplitude*-only object so that whereas previously the complexity of the problem lay in the object structure, now the complexity lies in the structure of the filter.

Pattern Recognition

The purpose of this chapter is to introduce the general terminology of two dimensional spatial filtering, and in particular, pattern recognition filters. The construction and principles of operation of a two dimensional reconfigureable Fourier plane filter (a Spatial Light Modulator or SLM) is detailed in chapters five and six. Specifically, optical pattern recognition experiments have been performed with this device, and form the subject of chapters seven and eight. As such, this chapter lays an essential framework within which the subsequent work of this project should be set against.

In the first section, the concept of two dimensional phase filters with a frequency dependent complex transmittance is introduced. Spatial filters which allow pattern recognition tasks to be performed optically, such as the classical Vander Lugt matched spatial filter, are defined. Examples of practical modern day filters (SLMs) are given in section two together with a mathematical background to such effects as filter pixellation. In section three a major constraint upon many spatial light modulators, that of single parameter modulation, is discussed with relevance to its effect on phase-only correlation filters. Finally, section four discusses the merit of alternative correlation techniques and their relevance to this project assessed.

1 Two Dimensional Spatial Filtering

The alteration of the spatial frequency spectrum of an object is known as `spatial filtering'. Appendix one provides a brief review of the standard optical processing bench on which such operations are usually performed. A single spatial frequency present in the object has both an associated amplitude and phase, which is to say a number describing the spatial offset of that component from the origin of the object plane. In the frequency plane, the amplitude information of a single frequency is represented by the Point Spread Function of the transform lens, which is usually a sharply peaked function, and the phase as a temporal delay of the light field at that point.

1.1 Mathematics of Correlation Filters

One particular optical processing operation has been studied extensively, that of optical correlation. Consider the spatial frequency spectrum of an object g(x,y) which is conventionally denoted by $G(\nu_x,\nu_y)$. The spectrum can be separated into a product of two functions, one entirely REAL and describing the amplitude of $G(\nu_x,\nu_y)$ and the other complex, describing the phase distribution so that

$$G(\nu_x,\nu_y) = |G(\nu_x,\nu_y)|\, e^{i\,\Phi(\nu_x,\nu_y)} \qquad (1)$$

It is well known [28] that a filter $H(\nu_x,\nu_y)$ of form

$$H(\nu_x,\nu_y) = |G(\nu_x,\nu_y)|\, e^{-i\,\Phi(\nu_x,\nu_y)} \qquad (2)$$

will result in an image plane amplitude distribution which is the auto-correlation of g(x,y), which will be defined shortly (equation 4.4). Such a filter is known as a `matched' filter, being matched to one specific object function, in this case g(x,y). If the filter amplitude and phase correspond to amplitude and (conjugate) phase of a different function p(x,y) then the image plane contains the cross-correlation of the functions g(x,y) and p(x,y), which again will be rigorously defined shortly.

These results are at the very heart of most pattern recognition filters, for if it is desired that a particular object is identified in an input scene to a processor, one way of detecting the presence of that object is to use a spatial filter having identical amplitude and conjugate phase to the object in the frequency plane. The function g(x,y) is called the `target' object. Suppose an input scene to the correlator contains several different, spatially distinct objects f_i(x,y). The image plane will contain the cross-correlations of the target g(x,y) with the input functions f_i(x,y). Should one of the f_i(x,y) prove identical to the target function then the image plane will contain the auto-correlation of g(x,y) instead. Mathematically, the cross-correlation of two functions g(x,y) and f(x,y) is

$$c_{gf}(\Delta_x,\Delta_y) = \int\!\!\int_{-\infty}^{+\infty} g(\,(x-\Delta_x),(y-\Delta_y)\,)\, f^*(\,x,y\,)\,dx\,dy \qquad (3)$$

which is the integral of the area of overlap of the two functions [1]. after they are spatially offset by amounts Δ_x and Δ_y in the x and y directions respectively. It is usual to normalise this function by dividing by the zero ordinate of the autocorrelation of g(x,y), defined as

$$c_{gg}(0,0) = \int\!\!\int_{-\infty}^{+\infty} g(\,x,y\,)\, g^*(\,x,y\,)\,dx\,dy \qquad (4)$$

Using this convention, the cross-correlation c_{gf} has a maximum value of unity occuring when f(x,y)=g(x,y).

If the *power* P contained in two functions g(x,y) and f(x,y) is identical, power defined in the usual way as

$$P = \int\!\!\int_{-\infty}^{+\infty} g^2(x,y)\,dx\,dy \qquad (5)$$

then it can be shown that the normalised auto-correlation c_{gg} of a function is greatest at $c_{gg}(0,0)$. For functions of identical power, the central value of the autocorrelation can be shown to always exceed any cross-correlation value. As the auto-correlation function is generally peaked at the origin, the brighter target object auto-correlation peak may be picked out from the dimmer cross-correlation peaks as illustrated in figure 4.1.

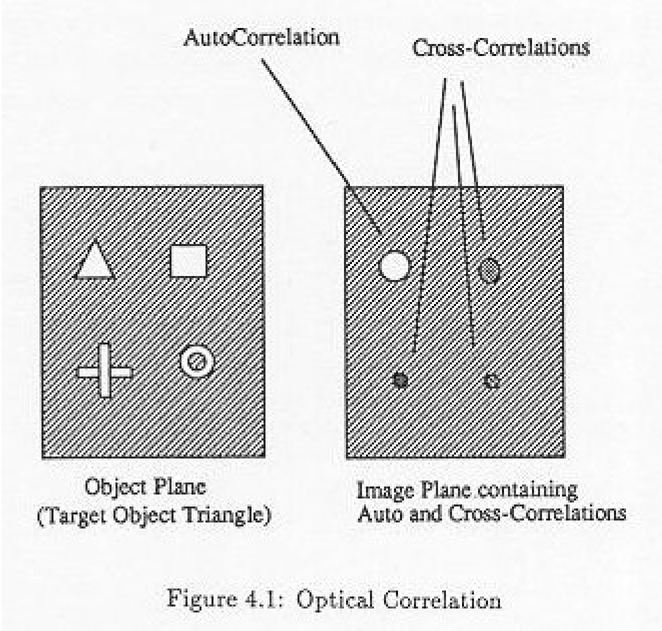

Figure 4.1: Optical Correlation

1.2 Vander Lugt Matched Filter

The first optical implementation of matched spatial filtering was carried out holographically by Vander Lugt [29]. The spectrum of a target object g(x,y) is caused to interfere with an off axis reference beam in the Fourier Plane and the resulting interference pattern recorder on photographic emulsion. Once the emulsion is developed, it may be reinserted into the Fourier plane to act as a matched spatial filter. If the input object is replaced by another of transmittance f(x,y), illuminated by a plane wave and the reference beam is removed, it can be shown [30] that three beams emerge from the far side of the holographic filter. One beam emerges at zero angle to the optical system and is focussed to the origin of the image plane, but contains no extractable information. Two further beams emerge at equal but opposite angles from the optical axis of the system and are again focussed to the image plane. One contains the *convolution* of the target function f(x,y) with the object function g(x,y) and the other contains the *cross-correlation* of the two functions. Until the advent of fast, reconfigurable spatial filters (to be described in section 4.2) much work on optical correlation utilised this experimental arrangement. See, for instance, the work of Casasent & Furman [31]. Although the matched spatial filter produces the highest signal-to-noise ratio (SNR, to be discussed shortly) in the image plane [32], there exist several disadvantages to its use in practice. Section 4.2 lists some of the drawbacks of this technique and describes alternative methods of implementing correlation optically. All of these, however, are modifications of the basic Vander Lugt matched spatial filter described here.

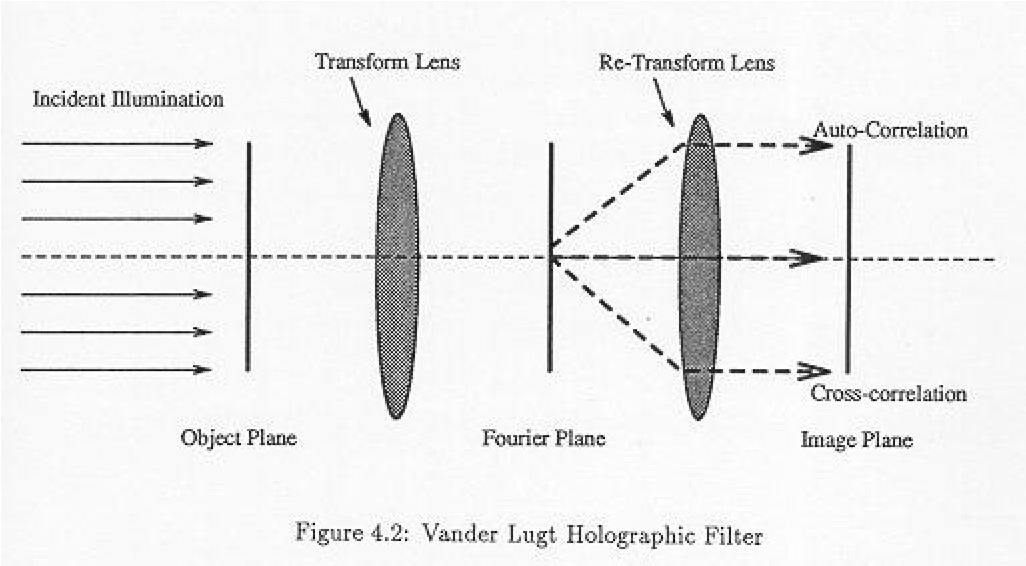

Figure 4.2: Vander Lugt Holographic Filter

Improved Discrimination

For many objects however the difference in brightness between the autocorrelation and cross-correlation arising from use of a matched spatial filter is quite small. It is thus difficult to discriminate between a *recognised* target by the brightness of the correlation peak compared with the surrounding cross-correlation peaks. Also, the peak need not necessarily be very sharp, so that even the auto-correlation may be a broad, dim function.

It has been shown however that discrimination is greatly improved if the input scene has undergone an edge enhancing operation before being used as input to the correlator (as commonly performed by the subtraction of the mean value of the object signal throughout the function, for example). The low spatial frequencies present in the object spectrum, which `fill in' large, fairly uniform areas of the object, are suppressed in such an operation in deference to the high spatial frequencies which define the edges of the object. Research performed by Horner and Bartelt [33] also indicates that binarisation of the input scene combined with a filter matched to the binary object, rather than a continuous tone object, results in a brighter correlation peak value and an improved signal to noise ratio (SNR). The precise specification of what the SNR measures, together with a parameter known as the optical efficiency η_H, are defined as follows:

1. **The SNR used was defined as the ratio of the peak correlation value to the rms noise outside of a 50% of peak threshold level.**

2. **HORNER efficiency η_H is defined as the ratio of the energy within the 50% of peak threshold level to the total energy falling on the output plane of the correlator.**

Table 4.1 compares the correlation peak intensity (as a fraction of that obtained by a classical matched filter), SNR and HORNER efficiency for several combinations of input and filter type. Subscripts $_{CTS}$ and $_{BIN}$ refer to the object type from which the filter was made. These results are from Horner and Bartelt, where the input scene was a doll's face, all results being computer simulations.

Input Object		$Filter_{CTS}$	$Filter_{BIN}$
Continuous	SNR	4.1	4.1
	R_0^2	1.0	1.3
	η_H	6.3%	3.1%
Binary	SNR	4.1	7.1
	R_0^2	1.1	3.4
	η_H	1.5%	0.7%

Table 1: Matched Filter: Three Parameters Affected by Input / Filter Type

Thus the highest SNR ratio and intensity of autocorrelation peak occur when a filter is made from a binarised input scene and the same binary object is presented at the input. In this case however optical efficiency decreases because although the peak is made sharper and higher, less energy goes into the peak region as a whole than it does with a classical matched filter using a continuous object.

In section 4.3.3, the same three parameters used here are again compared but for a filter which is matched only to the phase of the target object. It will be argued that a phase-only filter should give much improved results and a similar table to that shown above is presented, again quoted from Horner and Bartelt, to quantify this improvement.

2 Practical Filters

In practice, the matched spatial filters of Vander Lugt suffer from the severe drawback of lengthy preparation time and the need for a physically new filter for each target object. A significant increase in the speed of the whole process may be achieved by encoding several matched filters on a single piece of film. Such a filter is known as a \`frequency-multiplexed' or \`composite' filter [34], [35]. To avoid overlapping of the autocorrelations in the image plane, the holographically stored matched spectra of a number of different input objects are recorded using a plane wave reference beam with an object dependent angle to the optical axis. An important strength of this technique is the ability to store, on a single filter, a number of filters matched to both scaled and rotated versions of a single target object. Using computer generated holograms, this field has been investigated by Leger and Lee [36].

A practical use of an optical correlator has been suggested by Johnson [37] which serves to highlight this problem. If a small video camera views the entrance to a doorway it can send signals which, if suitably displayed, can be used as the input scene to a compact optical processor (The means whereby this might be effected are discussed shortly). The frequency spectrum of the input scene can then be filtered so as to produce a correlation peak whenever a particular face appears at the input to the system. This has obvious security implications. Due to the large number of variations in subject distance from camera, aspect ratios of the face and facial expressions, it would be required to scan through a number of filters in rapid succession for each time selected input scene, each filter pre-stored and calculated to recognise the face at various distances, etc. Holographic filters *could* be used if mechanically replaced rapidly enough but owing to the large number envisaged this scheme would be severely limited in the number of different people it could \`recognise', even if frequency-multiplexing were used.

Spatial Light Modulators

It would clearly be advantageous to update the filter pattern without physical removal of the filter. To date, a large number of devices exist which can act as reprogrammable spatial filters, and are known as \`Spatial Light Modulators' or \`SLM's. In general, a spatial light modulator is a device which may be used to impress information onto a wavefront, so that the information may represent an input image scene as an amplitude variation or a spatial filter as phase variations, for example.

Actual spatial light modulators are constrained by many factors to perform modulation of *either* amplitude or phase, but not both together. However, it is a frequent occurence for a phase modulator to also, sometimes unavoidably, introduce a small degree of amplitude modulation as well and vice versa for an amplitude modulator. These are aberrations of the filter and can usually be tolerated if small enough, but have led to some authors referring to specific SLMs as either \`phase-mostly' or \`amplitude-mostly' modulators [38], [39].

A further consideration in physical devices is that the modulation parameter frequently takes one or other of only two values, and such devices are said to have a binary mode of operation. For amplitude-only modulators this is usually adequate if it is desired either to pass or block a group of spatial frequencies, for example. Of far more concern is the effect this has on phase modulating filters. This forms the subject of section 4.3.3.

SLMs are subclassified as either pixellated, a necessity arising from the need to assign a stored memory location value to a particular position on the surface of the modulator, memory locations being discrete, or non-pixellated. The process of loading the information onto the SLM is known as \`addressing'. Pixellated devices are usually electrically addressed whereas non-pixellated devices are commonly addressed optically, though this *is* a generalisation.

4.1 Optical Addressing

As an example of what is meant by optical addressing, it will prove instructive to examine the operation of an SLM known as the \`Hughes Liquid Crystal Light Valve' [40], [41], the operation of which is depicted in figure 4.3.

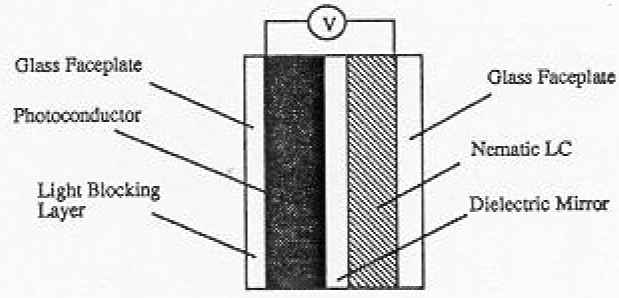

Figure 4.3: Operation of the Hughes LC Light Valve SLM

Figure 4.3: Operation of the Hughes LC Light Valve SLM

In this device, an image is focussed onto the front face of the SLM using an incoherent beam of light. A photoconductive layer of cadmium sulphide behind the glass substrate experiences a decrease in resistance as light falls on the surface, causing the voltage dropped across a layer of nematic liquid crystal to decrease. The light modulating effect is actually quite involved and uses a phenomena known as the hybrid field effect, but the specific effect is not of concern here. A coherent beam of light is reflected off the back face of the device and as it travels through the liquid crystal layer experiences a spatially varying optical effect which is primarily one of polarisation guidance in this case. By this means, the information of an incoherent signal is imprinted onto the wavefront of a coherent signal which can be used, for example, as the input stage to an optical processor.

4.2 Pixellated Devices

Whilst optically addressed devices are suitable for incoherent to coherent conversion in the input plane to an optical processor, they are of limited practical use as frequency filters with one notable exception - the Joint Transform correlator (See section 4.4). Most filters are calculated computationally and pixellation of the devices allows interfacing of the SLM to a computer for filter update. Consequently, most SLMs are pixellated devices requiring a discrete number of values to describe the modulation characteristics of each pixel. Figure 4.4 illustrates the general characteristics of a pixellated spatial light modulator in the frequency plane.

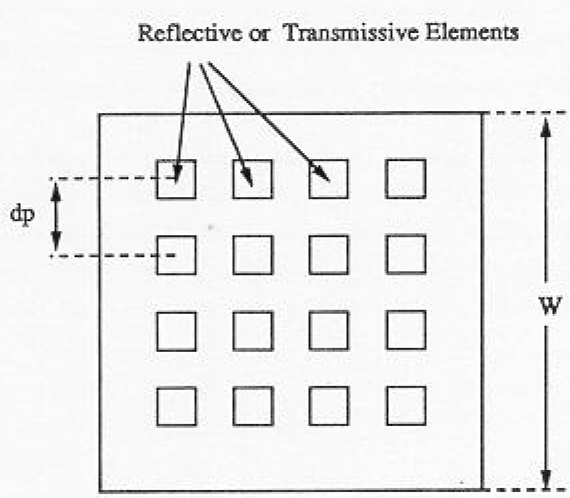

Figure 4.4: Characteristics of a Pixellated Spatial Light Modulator

Figure 4.4: Characteristics of a Pixellated Spatial Light Modulator

The pixels may either be transparent (a transmissive SLM) or reflecting (reflection mode SLM), in which case the reflected beam is isolated by using a beamsplitter placed in front of the device. The finite extent of the filter results in a maximum frequency passed by the device, each pixel covering a range of spatial frequencies which it

ideally modulates in an identical fashion. If `$F_L$' denotes the focal length of the lenses used in the processor, `x_T' the physical distance along the x-axis of the frequency plane then the frequency-distance relationship has been given as

$$\nu_x = \frac{x_T}{\lambda F_L} \tag{6}$$

The highest spatial frequency passed by the SLM with a width `W' is then

$$\nu\left(\frac{W}{2}\right) = \frac{W}{2\lambda F_L} \tag{7}$$

If the pixels are separated by a physical distance `$d_p$', the corresponding separation in frequency can be found. According to the `sampling theorem' [] a sampling in frequency space at interval $\delta\nu$ results in a replicated image of the object, each replication separated by a distance [$1/(2\delta\nu)$]. To avoid overlapping images the object width O_W must not exceed this value, so that

$$O_W = \frac{\lambda F_L}{d_p} \tag{8}$$

and

$$O_W \times \nu\left(\frac{W}{2}\right) = \frac{W}{d_p} \tag{9}$$

which in a pixellated device is merely the number of pixels along any one axis of the SLM, generally assumed here to be square. This result is known as the `space-bandwidth product' (or SBP), and states that a gain in image resolution (by passing higher frequencies) can be obtained but only at the expense of the maximum object size usable as input, if aliasing in the image plane is to be avoided.

This result shall be of central importance in chapter 7 and shall be discussed with reference to a specific optical processing system, and is noted here as a figure of merit of an SLM. Notably, the greater the number of pixels the better becomes the SBP becomes allowing either improved image resolution or larger object dimensions for the same degree of resolution.

4.3 Electrical Addressing

One commercially available electrically addressed SLM is known as the Litton magneto-optic SLM or LIGHT-MOD [42], [43] as it is more commonly known, utilises the effects of Faraday rotation as a light modulation mechanism. The LIGHT-MOD is a 48×48 pixellated transmissive array[2], the pixels defined by the areas of intersect of a network of `drive lines' in which currents are caused to flow. By suitably addressing the device, the resulting currents in the drive line network cause the magnetisation vector in the region of a pixel to lie in one of two possible directions. Consequently, a linearly polarised light field normally incident on the array has its polarisation vector rotated in either a positive or negative sense according to the magnetic field on each pixel by the Faraday effect. By a suitable arrangement of polarisers before and after the SLM either binary amplitude or binary phase modulation may be achieved.

This device has been used extensively in experimental studies, most notably by Flannery et al [44] in 1986 where the results of an initial investigation into optical binary phase-only filtering were published. `Excellent agreement' between computer simulation and the experimental correlations was found and a photograph was presented showing two bright, distinct spots of light in the output plane of the optical correlation bench. Further experimental correlation results from a very thorough investigation were published in 1988 where an impressive agreement with predictions from simulations was obtained [45].

The space-bandwidth product (as determined by the number of pixels along one axis of the filter) of this device is a factor of three times higher than that of the 16×16 array used in this project. Thus the results of Flannery et al, particularly those of reference [44], may be used as a benchmark for comparison with those of this project. More shall be said of this in chapter eight.

4.4 Information Function

In the mathematical description of a pixellated filter it is often useful to form an expression describing the modulation parameter of each pixel of the SLM. Individual pixel modulation parameter settings are described by an *information* function v(x,y), the local value of which is identical to the modulation parameter at any given pixel location. Consider the description of an SLM used in the object plane: the array of pixels is commonly described by a Dirac comb function [], convolved with a pixel function describing the shape of each pixel (assumed not to vary over the SLM). Reducing to one dimension for simplicity, and assuming the filter is so large that the summation may be extended to ±∞, the light modulation performed by the SLM can be written as

$$t(x) = \left[\, v(x) \sum_{n=-\infty}^{+\infty} \delta(x - n\Delta) \,\right] * P(x) \qquad (10)$$

where `*' denotes convolution, Δ is the separation of mirrors in the object plane and P(x) describes the nature of the pixel. Here it has been assumed that the light field over regions of the SLM not covered by a pixel is zero for simplicity. Upon Fourier Transformation, the shape and transmittance information of the pixel, described by P (x), results in a multiplicative envelope to the spectrum. For example, if the pixel is square of side `a' then in one dimension it may be represented by the rectangle function described by

$$R\left(\frac{x}{a}\right) = \begin{cases} 1 & |x| \le \dfrac{a}{2} \\ 0 & |x| > \dfrac{a}{2} \end{cases} \qquad (11)$$

which has Fourier Transform [17]

$$F(\nu) = \mathrm{sinc}\left(\frac{\nu}{a}\right) \qquad (12)$$

This particular function[3] gets broader as the pixel width a becomes narrower, and in general the envelope function attenuates the higher order replications caused by pixellation of the filter. The Fourier Transform of equation 4.10 is, ignoring the envelope function, given by

$$T(\nu) = V(\nu) * \sum_{n=-\infty}^{+\infty} \delta\left(\nu - \frac{n}{\Delta}\right) \qquad (13)$$

where V(ν) is the Fourier Transform of the information function. It is observed that the Fourier Transform of the information function is replicated in the frequency plane. This observation underpins the idea of using a pixellated SLM as input to an optical system, the ideal continuous object being sampled (perhaps a local average of the function is taken as the information function) and a bandlimited continuous spectrum is observed in the frequency plane for further processing. The central or zero order replica of V(ν) is commonly used for further processing, the envelope function generally having higher attenuation for the outer replications which are lowpassed so as not to take part in the formation of the image. Care must be taken in design of the SLM that too wide a pixel is not used for then the envelope function would strongly attenuate the outer regions of the zero order replication.

Optimum Sampling

Note that if the bandlimit of the information function is too wide (> [1/(Δ)]) then the spectral replicas overlap and aliasing occurs, so that high frequencies from the outer replications wrongly appear in lower frequency locations of the zero order replication. (Figure 4.5).

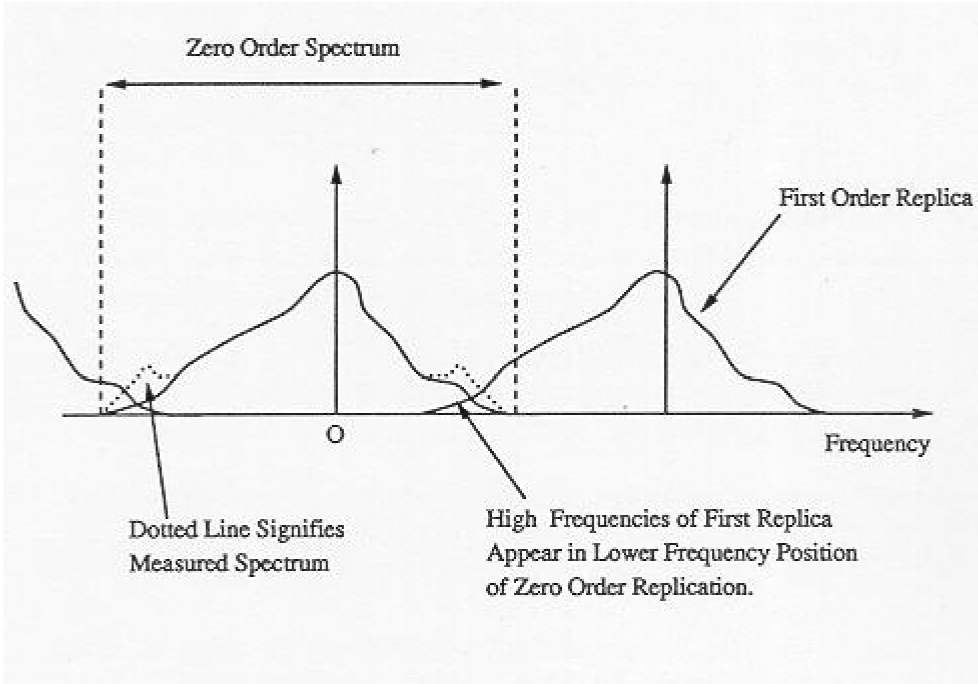

Figure 4.5: Aliasing arising from too low a sampling interval

This occurs if the SLM sampling interval Δ is too large as may be seen from equation 4.13. `Optimum sampling' refers to the situation where the spectral replications just touch at their peripheries, so that if ν_L is the highest frequency required in the information function then

$$\nu_L = \frac{1}{2\Delta} \tag{14}$$

3 Single Parameter Correlation Filters

The general characteristics of practical filters (pixellation, bandwidth, etc.) have been introduced. The primary limitation of many pixellated SLMs is the binary mode of operation. As stated in chapter one, it is a primary objective of this project to demonstrate the capabilities of a low space-bandwidth-product, pixellated, binary-mode SLM as an optical correlator. Therefore it is required now to ascertain the likely performance of correlation filters which have only two phase values allowed - binary phase-only filters or BPOFs, as they are commonly known. Specifically there are three points to be addressed.

1. **Given that the vast majority of SLMs available can function either as binary amplitude-only or binary phase-only filters, which modulation parameter should be chosen to represent a pixellated correlation filter ?**

2. **The classical `matched' correlation filter requires both amplitude and phase filtering operations to be performed. Whichever modulation parameter is chosen, how will the filter perform with only one modulation parameter ?**

3. **What effect on filter performance does the quantisation of the modulation parameter to only two values have ?**

3.1 Choice of Modulation Parameter

Consider an object centered in the object plane of an optical processor. At the image plane it is desired to obtain the cross-correlation of this object with some target object, which is a sharply peaked function centered about the origin of the image plane. The phase of the filter serves to cancel out the phase of the object spectrum so that no phase variations exist over the frequency plane. As such, the complex light field immediately behind the filter behaves like a plane wave but with a spatially varying amplitude over the wavefront. This wave is focussed down to the center of the image plane and the resulting sharply peaked image plane light distribution is the cross-correlation of the object with the target object from which the filter was calculated. It is likely that sharper still focussing would occur if the amplitude over the surface of the plane-type wave was uniform, so that the light field immediately after the filter perfectly resembles a plane wave in both amplitude and phase.

Looking at the process in another way, setting the phase of the spectrum to be zero for each and every Fourier component means that the spatial offset of each spatial frequency in the image is zero. Thus if the image amplitude $g(x_i,y_i)$ is described as a Fourier integral

$$g(x_i,y_i) = \int\int_{-\infty}^{+\infty} G(\nu_x,\nu_y)\, e^{i\Phi(\nu_x,\nu_y)}\, d\nu_x\, d\nu_y \qquad (15)$$

where $\Phi(\nu_x,\nu_y)=0$ the components will add all in phase and resulting in a very large value at the origin of the image plane. This central value is increased further by setting all spatial frequency amplitudes to a constant value. (Indeed, this is the mathematical definition of the spectrum of a δ-function).

As such, it would seem that the phase information of the filter is primarily responsible for the correlation process. Indeed, the phase spectrum of an object is unique, whereas the amplitude spectrum need not be [24]. Correlation performed with an amplitude-only filter is compared to that from phase-only filters in reference [46] where it is also concluded that an amplitude-only filter is virtually useless.

Having established which parameter is the more important, the effect of dropping the other modulation parameter - amplitude - is now considered. This will be aided by studying some particular types of phase filter.

3.2 Phase Only Filters

Actual correlation filters commonly differ from the matched filter thus far introduced. The matched filter, as described by equation 4.2, has an amplitude equal to that of the Fourier Transform of the target object. The phase-only filter however (POF) has an amplitude everywhere equal to unity, and is of form

$$H(\nu_x,\nu_y) = e^{-i\,\Phi(\nu_x,\nu_y)} \qquad (16)$$

This filter may be considered to be a product of two filters, one the matched spatial filter and the other having an amplitude transmittance of [$1/(|G(\nu_x,\nu_y)|)$]. As such, the second filter in the product emphasises the lower amplitude (and commonly high spatial frequency) components of the spectrum and acts in a similar manner to a high pass filter. The sharpness of the correlation peak derives mainly from high spatial frequencies in the image, and consequently the correlation sharpness is greatly improved. Some quantitative results will be quoted later in this section. Notice how the elimination of the amplitude modulation is expected to improve the filter performance rather than detract from it[4].

As mentioned earlier, practical filters (SLMs) most commonly modulate only a single parameter. Further, this parameter frequently is allowed to take on just one of two possible values. From the discussion above the choice of modulation parameter in such a device should be phase rather than amplitude in a practical optical correlation system. Knowing the relative importance of the phase information of a spectrum, it is important that a theoretical basis be laid which shows that binarisation of phase is an allowable procedure in the first place. Such an analysis has already been published and is summarised here.

Phase Quantisation

As mentioned earlier, the vast majority of phase modulating SLMs cannot perform continuous phase modulation. Their modulation capabilities are *quantised*, frequently to only two levels, and a proper understanding of the effects of phase quantisation is thus required.

The effects of phase quantisation of the Fourier Transform on the image has been analysed in detail by Goodman and Silvestri []. They found that the image was described by a series of terms relating to the number of phase quantisation levels. Specifically, if N denotes the number of quantisation levels and $g'(x)$ the actual image amplitude obtained then

$$g'(x) = \sum_{m=-\infty}^{+\infty} \mathrm{sinc}\left(m + \frac{1}{N} \right) g_m(x) \tag{17}$$

where

$$g_m(x) = \int_{-\infty}^{+\infty} |G(\nu)| \; e^{i\,(Nm+1)\,\Phi(\nu)} \times e^{i\,2\pi\nu x}\, d\nu \tag{18}$$

It can be seen that the m=0 term corresponds to an attenuated version of the ideal image, whose strength increases as the number of quantisation levels N increases. This image has been termed the `primary' image. Terms corresponding to values of m other than zero are termed `false' images, whose strength decreases as N increases.

This work forms the basis of a theoretical explanation as to why binary phase-only filters still work. The ideal Fourier Transform of a continuous phase-only filter should resemble the object function from which the filter was calculated. In reducing the allowable phase values to only two the above analysis reveals that the primary image still exists, though attenuated, together with a number of unwanted (modified) images superimposed. In the next subsection, simulations performed by Horner et al indirectly provide a quantitative comparison of the effects of these `false' images on the correlation arising from a binarised phase filter with the correlation arising from a continuous phase-only filter.

3.3 Binary Phase Only filters (BPOFs)

Binarisation of the filter phase must be done in many spatial light modulators due to their binary mode of operation, more of which shall be discussed later. A very large number of binarisation algorithms have been proposed, the aim of which is to produce a single number with which to characterise the phase of a single SLM pixel. This number most commonly takes the value 0 or π radians, arising partly because some SLMs can only retard one pixel by π radians relative to another. In order that a binary phase value be calculated, the target object is modelled on a computer and the Fourier Transform taken. The phases $\Phi(\nu_x,\nu_y)$ of all points lying in the physical extent of any pixel are binarised according to a relation such as

$$\Phi_B(\nu_x,\nu_y) = 0 \quad \Phi \geq 0$$

$$\Phi_B(\nu_x,\nu_y) = \pi \quad \Phi < 0$$

The performance of matched filters, phase-only filters and BPOFs has been studied by Horner et al [] to whom the above binarisation criteria is attributed. Choosing a 2-D object, the face of a doll, Horner has investigated the performance of each filter and characterised this by peak intensity, optical efficiency and signal to noise ratio in the same manner as for the classical matched filter of section 4.1.2. The peak intensity was measured in units whereby the matched spatial filter gives a peak intensity of unity. For both a continuous tone and binary input object, table 4.2 quotes Horner's results for the matched (MF), phase-only (POF) and binary phase-only (BPOF) filters.

	MF	POF	BPOF

$Filter_{CTS}$, Object Continuous	1	125	36
$Filter_{BIN}$, Object Binary	3.4	1191	471

Table 2: Correlation Peak Intensities relative to the Matched filter

These results suggest that phase only filters far outperform, with respect to peak intensity, the classical matched spatial filter. Also the BPOF, though not as good as the POF, produces correlation peaks several hundred times brighter than the classical matched filter if the target object is binarised and the filter made from the binary object also. Binary phase-only filters have recently been given serious consideration [48] as a means of guiding remote spacecraft to their landing site, specifically the Mars Rover Sample Return mission of NASA [49]. Although the correlation is proposed to be performed electronically rather than optically, the BPOF guidance technique has been demonstrated to match the best tracking algorithms to date and thus illustrates the power of this class of filter.

3.4 Binary Phase Filter Spectra

The complete analysis of Goodman and Silvestri may be complemented in this chapter by a related analysis on the spectra of binary phase objects and filters. In this case the information function v(x,y) of section 4.2.4 is phase only so that a binary phase filter would be represented by

$$v(x,y) = e^{i\,f(x,y)} \qquad (19)$$

where f(x,y) is a two dimensional, REAL binary function. It will now be shown that the spectrum of v(x,y) does not change form between describing a binary amplitude object or a binary phase object. The analysis is performed in 1-D for ease of discussion though the generalisation to 2-D is immediately apparent. Let f(x) be a REAL, binary function with minimum zero and maximum unity as shown in figure 4.6. This function might be the 1-D representation of a binary amplitude pattern displayed on an SLM, for instance.

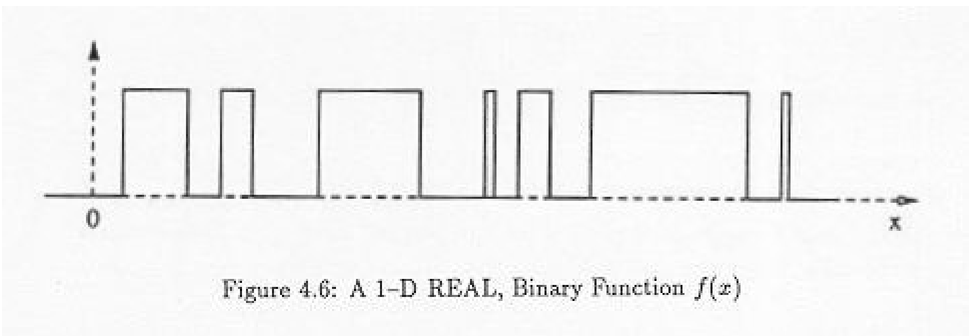

Figure 4.6: A 1-D REAL, Binary Function f(x)

The function comprises a set of rectangle functions placed along the x-axis, though these need not be placed at regular intervals nor do they need to have identical widths for this analysis. One may now describe a binary <u>phase</u> object g(x) with phase retardance α by

$$g(x) = e^{i\,\alpha f(x)} \qquad (20)$$

Using a Taylor expansion of the exponential and using the fact that $f(x)^k=f(x)$, f(x)=1 or 0 only, it is straightforward to show that

$$\begin{aligned} e^{i\,\alpha f(x)} &= \cos(\alpha f(x)) + i\sin(\alpha f(x)) \\ &= 1 + \left(\frac{\alpha^2}{2!} - \frac{\alpha^4}{4!} + \ldots\right) + i\sin(\alpha)\, f(x) \end{aligned} \qquad (21)$$

This may be further written as

$$e^{i\,\alpha f(x)} = 1 + [\cos(\alpha) - 1]\, f(x) + i\sin(\alpha)\, f(x)$$

$$= \quad 1 + [\, e^{i\alpha} - 1 \,]\, f(x) \tag{22}$$

If F(ν) and G(ν) represent the Fourier Transform of f(x) and g(x) respectively, it follows that

$$G(\nu) = \delta(\nu) + [\, e^{i\alpha} - 1 \,] F(\nu) \tag{23}$$

Therefore, the intensity of the phase filter spectrum is given by

$$I(\nu = 0) \quad = \quad \frac{1}{2}\, (\, 1 + \cos(\alpha)\,)\, F(\nu = 0) \tag{24}$$

$$I(\nu \neq 0) \quad = \quad 2\, (\, 1 - \cos(\alpha)\,)\, F(\nu = 0) \tag{25}$$

As can be seen, this analysis is not restricted to just one dimension and makes the following two predictions:

1. **The spectrum of an arbitrarily shaped binary phase object is identical in form to that of a binary amplitude object, save for a complex multipicative constant.**

2. **Variations of the phase retardance α serve only to vary the intensity of the zero frequency relative to the rest of the spectrum as a whole.**

Binary phase objects thus do not suffer from the effects of *ghost* spectral orders as defined in chapter two. This effectively defines the useful region of the Bessel function analysis to non-binary phase objects. However, for the sake of completeness, the area of overlap between the two regimes has been examined and is briefly presented here. Using the Bessel function convolution program, the properties of the spectrum resulting when the initial Fourier series coefficients are those of a square wave phase object are compared with the predictions of equation 4.44. Each Fourier coefficient will give rise to a Bessel comb, but from the work above it is known that no light must be found outwith the confines of the equivalent amplitude object spectrum. Therefore, the Fourier coefficients must be such that cancellation of the ghost orders occurs.

Twenty-one coefficients of the square wave series were used in the program and the amplitude of the spectrum at the frequency origin recorded for several phase modulation depths of the square wave which the coefficients describe. Figure 4.7 plots the theoretical spectral amplitude at ν = 0, from equation 4.41, together with the measured amplitude from the Bessel function program.

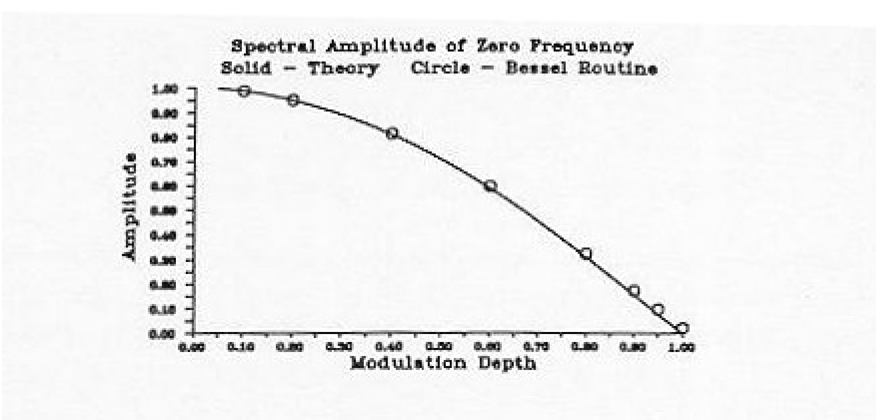

Figure 4.7: Amplitude at $\nu = 0$, Theory and Bessel Function Program Results.

Figure 4.7: Amplitude at ν = 0, Theory and Bessel Function Program Results.

As can be seen, excellent overall agreement between the Bessel function program and the theoretical expression is found. Further, the percentage of light energy outwith the region of the 21 Fourier coefficients was never observed to exceed 2% of the total energy of the spectrum indicating that cancellation of the ghost orders did indeed occur.

For a phase retardance of $\alpha = \pi$, a 1-D binary phase object with equal areas of 0 and π retardance should cause complete cancellation of the light field at the frequency plane origin. The evolution of the convolution process which leads to this cancellation in the Bessel function program is shown in figure 4.8, where the ordinate records the spectral amplitude after the N'th convolution stage.

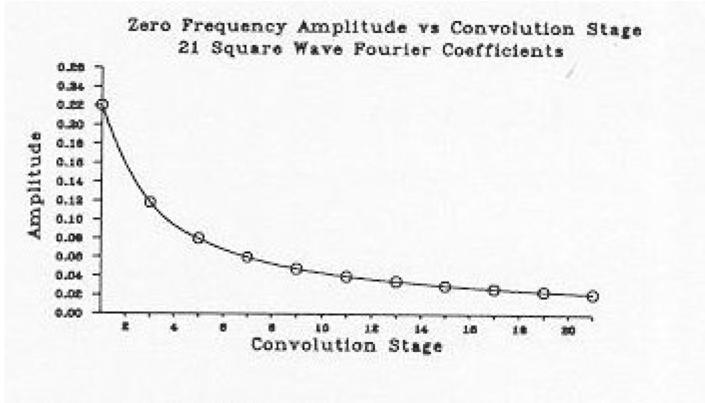

Figure 4.8: Evolution of Zero Frequency Cancellation

Figure 4.8: Evolution of Zero Frequency Cancellation

It is observed that cancellation occurs rapidly after just a few convolution stages have taken place, so that although the Bessel function analysis *can* be used in the analysis of binary phase filters, more convenient analysis techniques exist and the Bessel function analysis ends its usefulness for this project here.

Rigorous Imaging Equation

As shown in chapters two and three, a fully analytical solution for the spectrum of a general phase object is difficult to obtain, yet alone an expression for the image intensity. For a binary phase object, however, a fully analytical expression for the image intensity after a phase contrast operation on the spectrum may be deduced and is included here for completeness.

Let $F(\nu)$ be split into a two spatially exclusive components

$$F(\nu) = F(0)\,\delta(\nu) + F_s(\nu) \qquad (26)$$

where $F_s(\nu)$ is the whole spectrum minus the zero frequency component. Then, if $E(\alpha)=e^{i\alpha}-1$, equation 4.40 may be written

$$G(\nu) = [\,1 + E(\alpha)\,F(0)\,]\,\delta(\nu) + E(\alpha)\,F_s(\nu) \qquad (27)$$

so that performing a positive phase contrast operation gives a spectrum of form

$$G_p(\nu) = i\,[\,1 + E(\alpha)\,F(0)\,]\,\delta(\nu) + E(\alpha)\,F_s(\nu) \qquad (28)$$

The image field is given by the Fourier Transform of equation 4.45. If we set

$$f_s(x) = f(-x) - F(0) \qquad (29)$$

then the image field $g_p(x)$ which results is of form

$$g_p(x) = [\,f_s(x)\,(\cos(\alpha)-1) - F(0)\sin(\alpha)\,] + i\,[1 + F(0)\cos(\alpha) - F(0) + f_s(x)\sin(\alpha)\,] \qquad (30)$$

Squaring the REAL and IMAGINARY parts leads to an intensity $g_p^2(x)$ of

$$g_p^2(x) = 1 + 2f(-x)\,[\,\sin(\alpha) + 2\,A\,F(0) - A\,] - 4\,A\,F^2(0)$$

$$- \quad 2 \sin(\alpha)\, F(0) + 2\, A\, F(0) \tag{31}$$

where $A=\cos(\alpha)-1$.

It should be noted that

1. **The analysis is applicable to 2-D phase objects although it is conducted in 1-D.**
2. **Binary phase objects need not be periodic for this analysis to apply.**
3. **For a 1-D square wave phase grating of total width D having a total of n rectangle functions of phase retardance α and width W, the variable F(0) reduces to [nW/D].**

In the limit of small phase, equation 4.48 gives the intensity over a phase pixel as

$$g_p^{\,2}(x) \cong 1 + 2\alpha\,[1-F(0)] + \ldots \tag{32}$$

whereas the Taylor expansion of the exponential would suggest an intensity I(x) of

$$I(x) \cong 1 + 2\alpha + \ldots \tag{33}$$

The full analytical analysis presented here shows that not only does the intensity of a phase pixel depend on the phase retardance of that pixel (as the Taylor expansion shows) but also on F(0) which is proportional to the mean of f(x). This prediction has been verified by comparing the analytical prediction of pixel intensity of equation 4.48 with that obtained by computer simulation where the spectrum was computed using a Fast Fourier Transform algorithm. The results are shown in figure 4.9.

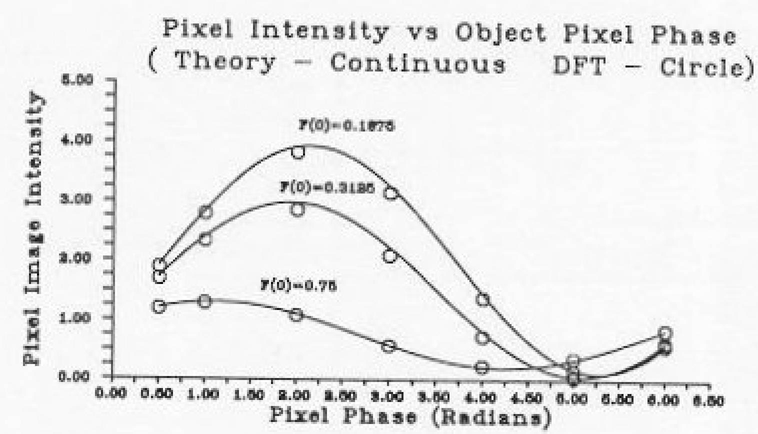

Figure 4.9: Pixel Intensity of a Binary Phase Object after Positive Phase Contrast Filtering: Theory and Simulation

Figure 4.9: Pixel Intensity of a Binary Phase Object after Positive Phase Contrast Filtering: Theory and Simulation

This short subsection has shown binary phase objects to be tractable to analytical solution in as far as certain phase visualisation operations are concerned. Further, the results presented here may be used as a basis of phase calibration for certain 2-D transmissive binary phase only filters if the variable F(0) is replaced by the *fill factor* of the device[5].

3.5 Noise Effects

Input scenes to a correlator usually suffer from a degree of noise, defined generally as small, random variations in amplitude transmission. Such noise will have an associated power spectrum. `White' noise, for example, is defined

as having an equally strong power spectrum at all spatial frequencies. The effect of a phase-only filter is to pass all frequency components without any attenuation of amplitude. Clearly, if the bandwidth of the input signal is much less than the pass-band of the spatial filter then if white noise is present a large amount of noise is passed to the image plane with no additional increase in signal (the object spectrum) as illustrated in figure 4.10.

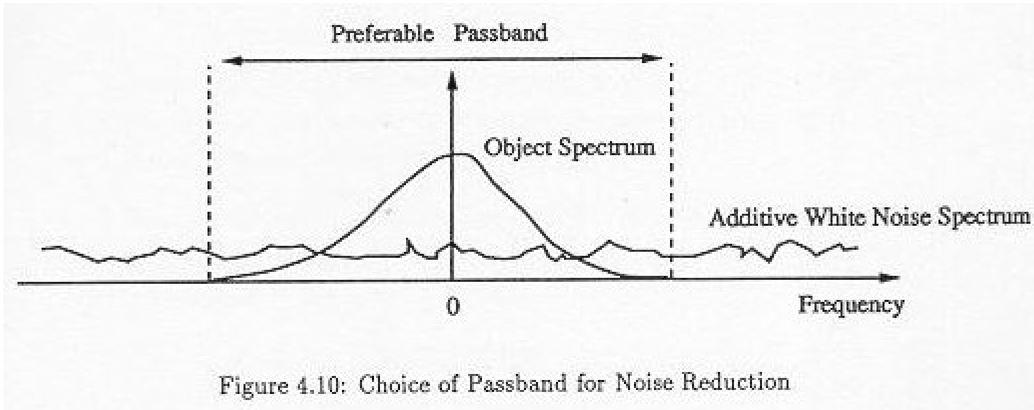

Figure 4.10: Choice of Passband for Noise Reduction

A trade off between the high spatial frequency components and the amount of noise let through in passing those frequencies should be made in calculation of an optimum passband for the phase-only filter. Several papers have been written on this subject [50], and it is mentioned here only in passing.

In computer simulations of correlations without input noise, the `signal to noise ratio' (SNR) defines noise as anything which is not part of the correlation peak but is present in the image. Horner, for example, uses a 50% of peak value threshold [33] to define what is noise in the image. Examination of filter performance without noise is useful in providing a clear, basic figure of comparison between filters. Table 4.3 again quotes from Horner the SNR and optical efficiency (as defined in section 4.1.2) for the classical matched filter, phase-only filter and the binary-phase-only filter. This time, only those results obtained from using both a binary object and the filters made from this object are tabled as this was the experimental situation realised in chapter seven of this thesis.

	MF	POF	BPOF
SNR	7:1	62:1	36:1
η_H	0.7%	19%	7.5%

Table 3: SNR and HORNER Efficiency, Three Types of Filter.

The signal to noise ratio, as might be expected, is best for the POF and not quite so good for the BPOF. Figures for both these filters however are very much better than for the classical matched filter even though it is operating as best it ever will, on a binary object.

The conclusions of this section are

1. Phase-only filters result in an enormous increase in peak intensity (and, though results are not shown here, peak sharpness also), and signal to noise ratio of the resulting correlation.

2. Binary phase-only filters have a reduced performance over their continuous counterparts but nonetheless are much superior to matched spatial filters with regards to the same parameters.

3. Optical efficiency - the fractional percentage of energy in the region of the peak to the total image plane energy - is much improved in both POFs and BPOFs over that obtained by matched spatial filtering.

These conclusions have been verified in numerous publications. It is not the intention of this section to review the very large field of optical phase-only correlation, rather to provide an introduction to the subject and quote the major results which allow the use of current binary spatial light modulators as phase filters.

4 Alternative Correlation Techniques

It should be said that there are other techniques available for performing optical correlation than the one described here. Most notable is a technique known as `Joint Transform Correlation', where both the target object and and input scene are displayed side by side in the object plane. The resulting *intensity* distribution in the frequency plane is an interference pattern between the individual Fourier Transforms of the input functions, and is used as input to a spatial light modulator. The input transmittance can be written as

$$t(x,y) = p(x,y+b) + h(x,y-b) \qquad (34)$$

where p(x,y) denotes the target function and h(x,y) the function to be correlated with. It can be shown [] that the image field contains the terms

$$[\ g(x,y)*h(x,y)\ ] ** \delta(x, y-2b) + [\ h(x,y)*g(x,y)\ ] ** \delta(x,y+2b) \qquad (35)$$

where * denotes correlation and ** convolution. As such, this system implements the classical Vander Lugt correlator system.

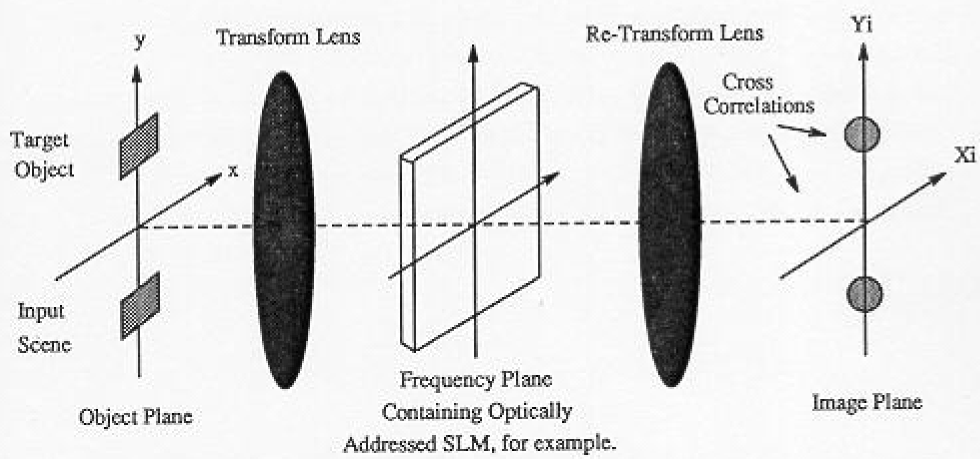

Figure 4.11: Joint Transform Correlator Configuration

Figure 4.11: Joint Transform Correlator Configuration

A most concise article on this subject is provided by Gregory and Loudin [51], who have successfully demonstrated this technique by using a modified liquid crystal television (LCTV) as the spatial light modulator. The LCTV provides the necessary continuous representation of the filter interference pattern, which binary mode devices clearly cannot. Several advantages of this technique over the `classical' one described in this thesis are presented, but perhaps most notable is the absence of computation required to find the filter pattern as it automatically appears as an interference pattern in the frequency plane. If one SLM is used to input the object scene and target, and another to display the intensity distribution of the resulting Fourier Transform patterns then correlation may be performed at video frame rates.

Although it is also possible to binarise the filter pattern (and thus make use of the binary mode devices available within the Applied Optics Group at Edinburgh) it was decided not to pursue this technique as a possible filtering algorithm for use in this device. As discussed in section 4.4.2 (space-bandwidth product), there exists a maximum object field size for most real spatial filters in current use. For the particular device used in this thesis (the subject of chapters 5 & 6) this size is very small indeed, so that the whole object space can readily accommodate only one (small) scene of limited resolution. This unfortunately means that the recent field of Joint Transform Correlation cannot be investigated experimentally. It is suggested that this form of correlator be investigated within this Group with SLMs having a much larger number of pixels, this directly increasing the space-bandwidth product.

4.1 Amplitude-Modulated Phase-Only Filter

Even more recently than the subject of Joint Transform Correlation, a new correlation filter has been proposed which combines both amplitude and phase modulation to give a significant improvement over even the phase-only

filter described in section 4.3.1.

Awwal et al [] propose an amplitude-modulated phase-only filter (AMPOF). If the target object spectrum is described by $G(\nu_x,\nu_y)$ which has a phase distribution characterised by $\Phi(\nu_x,\nu_y)$, the AMPOF is given by

$$H(\nu_x,\nu_y) = D \frac{e^{-i\,\Phi(\nu_x,\nu_y)}}{|G(\nu_x,\nu_y)| + a} \quad (36)$$

where `D' is a constant and `a' may be either constant or a function of frequency. This is a modified version of the (alternative) classical matched filter, the inclusion of `a' in the denominator ensuring the function does not blow up at very small (high frequency) values of $G(\nu_x,\nu_y)$. The effect of the denominator is to flatten out the amplitude of the spectrum immediately following the filter, so that each spatial frequency has a phase of zero and a very similar amplitude.

Table 4.4, taken from [52], compares the percentage energy in the correlation peak relative to the energy in the image plane as a whole, the normalised peak power (NPP), for the AMPOF and the POF, results quoted from AWWAL et al. The input and filter combinations are varied so as to compare both auto-correlation performance and filter discrimination for both filter types. The results were obtained from computer simulation using letters defined on a 64×64 object data grid. It is observed that the AMPOF results in much more energy being channeled into the correlation peak, and the percentage change in discrimination is much greater than for the POF.

Input	Filter	POF	AMPOF
E	E	57	134
F	F	62	131
G	G	70	163
0	0	55	144
E	F	41	81
F	E	44	83
G	O	52	105
O	G	44	99

Table 4: Normalised Peak Power for AMPOF and POF filters

The high performance obtained by using the AMPOF, as explained in their original paper, is roughly as follows. It can be shown [x] that the auto-correlation $C_{gg}(\Delta)$ of a function g(x) is given by

$$C_{gg}(\Delta) = \int_{-\infty}^{+\infty} |G(\nu)|^2\, e^{i\nu\Delta}\, d\nu \quad (37)$$

which is the Fourier Transform of $G(\nu)^2$ with respect to the variable Δ[6]. The central value of the auto-correlation ($\Delta = 0$) is then proportional to the mean value of $|G(\nu)|^2$, and as such a flatter spectral amplitude profile will have a larger mean value of $C_{gg}(0)$. The inclusion of $|G(\nu_x,\nu_y)|$ in the denominator serves to perform this operation. Although faultless, the author of this thesis suggests that the more physical interpretation as given under the heading `Choice of modulation parameter' of section 4.3.1, provides a better feel for what is going on.

At the time of publication, the authors intended to study the effects of binarisation on the filter performance. Although this type of filter is obviously very powerful, it relies on a filter capable of modulating both amplitude *and* phase. Cascading of two devices each operating in different modes (amplitude and phase) is possible though the alignment is liable to be an area of considerable experimental difficulty. This type of filter was not chosen for use with the SLMs available primarily because the project was already well underway before the publication of the original paper.

4.2 Review

This chapter has introduced the subject of optical correlation and the associated mathematics involved. The realisation of practical spatial filters is frequently constrained by the fact that only one parameter describing the complex light field (amplitude or phase) can be modulated by any one device. In summary,

1. **The general characteristics of practical `pixellated' filters have been introduced, the principle limitation of which is that the product of input object size with the highest spatial frequency passed by the device is a constant.**

2. **Where a single parameter only can be modulated in correlation filters, that parameter should be phase and in fact correlation filters employing phase-only modulation actually operate much better than the classical matched spatial filter.**

3. **Binarisation of the phase value is normally required by most phase modulating devices (SLMs), which is a special type of phase quantisation. However, quantisation of the filter phase to two states does reduce the correlation peak intensity among other things, but still provides a much improved correlation over the classical matched filter.**

4. **Alternative correlation techniques combining both amplitude *and* phase modulation offer much improvement over phase-only filtering but are unsuitable for this project.**

The subject of this chapter will form a background for the experimental work on optical correlation performed with a spatial light modulator designed within the Applied Optics Group at Edinburgh University. The design, operation and assembly of this device shall form the bulk of the next two chapters.

Footnotes:

[1] **The conjugate of the second function may be dropped if both functions are entirely REAL.**

[2] **Larger array sizes are also available to date though the 48×48 device was the first to become commercially available.**

[3] **The `sinc' function is defined here, according to Gaskill [17], as sinc(x)=[(sin(π x))/(π x)].**

[4] **The effects of noise on phase-only filters is considered in section 4.3.3**

[5] **A square pixel of side a has, if pixels are spaced d apart, a fill factor of order $([a/d])^2$ for instance.**

[6] **Indeed, this fact is fundamental to the operation of all matched correlator systems.**

File translated from T_EX by T_TH, version 3.04.
On 27 Oct 2001, 23:40.