Université de Neuchâtel
Institut de Microtechnique

Discrete incoherent optical correlator using diffraction gratings

Thèse

Présentée à la Faculté des sciences
pour obtenir le grade de docteur ès sciences

par

Giancarlo Pedrini

Imprimerie Le Franc-Montagnard S.A., 2726 Saignelégier
1990

Université de Neuchâtel

Institut de Microtechnique

Discrete incoherent optical correlator using diffraction gratings

Thèse

Présentée à la Faculté des sciences

pour obtenir le grade de docteur ès sciences

par

Giancarlo Pedrini

IMPRIMATUR POUR LA THÈSE

Discrete incoherent optical correlator
using diffraction gratings

de Monsieur Giancarlo Pedrini

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel
sur le rapport des membres du jury,

Messieurs R. Dändliker, F. Pellandini,
P. Chavel (Orsay) et R. Thalmann (Wabern)

autorise l'impression de la présente thèse.

Neuchâtel, le 4 septembre 1990

Le doyen:

Cl. Mermod

ABSTRACT

An optical correlator based on diffraction gratings is presented. The processor performs basically an incoherent, discrete correlation between an input data pattern and a predefined search symbol. The arrangement consists of a 4f system in which diffraction gratings are inserted behind the input pattern. Programmability of the search symbol is obtained by selecting the appropriate diffraction orders in the Fourier plane (spatial filtering). The described correlator can process input pattems of different encoding (intensity, polarization or spatial frequency). Some experimental results showing how this arrangement can be used as a pattern recognition system are reported.

The correlator can work with either completely or partially spatial coherent light. In fact, the use of partially coherent illumination turns out to be advantageous to obtain images of better quality, at the expense, however, of a certain loss of space bandwidth product.

The possibility to use various data encodings and the simplicity of programming make this processor suitable for use in digital optical computing (DOC), especially in symbolic substitution. The application of our correlator as recognition and substitution unit in a symbolic substitution processor is demonstrated. The optical system that we used to achieve symbolic substitution consists of the cascade of two discrete correlators separated by a nonlinear threshold device (we use a liquid crystal light valve). This symbolic substitution processor is well suited to implement the basic binary image algebra operations (erosion and dilation), especially for complex structuring elements. More complex morphological operations, like opening, closing, edge detection, noise suppression, can be performed using a combination of the basic operations. In this case we need a combination of symbolic substitution units. Some experimental results are reported to illustrate the capabilities of the system.

TABLE OF CONTENTS

1. INTRODUCTION

Optical systems have received considerable attention in various information processing applications since the 1960's. The attractive feature of optical systems is their capability to process two-dimensional signals in parallel, and this offers great potential in processing capacity and speed. A simple lens can be taken as an example of the capacity of optics to perform parallel processing. In fact, when it is used to form an image, it interconnects simultaneously every light ray coming from a point in the object plane, to the corresponding point in the image plane. In an object plane we can have millions of points that are conveyed in parallel to the image plane. Another advantage of optics is that the communications is non-interfering, this means that the rays can cross each other without interaction.

Image processing is in practice the manipulation of a given input image, to obtain a modified image. We can divide optical processors into two categories: coherent optical processors and incoherent optical processors. Coherent systems are linear in amplitude whereas incoherent systems are linear in intensity. In a coherent system the input and output functions represent amplitudes (complex functions) whereas for an incoherent system, intensity is processed (real positive functions).

The best known method for coherent optical processing is spatial filtering. This is based on the fact that a two dimensional Fourier transform relation exists between the front and the back focal plane of a lens. The spatial frequency spectrum of the input, (the input is a transparency placed in the front focal plane), is accessible in the back focal plane and can be modified. The experiments performed by Abbe in 1893 and later by Porter in 1906 provide a demonstration of an optical processor using the method of spatial filtering. The classical arrangement for a coherent spatial filtering system, uses two lenses in a confocal configuration. The first lens performs the Fourier transform of the input object (transparency modulated in amplitude and phase), that appear in the back focal plane (Fourier plane). The second lens performs the inverse Fourier transform so that in its back focal plane an inverted (possibly magnified) image of the input appears. By placing a filter in the Fourier plane we multiply the Fourier transform of the input object by the transmittance of the filter. In the back focal plane of the second lens we find the Fourier transform of this product that is equal to the convolution of the input mask with the Fourier transform of the filter (convolution theorem). The filter can be calculated to realize such operations as convolution, correlation, differentiation etc. The problem is that the spatial filter is generally complex-valued and this involve practical problem of realization. A.B. Vander Lugt [3] introduced a new technique, based on holography, for the realization of a complex-valued filter.

In an incoherent optical processor input and output need to be non-negative and real functions, and a Fourier relation cannot be used as in the coherent case. This limits the data-handling capacity of the processor. A proposed solution to this problem is to convert incoherent images into coherent images, and for this purpose many devices were developed that in general are very expensive. In certain cases it could be more convenient to process directly an incoherent image. Incoherent processors have some advantages compared with coherent processor. One of these advantages is given by the fact that usually the images that we would like to process are incoherent, for example a TV image or a live image. Another advantage of incoherent processing is that it is not troubled by the coherent noise produced by dust or scratches on the optical components, that can have a dramatical effect on the output.

The optical correlator, a processor that performs a correlation operation, has a particular place in the world of optical processing. The importance of correlation is derived by the fact that it specifies a relationship (it makes possible a comparison) between two functions. It is also useful in those aspects of pattern and character recognition that consist of determining the presence and location of a reference pattern in an input image. The correlation process of two functions involves multiplying, shifting and summing procedures. When the two functions are continuous the process can be represented mathematically by an integral. If one of the functions consists of a set of delta functions, we have a discrete sum, and we will call this a "discrete correlation".

An optical arrangement that allows us to perform a discrete correlation is presented in chapter 3. The arrangement consists of a 4f system in which diffraction gratings are inserted behind the input pattern. The system is based on shifting and superposition of multiple copies of an input image produced by diffraction gratings. One of the functions to correlate (this function can be continuous) is represented by the input image, whereas the function consisting of delta pulses is defined in the Fourier plane using spatial filtering. We will show that the illumination can be coherent or partially coherent. In spite of this, the processor has to be considered as incoherent because it is linear in intensity and not in complex amplitude. A characteristic of the proposed correlator is given by the possibility to use various data encoding techniques for the functions which are being correlated: intensity encoding, polarization encoding, spatial frequency encoding (theta modulation). For these reasons a large number of the existing spatial light modulators (SLM) can be used as an input mask and as spatial filters in Fourier plane. The system is space invariant, which means that a change of the input position involves a change in the output location without changing its functional form. Some experimental results are given in chapter 5, in which it is shown how this arrangement can be used as a pattern recognition system.

The possibility to use various data encodings and the simplicity of programming make this processor suitable to use in digital optical computing (DOC), especially in symbolic substitution. Symbolic substitution, as first proposed by Huang [5], is a concept for performing parallel and space invariant logic operations on an input pattern (usually binary intensity or binary polarization encoded). It consist of recognizing all occurrences of a set of search symbols and in substituting them by other predefined symbols. All symbolic substitution processors can be considered as a cascade of correlators. The application of our correlator as recognition and substitution unit in a symbolic substitution processor was demonstrated [10]. The optical system that we used to achieve symbolic substitution consists of the cascade of two discrete correlators separated by a nonlinear thresholds device (we use a liquid crystal light valve). The first correlator performs the recognition, the nonlinear device restores the output pattern of the recognition unit to binary values, and the second correlator performs the substitution. We have shown the possibility of recognizing and substituting in parallel different symbols (multiple channel symbolic substitution) and this was used to build a processor that can perform addition and subtraction in one cycle [11].

In section 5.3 is presented an application of symbolic substitution to mathematical morphology. Mathematical morphology is a method for image analysis which has increased in importance in recent years. Using electronic image processing systems, it has been successfully applied, for example, for edge detection, noise suppression, and smoothing. The theoretical foundations of mathematical morphology were developed essentially by Serra and Matheron [12] and are actually well established. Binary image algebra is a special case of morphological transformation where input and output are binary discrete images, represented by 2-D sets in plane Z^2. In a recent paper [13], Goodman and Rhodes show how optical symbolic substitution can be used to perform the morphological operations. The symbolic substitution processor, which has been described in [10], is well suited to implement the basic binary image algebra operations (erosion and dilation), especially for complex structuring elements. More complex morphological operations, like opening, closing, edge detection, noise suppression, can be performed using a combination of the basic operations. In this case we need a combination of symbolic substitution units. Some experimental results are reported to illustrate the capabilities of the system.

2. DISCRETE CORRELATION: GENERAL CONCEPTS AND OPTICAL IMPLEMENTATION

2.1. General Introduction to discrete correlation

In this section we describe an operation that we will call "discrete correlation". This correlation operation is not between continuous functions but between discrete functions, and the result is a discrete function as well. A mathematical description of this operation will be given in section 3.1, and we will see that it is a discrete sum and not an integral. We must emphasize that this concept is not new, and we cannot say who used it at first.

A simple example will introduce the concept and show at the same time how it can be applied to perform the recognition and location of a certain distribution of pixels (symbol) in a discrete binary encoded input pattern. Our interest in discrete correlation is in the application for pattern recognition and during all this work it will be used for this purpose only.

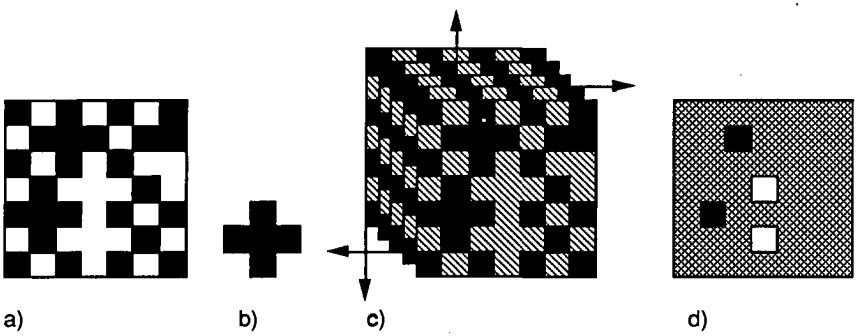


Fig.2.1. Illustration of discrete correlation. a) input pattern, b) search symbol, c) five copies of the input pattern, d) output obtained by superposition of shifted copies.

Figure 2.1a) shows a binary input pattern composed by black and white square pixels. We assume that black pixels have logical value "0" and white pixels have logical value "1". The problem which we would like to solve, is to recognize a given symbol inside the input pattern. In the example we will recognize a cross composed by five dark pixels (search symbol of Fig.2.1b)). The symbol to be recognized is defined relative to an origin, in our example we choose the origin to be the central pixel of the cross, that we call the reference pixel. The first step is to divide the input pattern into as many copies as there are pixels in the search symbol. In our case we produce also five copies, the pixel which in the input pattern had value "1", will have value "1/5" in each copy (Fig.2.1c)). The next step consist in the shift and superposition of the five copies. We associate the first copy to the central pixel (reference pixel) and we call it

the reference copy, the second copy is associated to the left pixel and it is shifted by one pixel right with respect to the reference copy (the shifting direction is indicated by the arrow), the third copy is associated to the right pixel and is shifted by one pixel left, the fourth copy is associated to the top pixel and is shifted by one pixel down, and the fifth copy is associated to the bottom pixel and shifted by one pixel up with respect to the reference copy. The output pattern is obtained by summation of the values of the overlapping pixels (Fig.2.1.d)). The output contains six brightness values "0", "1/5", "2/5", "3/5", "4/5", "1". In Fig.2.1d), the values "1/5", "2/5", "3/5", "4/5", are drawn in the same manner. It is easy to see, that each dark ("0") pixel in Fig.2.1d), corresponds to the presence of the search symbol (Fig.2.1b)) in the input pattern. Analogously, the presence of a bright pixel ("1") in the output indicates that in the input pattern we have a bright search symbol, this means the binary complement of the symbol of Fig.2.1b).

Using the same method, it is possible to recognize more complex symbols (symbols composed of many pixels) present in a binary input pattern. To achieve this, it is necessary to divide the input pattern into many copies that will be superimposed with appropriate shifts in the output plane, we can also imagine to recognize symbols composed of 100, 1000 ...pixels. Since each image can be seen as composed of small pixels, we can say that using this method we can recognize each symbol in a binary input pattern.

This kind of recognition can be interesting when performed in parallel, this means that all the pixels have to be processed at the same time. We know that optics is well suited to process two dimensional patterns in parallel. A binary input can be easily implemented with a spatial light modulator or a photographic plate having transparent and opaque pixels. Splitting of an image can be performed using beam splitters or gratings, and the shift operation can be performed e.g. using mirrors. Another feature of optics is that when two or more incoherent images are superimposed, the resulting intensity is the sum of the intensities of each single image. Optics seems also well suited to perform the operation described above.

An example of an optical implementation for the recognition of a search symbol composed of two pixels can be found in Ref.6. The set-up used is essentially a Michelson interferometer, in which the beamsplitter produce two copies of the input pattern, and the tilt of the mirrors determines the shifts of the two images with respect to each other.

In the next section we describe an optical arrangement that can perform a discrete correlation operation, where the splitting of the input pattern into many copies is produced by diffraction gratings.

2.2. Optical implementation of discrete correlation operation

To perform the discrete correlation operation described in section 2.1, we propose the optical arrangement sketched in Fig.2.2. We described this arrangement at first in Ref.10, where it was used as a recognition unit for a symbolic substitution processor. During this work, we will give a complete analysis of this optical system, and we will present some applications.

It is essentially a 4f Fourier filtering system in which a diffraction grating is inserted. A transparency, representing the input function, is illuminated with a monochromatic plane wave of wavelength λ (in sections 3.5 and 3.6 are discussed the coherence requirements for the illumination beam). A two-dimensional diffraction grating splits the input data pattern into different equal diffraction orders to produce many copies of the input. The distance *a* between the grating and the input pattern determines the relative shift of the multiple images in the output plane. If the period *d* of the grating is much smaller than the pixel size of the input mask, the spatial frequency spectrum of the mask is split into multiple spectra in the Fourier plane, corresponding to the diffraction orders of the gratings. Through spatial filtering, the different diffraction orders are multiplied by a transmittance a_{mn}, in this manner it is possible to suppress some of the produced copies. The system does not perform exactly the operation described in section 2.1, in fact it must produce a number of copies equal or greater than the number of pixels in the search symbol and it must suppress the undesired copies by spatial filtering. This has the disadvantage that we loose power, because we create copies that we do not need, but it has the advantage that the search symbol can be easily programmed by introducing a spatial light modulator in the Fourier plane. The result of the correlation operation is given by the intensity in output plane (x_o,y_o), in which the shifted copies are superimposed. We will show that the result obtained is dependent on the spatial coherence of the illumination.

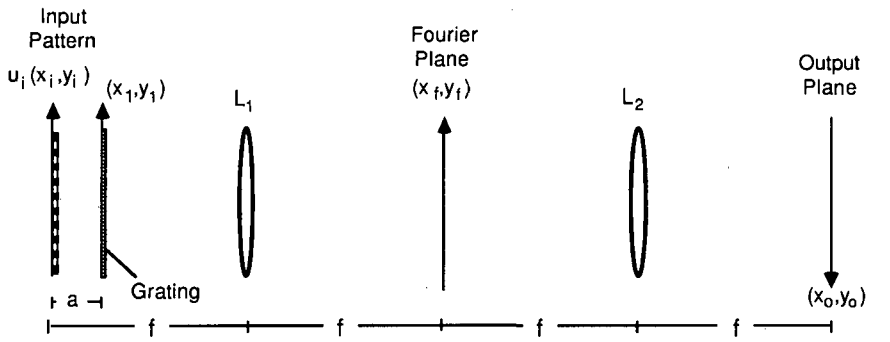


Fig.2.2. Optical arrangement for discrete correlator.

3. ANALYSIS OF THE DISCRETE INCOHERENT OPTICAL CORRELATOR

This chapter reports the analysis of the optical arrangement presented in section 2.2. Section 3.1 is a mathematical introduction to the discrete correlation operation. The calculations reported in section 3.2, will show that this operation can be optically implemented using the proposed arrangement. In the other sections of the chapter we will discuss the encoding of the input mask, the illumination of the input and the SBWP of the system.

3.1. Mathematical description of discrete correlation

We begin with the definition of two integral transforms, that we will use during all this work. The convolution operation between two 2-dimensional functions f(x,y) and g(x,y) is mathematically described by

$$f(x,y)*g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-x',y-y')g(x',y')dx'dy' \quad . \tag{3.1}$$

The physical interpretation of convolution is quite simple. The shifted and folded function f(x–x',y–y') is multiplied by g(x',y') and the "volume" under the "surface" f(x–x',y–y')g(x',y') gives the value of f(x,y) * g(x,y).

The correlation between f(x,y) and g(x,y) is described by

$$f(x,y) \star g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x'+x,y'+y)g^*(x',y')dx'dy' \quad . \tag{3.2}$$

* indicates a complex conjugation. The relation between correlation and convolution can be summarized as follows

$$f(x,y) \star g(x,y) = f(x,y) * g^*(-x,-y). \tag{3.3}$$

We discuss now a particular case in which one of the functions to correlate (or to convolve) is represented by an array of its sampled values taken on a discrete set of points in the xy-plane. We choose the points on a 2-dimensional cartesian grid (mΔ,nΔ), where Δ is the period of the grid and (m,n) are integer numbers. Thus we approximate the function g(x,y) by its discrete sampled function

$$g_d(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{mn}\delta(x - m\Delta, y - n\Delta). \tag{3.4}$$

a_{mn} are the values of g(x,y) at the points (mΔ,nΔ). In this work we will use only the discrete form of correlation, for this reason we discuss only this case, but it is clear that the same discussion can be done for discrete convolution. Thus by substituting g(x,y) by $g_d(x,y)$ in Eq.(3.2) one gets

$$f(x,y) \star g_d(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x'+x,y'+y) \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} (a_{mn})^* \, \delta(x'-m\Delta,y'-n\Delta)\, dx'dy', \tag{3.5}$$

and using the property of the delta function

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x'-m\Delta,y'-n\Delta) f(x'+x,y'+y)\, dx'dy' = f(x+m\Delta,y+n\Delta) \tag{3.6}$$

we obtain the simple relation

$$f(x,y) \star g_d(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} (a_{mn})^* \, f(x+m\Delta,y+n\Delta). \tag{3.7}$$

This equation describes a superposition of shifted copies of the function f(x,y). When the points (mΔ,nΔ) are a densely packed array (Δ small), we can approximate the correlation relation for continuous functions (Eq.(3.2)), by superimposing a large number of shifted copies of f(x,y).

We consider now the case in which f(x,y) is a set of elementary cells (pixels) located on the cartesian grid (mΔ,nΔ). Thus we can write

$$f_d(x,y) = \sum_{m'=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} b_{m'n'} \, f_0(x - m'\Delta, y - n'\Delta)\,, \tag{3.8}$$

where $b_{m'n'}$ is the value corresponding to the pixel located in (m'Δ,n'Δ). $f_0(x,y)$ is the function representing a pixel with dimension ΔxΔ located in (0,0), it has the value 1 inside the pixel and 0 elsewhere. The delta pulses of $g_d(x,y)$ and the pixels of $f_d(x,y)$ are located on the same grid, the correlation function $f_d(x,y) \star g_d(x,y)$ and the input pattern $f_d(x,y)$ have also the same pixel structure. Introducing Eq.(3.8) in Eq.(3.7) we obtain

$$f_d(x,y) \star g_d(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \left\{ (a_{mn})^* \sum_{m'=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} b_{m'n'} f_0(x+m\Delta-m'\Delta, y+n\Delta-n'\Delta) \right\}. \quad (3.9)$$

We can better understand this equation, if we look at the value of a single pixel of $f_d(x,y)\star g_d(x,y)$, located at position $(k\Delta,j\Delta)$, where k and j are integer numbers. The pixel function $f_0(k\Delta+m\Delta-m'\Delta, j\Delta+n\Delta-n'\Delta)$, take the value 1 only when $\Delta(k+m-m') = 0$ and $\Delta(j+n-n') = 0$, this involves also $m' = m+k$, $n' = n+j$. Using this substitution in Eq.(3.9), we find the discrete form of the correlation operation

$$f_d(k\Delta,j\Delta)\star g_d(k\Delta,j\Delta) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} (a_{mn})^* b_{m+k,n+j}. \quad (3.10)$$

The operation introduced with the example of Fig.2.1 consists of a superposition of shifted images and can be mathematically described using Eq.(3.7). The input pattern was a binary image composed of square pixels, which can be mathematically represented by Eq.(3.8), where the coefficients $b_{m',n'}$, have binary values 0 and 1. The result (Fig.2.1d)) can thus be described by the correlation operation of Eq.(3.10). The recognition operation is a discrete correlation between a function describing the input pattern, and another function consisting of delta pulses (by five delta pulses in the example of Fig.2.1).

3.2. Rigorous general calculation of the optical system

In this section we show that the optical arrangement presented in section 2.2 can be used to perform a discrete correlation operation .

The complex amplitude in the output plane of the optical arrangement presented in section 2.2, will be rigorously calculated. For this calculation, we suppose that the input mask is illuminated with a monochromatic plane wave of wavelength λ (coherent illumination). We will show in section 3.3 that the input transparency can be coded in intensity, polarization or in spatial frequency. In the general case, the complex amplitude is thus represented by a vectorial function $\mathbf{u}(x,y)$, (vectors are denoted with bold characters).

For the Fourier transform of a two dimensional vectorial function $\mathbf{f}(x,y)$ we use the notation

$$FT\{\mathbf{f}(x,y)\} = \mathbf{F}(p_x,p_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{f}(x,y)\, \exp\left(-2\pi i(xp_x + yp_y)\right) dxdy. \quad (3.11)$$

Consider the configuration depicted in Fig.2.2. A monochromatic plane wave propagating in z direction is incident on the input mask, thus just behind the mask we will have a complex

amplitude $\mathbf{u}_i(x_i,y_i)$. The complex amplitude in front of the grating, $((x_1,y_1)$–plane), is given by Fresnel diffraction

$$\mathbf{u}_1(x_1,y_1) = \frac{e^{2\pi ia/\lambda}}{i\lambda a}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{u}_i(x_i,y_i)\exp\left(\frac{2\pi i[(x_1-x_i)^2+(y_1-y_i)^2]}{2\lambda a}\right)dx_i dy_i\,, \tag{3.12}$$

which can be written as a convolution (see Eq.3.1), thus

$$\mathbf{u}_1(x_1,y_1) = \frac{e^{2\pi ia/\lambda}}{i\lambda a}\,\mathbf{u}_i(x_1,y_1) * \exp\left(\frac{2\pi i(x_1^2+y_1^2)}{2\lambda a}\right). \tag{3.13}$$

If $c(x_1,y_1)$ is the complex transmission of the grating, just behind the grating the complex amplitude is $\mathbf{u}_1(x_1,y_1)c(x_1,y_1)$. Aside from a multiplicative factor, the complex amplitude in the Fourier plane (x_f,y_f) is given by the Fourier transform of $\mathbf{u}_1(x_1,y_1)c(x_1,y_1)$, evaluated at the spatial frequencies $p_x = x_f/\lambda f$ and $p_y = y_f/\lambda f$. The multiplicative factor is

$$\frac{1}{(i\lambda f)}\exp\left(\frac{\pi ia(x_f^2+y_f^2)}{\lambda f^2}\right) \equiv \frac{1}{(i\lambda f)}\exp\left(\pi i\lambda a(p_x^2+p_y^2)\right). \tag{3.14}$$

The quadratic phase factor appears because $\mathbf{u}_1(x_1,y_1)c(x_1,y_1)$ is not in the front focal plane of the lens L_1 but at distance a behind the input mask.

We calculate at first the Fourier transform of $\mathbf{u}_1(x_1,y_1)$. The convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of their individual transform, thus

$$FT\{\mathbf{u}_1(x_1,y_1)\} = \frac{e^{2\pi ia/\lambda}}{i\lambda a}\,FT\{\mathbf{u}_i(x_1,y_1)\}\,FT\{\exp\left(\frac{2\pi i(x_1^2+y_1^2)}{2\lambda a}\right)\}. \tag{3.15}$$

We know that $FT\{\exp(ix^2)\} = (1+i)\sqrt{\pi/2}\,\exp(-i(\pi p)^2)$ [14]. Thus after some calculations and after omission of the phase term $e^{2\pi ia/\lambda}$, we obtain

$$FT\{\mathbf{u}_1(x_1,y_1)\} = \mathbf{U}_i(p_x,p_y)\exp\{-i\lambda a\pi\,(p_x^2+p_y^2)\}. \tag{3.16}$$

Note that in the following calculations the phase factors depending only on z (propagation direction) are omitted.

From the convolution theorem it follows that the Fourier transform of a product of two functions is equal to the convolution of their Fourier transforms. The complex amplitude $U_f(p_x,p_y)$ in the spatial frequency plane can also be written as

$$U_f(p_x,p_y) = \frac{1}{(i\lambda f)} \exp\left(\pi i\lambda a(p_x^2+p_y^2)\right) FT\{u_1(x_1,y_1)c(x_1,y_1)\}$$
$$= \frac{1}{(i\lambda f)} \exp\left(\pi i\lambda a(p_x^2+p_y^2)\right)\left\{\left[U_i(p_x,p_y)\exp\left(-i\pi\lambda a(p_x^2+p_y^2)\right)\right] * C(p_x,p_y)\right\}. \quad (3.17)$$

Let us consider the case of a two-dimensional grating with transmission $c(x_1,y_1)$ which produces the splitting of the input into NxN different diffraction orders with the same intensity. Let the spacings (period of the grating) be *d* for both x and y directions. The Fourier transform of such a grating is a sum of delta functions, described by

$$C(p_x,p_y) = \frac{1}{N}\sum_{n=-K}^{K}\sum_{m=-K}^{K}\delta(p_x-m/d,p_y-n/d). \quad (3.18)$$

The above summation is from –K to K, where K = (N–1)/2. m/d and n/d denote the spatial frequencies of the (m,n) diffraction order.

By inserting Eq.(3.18) in Eq.(3.17), it appears clear that the spectrum of the input mask is split into multiple spectra in the Fourier plane, corresponding to the diffraction orders of the grating. If the period of the grating is smaller than the size of the cell in the input mask ($d<\Delta$), it is possible to filter independently each diffraction order. Through spatial filtering, the different diffraction orders are multiplied by an amplitude transmittance A_{mn}. A is a discrete function which has the constant value A_{mn} within the diffraction halo of the input pattern around the diffraction order (m,n) of the grating. Using Eqs.(3.17–18) we obtain for the complex amplitude behind the filter

$$U_f(p_x,p_y) = \frac{1}{i\lambda f N}\sum_{n=-K}^{K}\sum_{m=-K}^{K}\left\{A_{mn}U_i(p_x-m/d,p_y-n/d)\exp\{2\pi ia\lambda(p_x m/d+p_y n/d)\}\right.$$
$$\left.\times\ \exp\{-\pi ia\lambda((m/d)^2+(n/d)^2)\}\right\}. \quad (3.19)$$

Some phase factors depending on the diffraction orders appear in Eq.(3.19). $\exp\{2\pi ia\lambda(p_x m/d+p_y n/d)\}$ is a linear factor accounting for the fact that the diffraction orders arrive on Fourier plane with different incidence angles. $\exp\{-\pi ia\lambda((m/d)^2+(n/d)^2)\}$ is a phase factor depending only on m/d and n/d, it is thus constant inside the (m,n) order. If the different diffraction orders do not overlap, the intensity behind the filter is given by

$$|U_f(p_x,p_y)|^2 = \frac{1}{(\lambda f N)^2} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \{ |A_{mn}|^2 |U_i(p_x - m/d, p_y - n/d)|^2 \}. \qquad (3.20)$$

If there is overlapping of the diffraction orders, an interference term appears in Eq.(3.20). Fig.3.1 shows, how the spatial filtering is performed in the Fourier plane.

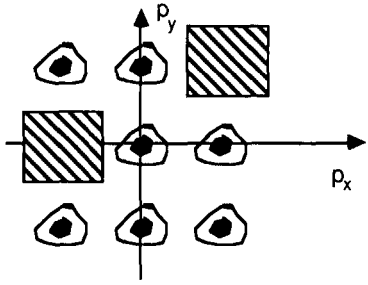


Fig.3.1. Example of filtering of the diffraction orders in the Fourier plane.(A_{11}=A_{0-1}=0).

The complex amplitude in the output plane (x_o,y_o) is given by the Fourier transform of $U_f(p_x,p_y)$, multiplied by the factor $i\lambda f$. (Note that our coordinate system in the image plane is reversed (Fig.2.2), which allows to use Fourier transform rather than inverse Fourier transform to calculate the output). To perform the Fourier transform of Eq. (3.19) we use the convolution theorem, the shift theorem and the properties of delta functions, and after some manipulations of the phase factors we obtain

$$u_o(x_o,y_o) = \frac{1}{N} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \{ A_{mn}\, u_i(x_o+\lambda a m/d,\, y_o+\lambda a n/d)\ \exp\{2\pi i(x_o m/d + y_o n/d)\} \times \exp\{\pi i a \lambda[(m/d)^2+(n/d)^2]\} \}. \qquad (3.21)$$

The different images of u_i in (x_o,y_o) are shifted by $(\xi_m, \eta_n) \equiv (\lambda a m/d, \lambda a n/d)$. The first, linear phase factor in Eq.(3.21) is due to the angles of incidence $(\alpha_m,\alpha_n) \equiv (\lambda m/d, \lambda n/d)$ of the different images. The second, constant phase factor expresses the fact that the optical path lengths for the superimposed images are different.

If we denote with $I_i(x_i,y_i) \equiv |u_i(x_i,y_i)|^2$ the intensity in plane (x_i,y_i), then the total intensity in plane (x_o,y_o) can be written as

$$I_o(x_o,y_o) = |u_o(x_o,y_o)|^2 = \frac{1}{N^2} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \{ |A_{mn}|^2\, I_i(x_o+\xi_m,\, y_o+\eta_n) \} + I_t(x_o,y_o). \qquad (3.22)$$

The first term describes the discrete correlation operation of Eq (3.7). The shift ξ_m,η_n, depends on the wavelength λ, the spatial frequencies of the grating m/d, n/d and the distance a. $I_i(x_o,y_o)$ is an interference term produced by the diffraction orders arriving at the output plane with different angles of incidence. We would like to suppress this interference term, because it prevents us from having a pure correlation. We will see that this can be achieved by using partially coherent light, this will be discussed in section 3.5 and 3.6 when we will treat the coherence requirements. Another solution is to make the interference fringes so close together that the detection system cannot resolve them, in chapter 5 we will present some examples. For the moment we will simply omit this term and we write Eq.(3.22) as a discrete correlation between the intensity $I_i(x_o,y_o)$ and a function composed of a set of delta pulses

$$I_o(x_o,y_o) = I_i(x_o,y_o) \bigstar g_d(x_o,y_o)\,, \tag{3.23}$$

where

$$g_d(x_o,y_o) = \frac{I}{N^2}\sum_{n=-K}^{K}\sum_{m=-K}^{K} |A_{mn}|^2\delta(x_o-\xi_m,y_o-\eta_n) \tag{3.24}$$

indicates a sum of delta pulses located on the grid $(\xi_m,\eta_n) \equiv (\lambda am/d,\lambda an/d)$, multiplied by $|A_{mn}|^2$. Notice that by substituting $|A_{mn}|^2$ with N^2a_{mn}, we obtain the function g_d described by Eq.(3.4).

We consider now the input pattern as composed by pixels located on the grid $(m\Delta,n\Delta)$. The intensity in the input plane can thus be written in a form like that of the function $f_d(x,y)$ of Eq.(3.8). The distance a between the diffraction grating and the input mask can be chosen such that the image shifts correspond to the pixel spacings of the input pattern, this means $(\xi_m,\eta_n)=(m\Delta,n\Delta)$, which yields the condition

$$a = \frac{\Delta d}{\lambda}\,. \tag{3.25}$$

In this case $I_o(x_o,y_o)$ (result of Eq.(3.23)) has the same pixel structure as the input pattern $I_i(x_o,y_o)$ (the pixels are located on a cartesian two-dimensional grid of period Δ). The values of the (k,j) pixels of $I_o(x_o,y_o)$ can thus be expressed by Eq.(3.10).

All these calculations show that the discrete correlation operation described mathematically in section 3.1 can be performed optically using the arrangement depicted in Fig.2.2. The correlator works on the principle of the sum of optical intensities. It is also linear in intensity and must be considered incoherent. One of the functions to correlate is defined in the input plane and the other is defined in the Fourier plane. Both functions are easily programmable

using spatial light modulators (SLM's). Since only the intensity plays a role, the SLM's do not need to have high optical quality. These features will be discussed in detail in chapter 4.

We will now see which result is obtained when a second grating, having transmittance c'(x,y), is introduced at a distance a' behind the input pattern. We will discuss later that there exist good reasons to introduce a second grating. The calculation of the system with two diffraction gratings is quite complex, for this reason we give here only the most significant results.

The first diffraction grating located at distance a behind the input mask has NxN diffraction orders and its Fourier transform is described by Eq.(3.18). The second grating located at distance a' behind the input pattern has N'xN' diffraction orders and its Fourier transform is given by

$$C'(p_x,p_y) = \frac{1}{N'} \sum_{n'=-K'}^{K'} \sum_{m'=-K'}^{K'} \delta(p_x-m'/d',p_y-n'/d') \ , \tag{3.26}$$

where K'= (N'–1)/2. If a transparency $u_i(x_i,y_i)$ is introduced in such a system, its spectrum is split in NxNxN'xN' spectra in the Fourier plane. If the periods of the gratings are smaller than the size of the cells in $u_i(x_i,y_i)$, it is possible to filter each order independently. Thus the complex amplitude in the Fourier plane after filtering is described by the equation

$$U_f(p_x,p_y) = \frac{1}{i\lambda f(NN')} \sum_{n'=-K'}^{K'} \sum_{m'=-K'}^{K'} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \{ A_{mnm'n'} U_i(p_x-m/d-m'/d',p_y-n/d-n'/d') \times PH_{mnm'n'} \}. \tag{3.27}$$

$A_{mnm'n'}$ denotes the strength of the (mnm'n') diffraction orders after filtering. $PH_{mnm'n'}$ represents a phase factor depending on (p_x,m/d,m'/d',p_y,n/d,n'/d',a,a') and accounts for the fact that the diffraction orders arrive with different incidence angles on plane (x_f,y_f), and that each order have a multiplicative phase factor. The Fourier transform of (3.27) gives the complex amplitude distribution in image plane, and from this it is possible to calculate the intensity, which after omission of an interference term becomes

$$I_o(x_o,y_o) = \frac{1}{(NN')^2} \sum_{m'=-K'}^{K'} \sum_{n'=-K'}^{K'} \sum_{n=-K}^{K} \sum_{m=-K}^{K} |A_{mnm'n'}|^2 I_i(x_o+\lambda am/d+\lambda a'm'/d',y_o+\lambda an/d+\lambda a'n'/d') \tag{3.28}$$

The shift in the image plane depends on the locations of the gratings (distances a and a'), on their spatial frequencies, and on the wavelength λ. We note that using two diffraction gratings it is possible to increase dramatically the number of diffraction orders, and this increases the

number of superimposed images in the output plane. We chose a=Δd/λ (Eq.(3.25)) and a'= NΔd'/λ. Equation (3.28) can thus be written as

$$I_o(x_o,y_o) = \frac{1}{(NN')^2} \sum_{m'=-K'}^{K'} \sum_{n'=-K'}^{K'} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \left\{ |A_{mnm'n'}|^2 \, I_i(x_o+(m+m'N)\Delta,\, y_o+(n+n'N)\Delta) \right\}, \tag{3.29}$$

(m+m'N) and (n+n'N) are integer numbers that determine the shift of the (m,m',n,n') copy in the output plane. The result of Eq.(3.29) can be obtained using a unique grating having period *d*, and located at distance *a* behind the input pattern, but in this case the grating must have (NxN')x(NxN') diffraction orders. Since gratings having many orders are practically more difficult to produce (cf. section 4.1), it is in general suitable to use more gratings having few diffraction orders.

The second diffraction grating can be used to build a multiple channel correlator. In fact the distance *a'* can be chosen in order to produce spatially separated images in the output plane (Fig.3.2). In this case the shift must be greater than the dimension of the input that we call Y, thus *a'* and *d'* have to be chosen in order to satisfy the relation a'λ/d' > Y.

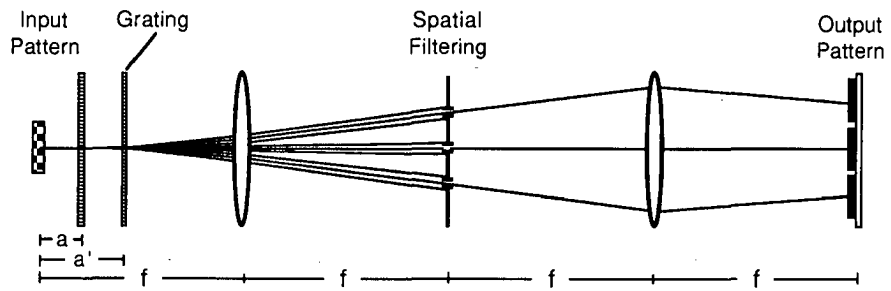


Fig.3.2. Multiple channel correlation system for 3 channel operation.

In this manner it is possible to correlate the input pattern with different functions (one function for each channel) at the same time. The different functions are chosen by spatial filtering in the Fourier plane.

3.3. Data encoding and corresponding spatial filtering in Fourier plane

With the optical system shown in section 2.2, it is possible to superimpose a large number of shifted copies of the input. This is equivalent to perform the discrete correlation described by Eq.(3.7). In the general calculation of the system we did not say anything about the encoding of the input pattern. We show now that the pixels of the input mask can be encoded in different ways, and that the filtering in the Fourier plane is dependent on this. In what follows, we suppose that the period d of the diffraction grating is much smaller than the pixel size Δ, the orders in Fourier plane are spatially separated, and we can introduce a filter A_{mn}.

3.3.1. Intensity encoding

We discuss at first the case in which the input mask is a set of elementary cells with different transmission placed on a cartesian grid. Figure 3.3 shows a binary encoded and a multilevel encoded input mask.

a) b)

3.3. Intensity encoded input masks. a) binary encoding, b) multilevel encoding.

The filter in Fourier plane is represented by a transparency having different transmission $|A_{mn}|^2$ for each diffraction order. We choose $a = \Delta d/\lambda$ (Eq. (3.25)), thus the image shift corresponds to the pixel spacing Δ. The intensity in the image plane is the discrete correlation given by Eqs.(3.22-23) which we rewrite omitting the interference term as

$$I_o(x_o,y_o) = \frac{1}{N^2} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \left\{ |A_{mn}|^2 \, I_i(x_o + m\Delta x,\, y_o + n\Delta y) \right\} = I_i(x_o,y_o) \bigstar g_d(x_o,y_o). \quad (3.30)$$

We have shown that this equation can be seen as a correlation between the intensity in the input plane $I_i(x,y)$ and a function $g_d(x,y)$ consisting of delta pulses with multipling coefficients $|A_{mn}|^2$.

The general case, in which the pixels of the input mask and the filter have a multilevel transmittance, is not very interesting, in fact we do not see any application where the calculation

of a discrete correlation between two multilevel functions could be used. The most common application is where I_i and A_{mn} are both binary functions. In fact, in this case the correlation can be used for pattern recognition.

We present now one experimental result obtained with intensity encoded input masks. The masks used are physically represented by photographic glass plates. In chapter 4 we will describe more in detail how the input pattern can be implemented. The input contains only five transparent square pixels arranged to form a cross. Figure 3.4a) shows a logical representation of the input pattern (middle figure) and a photograph of the input mask taken from a TV monitor (right figure), which displays at the same time a profile through the pattern. Using two crossed phase gratings, having each three diffraction orders (NxN = 3x3) with the same intensity, it was possible to obtain the correlation of the input pattern with a set of delta pulses. (The considerations about the features and the fabrication of the phase gratings are reported in section 4.1.)

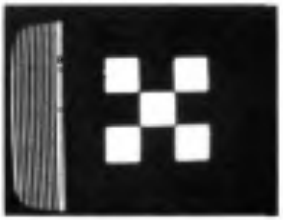
Figure 3.4b) shows the result of a correlation of the input with a function $g_d(x,y)$ composed by nine delta pulses located in the centers of the pixels of the input pattern. The left side of the figure shows a sketch of the position of the nine delta pulses of the function $g_d(x,y)$, the symbol "*" means that a delta pulse is present. In the middle figure we find a theoretical calculation of the correlation and in right part of the figure the experimental result. (Note that all the calculated results reported in the figure must be divided by a factor $N^2 = 9$). When the input pattern is correlated with a set of delta pulses located in the centers of the bright pixel of the input, we obtain the result of Fig.3.4c). In both examples (b and c) the central pixel is brighter than the others and has the value 5/9. Figure 3.4d) shows another example in which we correlate the input pattern with a set of delta pulses located in the centers of the dark pixels. In this case we obtain a result where we have a dark pixel in the middle of the correlation plane.

The result of a discrete correlation between a binary input and a function g_d having binary coefficients A_{mn} can be used for pattern recognition. From Eq.(3.30) follows that an absolute maximum (correlation peak) having value n/N^2 (n is the number of "1" pixels in the input pattern) appears when the pixels of the input pattern match with the values A_{mn} of g_d. The presence of the absolute maximum is a necessary but not sufficient condition for recognition, as the results of Fig.3.4b) and c) demonstrate. In fact the correlation peak says only that all the "1" pixels of the input match with the "1" coefficient of g_d. The correlator yields a dark peak where the "0" pixels of the input match with the coefficient of g_d having value 0.

With the arrangement, as it has been described so far, using intensity coded binary patterns only, either dark or bright symbols can be recognized. The recognition of symbols consisting of dark and bright pixels is not possible with the binary correlator thus far described.

a)

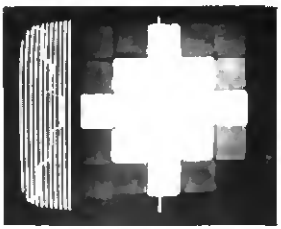
1	0	1
0	1	0
1	0	1



b)

*	*	*
*	*	*
*	*	*

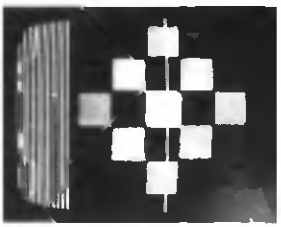
1	1	2	1	1
1	2	3	2	1
2	3	5	3	2
1	2	3	2	1
1	1	2	1	1



c)

*		*
	*	
*		*

1	0	2	0	1
0	2	0	2	0
2	0	5	0	2
0	2	0	2	0
1	0	2	0	1



d)

	*	
*		*
	*	

0	1	0	1	0
1	0	3	0	1
0	3	0	3	0
1	0	3	0	1
0	1	0	1	0

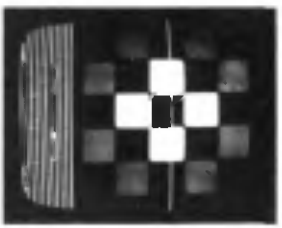


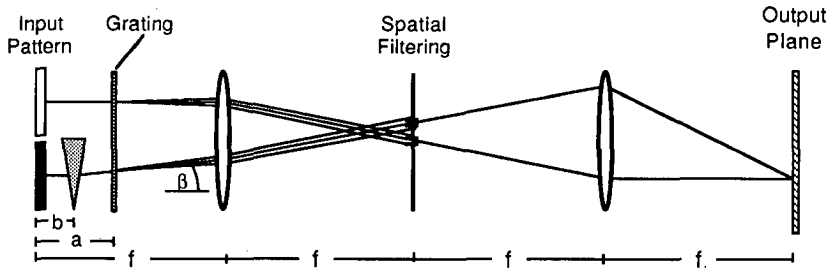
Fig.3.4. Examples for binary correlations of the input pattern a) with different discrete functions represented by sets of delta pulses (left column). Theoretical patterns are represented in the center column and experimental results at the right. Note that the maximum correlation peak (5) is obtained in example b) and c). Note that all the calculated results must be divided by $N^2=9$.

For recognition of both "0" and "1" binary states, the arrangement of Fig.3.5 can be used. In this dual-rail system, the positive and negative input mask are located, one beside the other in the input plane. Since the prism deviates the collimated light of one of the two parts, the Fourier spectra of the two masks (additionally diffracted by the grating) are spatially separated and can thus be filtered separately with two complementary filters. In this manner we have a system that can recognize, at the same time, the dark pixels of a search symbol in the positive mask, and the dark pixels of the binary complement of the search symbol in the negative mask. The distance b between the prism and the input mask is chosen such that the two correlation patterns recombine in output plane. A dark pixel appears at the place where the symbol is recognized. The superposition of the two correlation patterns is mathematically described by

$$I_o(x_o,y_o) = \frac{1}{N^2}\sum_{n=-K}^{K}\sum_{m=-K}^{K}\left\{|A_{mn}|^2\, I_i(x_o+ m\Delta x,\, y_o+ n\Delta y) + |\bar{A}_{mn}|^2\, \bar{I}_i(x_o+m\Delta x, y_o+ n\Delta y)\right\}$$

$$= I_i(x_o,y_o) \bigstar g_d(x_o,y_o) + \bar{I}_i(x_o,y_o) \bigstar \bar{g}_d(x_o,y_o), \qquad (3.31)$$

where $\bar{A}_{mn}$ is the binary complement of A_{mn}, and $\bar{I}_i(x,y)$ is the binary complement of $I_i(x,y)$. An experimental result obtained with the dual rail system will be presented in chapter 5.

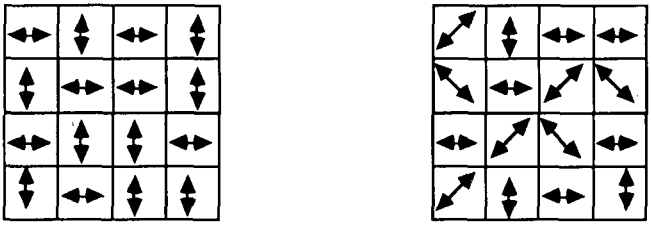


3.5. Optical arrangement for dual rail correlator

3.3.2. Polarization encoding

Logical state of the input data pixels can also be encoded in polarization state instead of intensity coding, e.g. two linear orthogonal states for two binary levels "0" and "1". Fig.3.6a) shows an example in which the pixels of the input pattern are binary polarization encoded.

We saw in 3.3.1 that a dual rail correlator can be used to recognize a symbol consisting of "0" and "1" pixels in a binary intensity encoded input mask. We will see now, that using our correlator, we can recognize a symbol composed of pixels having different polarization state in a binary polarization encoded input pattern, in which a logical "0" is coded by horizontally polarized light and a logical "1" is coded by vertically polarized light.



a) b)

Fig.3.6. Polarization encoded input mask. a) binary encoded, b) multilevel encoded.

The implementation of polarization coding in the above 4f correlator is straightforward. The amplitude of the input pattern is described by the vector $\mathbf{u}_i(x_i,y_i)$, which can take the two orthogonal polarizations :

$$\mathbf{u}_i \equiv \begin{pmatrix} u_{ix} \\ u_{iy} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{ or } \mathbf{u}_i = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.32}$$

In the optical processor, instead of introducing a binary transmission function A_{mn} in the Fourier plane (blocking the appropriate diffraction orders), as it is the case for intensity coded patterns, the polarization states are either inverted or not. This is done by introducing half wave plates in the appropriate diffraction orders in the Fourier plane. A polarizer is placed in front of the output plane, oriented along either of the two polarization axes. The optical amplitude at the output of the correlator is then described by

$$\mathbf{u}_1(x_o,y_o) = \mathbb{P} \sum_{n=-K}^{K} \sum_{m=-K}^{K} A_{mn}\, \mathbf{u}_i(x_o + m\Delta,\, y_o + n\Delta)\, \exp\left(i\phi_{mn}(x_o,y_o)\right). \tag{3.33}$$

ϕ_{mn} describes the phase factors of Eq.(3.21), $\mathbb{P}$ the Jones matrix of a linear polarizer along the x- or y-axis, respectively

$$\mathbb{P}_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ or } \mathbb{P}_y = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} . \tag{3.34}$$

The matrix $\mathbb{A}_{mn}$ can be described using the binary coefficients A_{mn} as

$$\mathbb{A}_{mn} = A_{mn}\, \mathbb{A}_{mn,0} + \bar{A}_{mn}\, \mathbb{A}_{mn,45} , \tag{3.35}$$

where $\bar{A}_{mn}$ is the binary complement of A_{mn}. That is to say, $\mathbb{A}$ operates either as a half wave plate at 45^0 or at 0^0. The latter has no effect on the input polarization state and is equivalent to an identity operation. Representation by Jones matrices yields

$$\mathbb{A}_{mn,45} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ or } \mathbb{A}_{mn,0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{3.36}$$

After completion of the matrix multiplications in Eq.(3.33), for the case of a $\mathbb{P}_y$ polarizer, the output amplitude becomes

$$u_o(x_o,y_o) = \frac{1}{N^2} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \Big\{ \big[A_{mn}\; u_{ix}(x_o + m\Delta,\, y_o + n\Delta) + \bar{A}_{mn}\, u_{iy}(x_o + m\Delta,\, y_o + n\Delta) \big] \times \exp\big(i\phi_{mn}(x_o,y_o)\big) \Big\} . \tag{3.37}$$

$|u_{ix}|^2 \equiv I_{ix}$ and $|u_{iy}|^2 \equiv I_{iy}$. From Eq.(3.32) follows that I_{iy} is the binary complement of I_{ix} ($I_{iy}(x,y) = \bar{I}_{ix}(x,y)$), thus after omission of the interference term, one gets finally for the correlation intensity

$$I_o(x_o,y_o) = \frac{1}{N^2} \sum_{n=-K}^{K} \sum_{m=-K}^{K} \Big\{ |A^{*}_{mn}|^2\, I_{ix}(x_o + m\Delta,\, y_o + n\Delta) + |\bar{A}_{mn}|^2\, I_{iy}(x_o + m\Delta,\, y_o + n\Delta) \Big\}$$

$$= I_{ix}(x_o,y_o) \bigstar g_d(x_o,y_o) + \bar{I}_{ix}(x_o,y_o) \bigstar \bar{g}_d(x_o,y_o). \tag{3.38}$$

The correlator yields thus a dark spot, where a y-polarized pattern correlates with A_{mn} and, at the same place, an x-polarized pattern correlates with $\bar{A}_{mn}$. Both binary states are thus recognized, as in a dual-rail processor. Note that a $\mathbb{P}_x$ polarizer at the system output yields the complementary result of Eq.(3.38). Using a polarizing beam splitter, both complementary results can be obtained for a dual-track system.

The experimental results of Fig.3.7 illustrate the recognition of a symbol composed by three pixels in a binary polarization encoded input (two polarization states). The polarization coded input data plane is given by a 32x84 element graphic liquid crystal device LCD [21]. (We will say more about this LCD in chapter 4). An area of 30x30 pixels has been illuminated with spatially coherent polarized light (expansion of a He-Ne laser beam). Fig.3.7a) shows the image of the input pattern passed across an analyzer which transforms the polarization pattern into an intensity pattern. Dark and bright pixels correspond to the two polarization states.

Fig.3.7b) shows the superposition of three images of the input pattern of Fig.3.7a). The first image (contained in the zero order) is not shifted, the second image is shifted by one pixel to the left, and the third is shifted by one pixel to the right. A dark pixel appears where the symbol composed by three dark pixels is recognized. This case is not new, in fact it corresponds to a single rail intensity encoded recognition and could be performed without polarization encoding. By changing the polarization state of one order in Fourier plane we invert the contrast of one of the three images (that shifted by one pixel to the left), and we obtain the result shown in Fig.3.7c). A dark pixel appears where a symbol composed by two dark pixels and a bright pixel is recognized. Note that to change the polarization in the Fourier plane we used an LCD (for detail concerning this device see chapter 4), in this manner the recognition symbol was programmable. Other examples of recognition are shown in Fig.3.7d)..i). Since the symbols to recognize are composed by three pixels and we use two states of polarization, the possible combinations are $2^3=8$ and are all represented in Fig.3.7. Note that a dark pixel indicates recognition of a symbol whereas a bright pixel indicates the recognition of the negative (inverted) symbol. Note that we have in Fig.3.7a)..i) four levels of the pixel intensity. Dark where three dark pixels are superimposed, bright of "level 1" where two dark pixels and one bright pixel are superposed, bright of "level 2" where two bright pixels and one dark pixel are superposed, bright of "level 3" where three bright pixels are superimposed.

Polarization coded optical processors are not restricted to binary data, it is possible to process more than two logic states. In our arrangement we can introduce a mask that is multilevel encoded (Fig.3.6b), and recognize a multilevel encoded search symbol. However, it is difficult to create masks or to find spatial light modulators which provide more than two well defined polarization states.

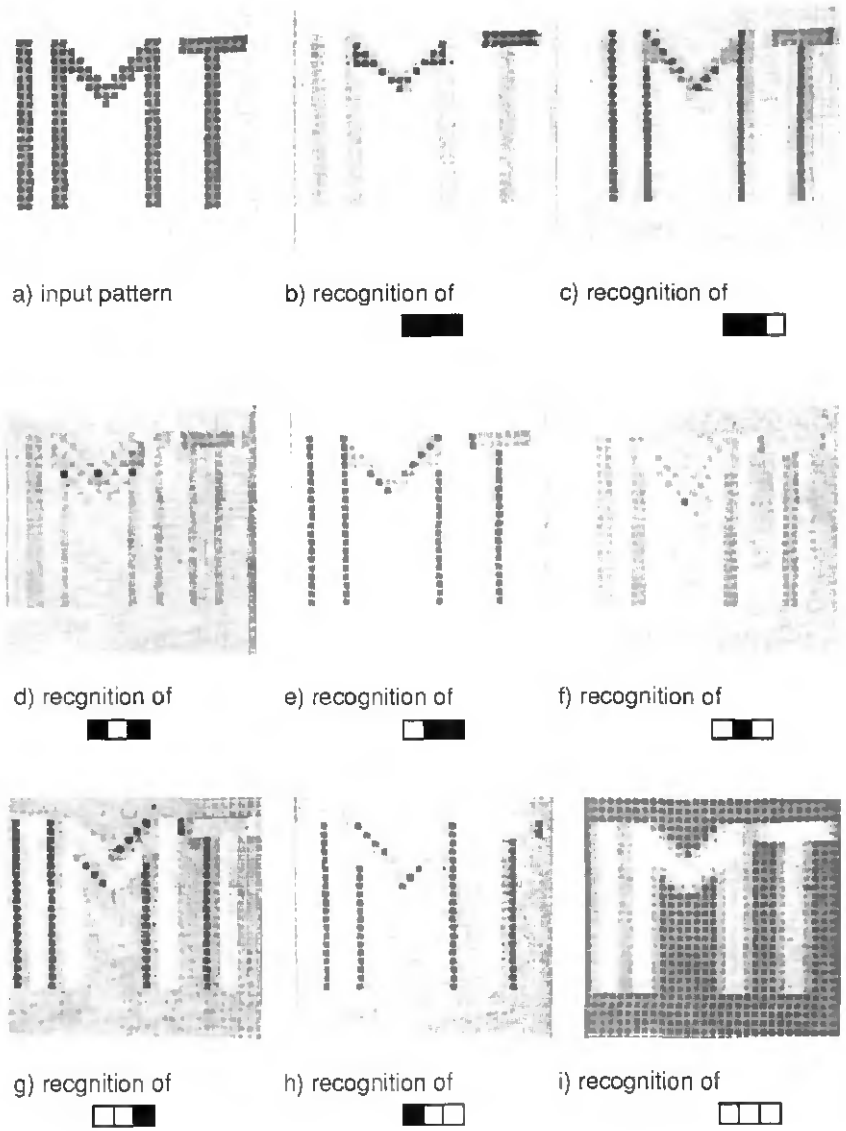


Fig.3.7 Experimental results of dual-rail correlation using liquid crystal displays as polarizing spatial light modulators. a) input pattern, b)..i) correlation patterns that allow recognition of different symbols.

3.3.3. Spatial frequency encoding

The basic idea of the so called theta modulation [15] is to encode optical data cells by the orientation of a grating structure (spatial frequency). Figure 3.8 shows an example.

a) b)

Fig.3.8. Theta encoded input mask. a) binary encoded, b) multilevel encoding.

The discrete correlator described before can also be used for theta modulated input data masks. For the ease of explication we suppose binary encoded input data with data cells forming either a horizontal or vertical grating which diffracts the illuminating plane wave into one diffraction order only, along the y and x direction, respectively. This represents the case of blazed gratings, or of cells are formed by small prisms. We describe the binary theta encoded input mask as a sum of two functions

$$u_i(x_i,y_i) = u_{ix}(x_i,y_i)t_x(x_i,y_i) + u_{iy}(x_i,y_i)t_y(x_i,y_i) \tag{3.39}$$

in which $u_{iy}(x_i,y_i)$ represents a binary function having value 1 in the horizontally encoded cells and 0 elsewhere, $u_{ix}(x_i,y_i)$ is the binary complement of $u_{iy}(x_i,y_i)$ and has a value of 1 for the vertically encoded cells. $t_x(x_i,y_i)$ is the transmission of a vertical grating which deflects the incident beam in x direction by an angle $\alpha_x=\lambda p_{1x}$ (p_{1x} is the spatial frequency of $t_x(x_i,y_i)$). $t_y(x_i,y_i)$ is the transmission of a horizontal grating which deflects the incident beam in y direction by an angle $\alpha_y=\lambda p_{1y}$ (p_{1y} is the spatial frequency of $t_y(x_i,y_i)$). The spectrum of $u_i(x_i,y_i)$ is the convolution of the Fourier transform of $u_{ix}(x_i,y_i)$ with the spectrum of the grating t_x plus the convolution of the Fourier transform of $u_{iy}(x_i,y_i)$ with the spectrum of the grating t_y. The period of the grating is smaller than the dimension Δ of the cell, the two spectra are also spatially separated in Fourier plane. This is illustrated in Fig.3.9a). When the theta encoded transparency is introduced in the 4f correlator arrangement, the gratings placed at distance *a* behind the mask produce many diffraction orders, and the spectrum in Fourier plane (x_f,y_f) can be described by a convolution of the spectrum of $u_i(x_i,y_i)$ with the spectrum of the diffraction grating, multiplied by a phase factor depending of the (m,n) diffraction order. Figure 3.9b) represents the convolution of the spectrum of a binary theta encoded transparency with the spectrum of a grating having 9 orders (3x3).

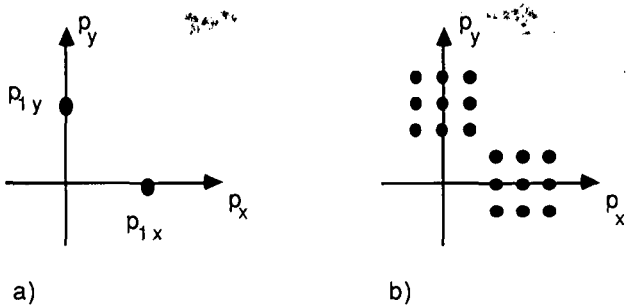


Fig.3.9. a) Spatial frequency spectrum of a binary theta modulated input mask. b) Spatial frequency spectra of a binary theta encoded input mask introduced in the correlator of Fig.2.2.

Mathematically we describe the complex amplitude in the Fourier plane by substituting $U_i(p_x,p_y)$ in Eq.(3.19) with

$$FT\{u_i(x_i,y_i)\} = U_{ix}(p_x-p_{1x},p_y) + U_{iy}(p_x,p_y-p_{1y}). \tag{3.40}$$

When the spectra are spatially separated, the different orders can be filtered. We describe the filter by a binary transparency, in which we represent with A_{mn} filtering of the spectrum of u_{ix}, and with $\bar{A}_{mn}$ filtering of the spectrum of u_{iy}, where $\bar{A}_{mn}$ are the binary complement of A_{mn}. $|u_{ix}|^2 \equiv I_{ix}$ and $|u_{iy}|^2 \equiv I_{iy}$. Since u_{iy} is the binary complement of u_{ix}, it follows that I_{iy} is the binary complement of I_{ix} ($I_{iy}(x,y) = \bar{I}_{ix}(x,y)$), thus after omission of the interference term, one gets finally for the correlation intensity

$$I_o(x_o,y_o) = \frac{1}{N^2}\sum_{n=-K}^{K}\sum_{m=-K}^{K}\left\{|A_{mn}|^2\,|u_{ix}|^2(x_o+m\Delta x,y_o+n\Delta y) + |\bar{A}_{mn}|^2|u_{iy}|^2(x_o+m\Delta x,\, y_o+n\Delta y)\right\}$$

$$= I_{ix}(x_o,y_o) \bigstar g_d(x_o,y_o) + \bar{I}_{ix}(x_o,y_o) \bigstar \bar{g}_d(x_o,y_o). \tag{3.41}$$

This equation is equivalent to Eqs.(3.31) and (3.38). This is analogous to a two channel correlator, in which in one channel we recognize a given spatial distribution of horizontally encoded pixels, and in the other channel we recognize the spatial distribution of vertical encoded pixels.

Figure 3.10 shows an example of filtering in the Fourier plane that allows the recognition of a given search symbol. Note that this filtering takes the double of the surface compared with that used for the recognition of a polarization encoded symbol having the same number of

pixels and the same surface as the dual rail correlator that recognizes black and white pixels (this problem will be discussed in more detail in section 3.8).

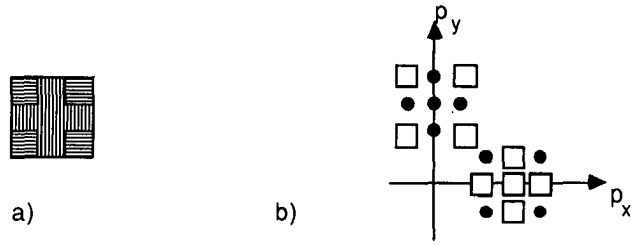


Fig.3.10. Example of filtering for the recognition of a theta encoded symbol. a) search symbol. b) filtering for the recognition of the search symbol .

The practical implementation of an ideal theta encoded mask in which the gratings encoding the cells are blazed, is difficult. For this reason in most cases we will encode the cells of the input pattern with gratings having more than one diffraction order. This will increase the number of orders in the Fourier plane and consequently decrease the space bandwidth product (SBWP) of the system (cf. section 3.8).

Theta encoded optical processors are, like polarization encoded processors, not restricted to binary data. It is possible to process more than two logic states. The following example illustrates an experimental result of the recognition of a symbol for theta encoded input data. The image of the theta encoded mask, taken in the output plane of the 4f system, is shown in Fig.3.11a). It contains 8x8 pixels and its original dimension is 11x11 mm^2. Four orientations of the gratings are used to encode the pixels (0^0,45^0,90^0,135^0). Its Fourier spectrum is represented in Fig.3.11b). It was obtained by illuminating the input with a collimated laser beam. This spectrum clearly shows the different diffraction orders produced by the gratings having different orientation. By filtering the Fourier spectrum as shown in Fig.3.11c), (blocking of the spatial frequencies along a line passing through the center of the spectrum) the pixels encoded by vertical gratings become dark in the output plane, whereas all the others remain bright. Analogously, it is possible to turn the filter in the other directions and to suppress the pixels encoded with gratings having other orientations. When two crossed gratings, having each three equal diffraction orders, are inserted behind the input mask, the distribution of the intensity in Fourier plane is also given by nine shifted Fourier power spectra, as shown in Fig.3.11.d). Now the nine images (each of them is contained in a diffraction order), can be independently filtered in the Fourier plane. Due to the four different spatial frequencies used for encoding, it is possible to recognize a symbol which in our case is composed by 3x3 pixels, where each pixel can have four levels. An example of recognition is

shown in Fig.3.11g). The dark pixel appears where the search symbol (Fig.3.11f), defined by the spatial filter sketched in Fig.3.11e), is recognized. In our case (with the symbol composed of 3x3 pixels and four levels), the possible filtering combinations are $4^9 = 262144$. Using theta encoding it is thus possible to implement multiple levels, which increases the complexity of the symbols that can be recognized.

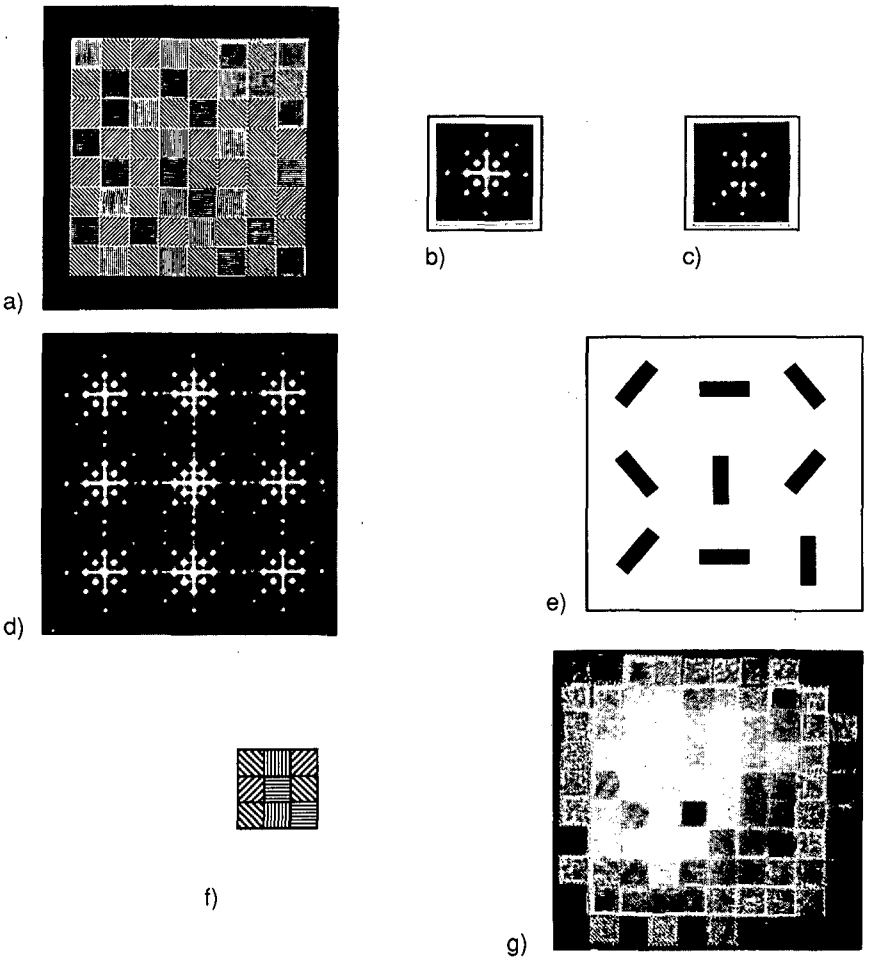
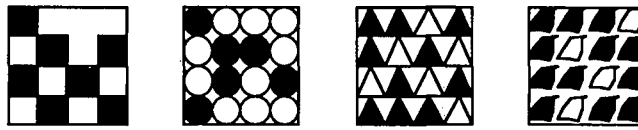


Fig.3.11. a) Theta encoded input mask, b) spectrum of the theta encoded mask, c) filtered spectrum of the theta encoded mask, d) spectrum after introduction of two crossed phase grating behind the input mask, e) schematic representation of a filter for the given search symbol f), g) recognition of the search symbol.

3.4. Geometrical arrangement of the pixels in the input mask

We present now some considerations about the geometrical arrangement and shape of the pixels in the input mask. These considerations can be applied to any encoding described in section 3.3.

The first consideration concerns the form of the pixels, that we have denoted with the pixel function $f_0(x,y)$ in Eq.(3.8). It is obvious that $f_0(x,y)$ can describe any shape (triangular, hexagonal, round,..) and not only the square shape used for the examples in section 3.3. Notice that using a cartesian grid and pixels having not square shape, it is not possible to cover the complete surface with pixels (Fig.3.12).



a) b) c) d)

Fig.3.12. Some example of binary intensity encoded input masks, in which the pixels have different shapes.

In 3.3. we discussed only examples in which the pixels of the input mask were arranged on a regular cartesian grid, and the discrete function to correlate with the input had the same structure. Consider now a diffraction grating having the Fourier transform

$$C(p_x,p_y) = \frac{1}{N}\sum_{n=-K}^{K}\sum_{m=-K}^{K}\delta(p_x-p_{xmn},p_y-p_{ymn}), \tag{3.42}$$

where (p_{xmn},p_{ymn}) are the coordinates of the (mn) delta pulse. The total number of delta pulses is N^2, and they are located on a periodic 2-dimensional structure that can be for example: cartesian, hexagonal, triangular... The grating is inserted in the optical arrangement of Fig.2.2, at a distance a behind the input pattern. A filter having transparency A_{mn} in (p_{xmn},p_{ymn}) performs a filtering of the different diffraction orders in the Fourier plane. The intensity in the output plane is obtained by generalizing the calculation reported in section 3.2 for the cartesian case. Thus by substituting $(m\lambda/d,n\lambda/d)$ with the more general expression (p_{xmn},p_{ymn}) we get

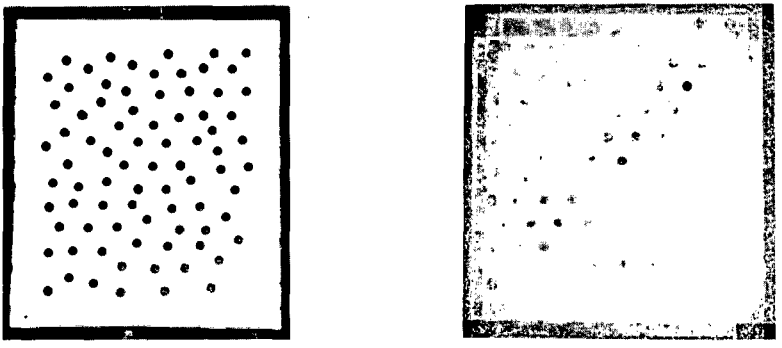
$$I_o(x_o,y_o) = \frac{1}{N^2}\sum_{n=-K}^{K}\sum_{m=-K}^{K}\left\{|A_{mn}|^2\, I_i(x_o+a\lambda p_{xmn},\, y_o+a\lambda p_{xmn})\right\}. \tag{3.43}$$

The (mn) copy is shifted by $(\xi_{mn}, \eta_{mn}) = (a\lambda p_{xmn}, a\lambda p_{xmn})$. Eq.(3.42) can be written as a correlation between the input pattern I_i and the function

$$g_d(x_o,y_o) = \frac{1}{N^2} \sum_{n=-K}^{K} \sum_{m=-K}^{K} |A_{mn}|^2 \delta(x_o - a\lambda p_{xmn}, y_o - a\lambda p_{ymn}) . \qquad (3.44)$$

When the pixels of the input pattern match with the delta pulses of the function g_d, the system can be used to recognize a given distribution of pixels (symbol) in the input pattern.

The experimental result of a recognition is shown in Fig.3.13. Symbols composed of seven dark pixels, placed on the vertex and in the center of a hexagon, are hidden in a random distribution of pixels (noise), as shown in Fig.3.13a). The system performs the correlation between the input pattern and a function g_d composed of seven delta pulses located in the vertices and in the center of a hexagon (search symbol). The copies were produced by using two linear diffraction gratings rotated by 60^0 one with respect to the other. At the location of the recognized symbol a dark pixel appears (Fig.3.13b)). The randomly distributed pixels do not have a structure and for this reason, after superposition of seven shifted images, the noise pixels are erased by the bright background.



a) b)

Fig.3.13. Recognition of hexagons hidden in a random distribution of pixels. a) input pattern, b) recognition.

We have shown (section 3.2) that two gratings c and c' can be inserted in our optical arrangement at a distance a and a' behind the input pattern. From Eq.(3.19) it follows that a combination of two diffraction gratings allows to increase dramatically the number of produced copies and the geometrical complexity of the structure.

In Fig.3.14a) and b) are represented the spectra of two gratings, having two and four diffraction orders, respectively. In Fig.3.14c) is represented the power spectrum that appears in the Fourier plane (convolution between the spectra of a) and b)) for this combination of gratings. By correct choice of a and a', such a system allows to recognize the spatial distribution of the pixels like that sketched in Fig.3.14d). By changing the location of the grating with two diffraction orders, it allows for example to recognize the symbol in Fig.3.14e).

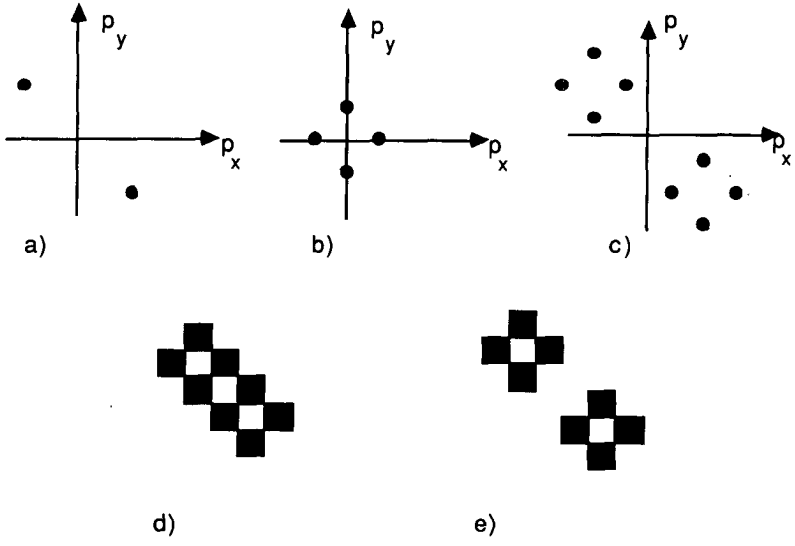


Fig.3.14. a) spectrum of a one dimensional grating having two diffraction orders. b) spectrum of a two dimensional grating having four diffraction orders. c) spectrum of the combination of the two gratings. d) and e) examples of symbols that can be recognized using a simple combination of gratings .

3.5. Discussion on the spatial coherence of the illumination

In the calculation of the system presented in section 3.2, we assumed that the input mask is illuminated by a monochromatic plane wave. This kind of coherent illumination can be realized in practice using an expanded laser beam. It was emphasized that using coherent illumination, interference fringes appear in the output plane of the correlator. This phenomenon is described by the interference term $I_t(x_o,y_o)$ in Eq.(3.22).

In this section we will study what happens when the illumination of the input mask is partially spatial coherent. In particular, we will study what happens when the degree of coherence of the illumination decreases, and what are the coherence requirements to allow the correlator to work. At first, we introduce some general concepts about partially coherent illumination. A typical arrangement that can be used to produce a partially spatial coherent illumination is sketched on Fig.3.15. The light coming from an extended quasi-monochromatic source is "collimated" by a lens L_S having focal length f_S. The radiation process for each point of the source is supposed to be independent of the other points. This kind of quasi-monochromatic extended sources can be easily approximated in practice by a mercury lamp, or an incandescent lamp with a filter that reduces the bandwidth of the spectrum.

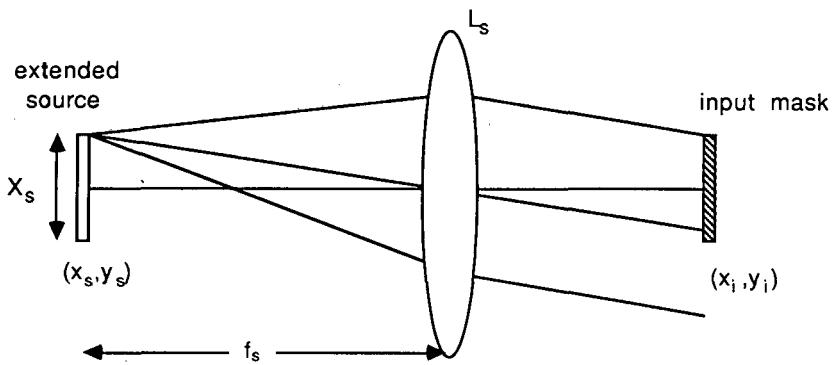


Fig.3.15. Partially coherent illumination of an input mask, using an extended monochromatic source.

In order to explain in more detail the nature of the partially coherent illumination produced by the arrangement depicted in Fig.3.15, we restrict ourselves to a one dimensional case, the generalization to the two dimensional case is straightforward. A spherical wave originating from point x_S of the source, is transformed by the lens L_S into a plane wave which arrives in the plane of the mask (x_i,y_i). This plane wave can be mathematically described by

$$s_S(x_i) = K\sqrt{I(x_S)}\exp\left(-2\pi i\,\frac{x_i x_S}{\lambda f_S} + i\phi(x_S)\right), \tag{3.45}$$

K is a constant proportional to $1/f_S$, $I(x_S)$ is the intensity at the point of the source x_S, and $\phi(x_S)$ is a phase depending on the corresponding point of the source. Since each point emits independently, there is no correlation between the phases. The illumination field in front of the input mask, is given by a superposition of plane waves,

$$s(x_i) = K\int_{-\infty}^{\infty}\sqrt{I(x_S)}\exp\left(-2\pi i\,\frac{x_i x_S}{\lambda f_S} + i\phi(x_S)\right)dx_S\,. \tag{3.46}$$

The plane wave of maximum inclination with respect to the optical axis, has the spatial frequency $X_S/(2\lambda f_S)$. Note that Eq.(3.46) represents a Fourier transform of $\sqrt{I(x_S)}\exp\{i\phi(x_S)\}$.

The nature of the illumination in the plane (x_i,y_i) can be conveniently described by the mutual intensity function (see e.g. Ref.4), that is defined as the correlation between the vibration at two points in the wave field. We choose two points X_i and X_i' in plane (x_i,y_i), thus the mutual intensity function can be written as

$$J(X_i,X_i') = \langle s(X_i)\, s^*(X_i')\rangle\,. \tag{3.47}$$

We consider the quasi-monochromatic condition thus the phase $\phi(x_S)$ is slowly time dependent. The $\langle\ \rangle$ brackets denote a time average process over a time-span greater than the coherence time of the light. The light vibrations arising from different points of the source may be assumed to be statistically independent (mutually incoherent) thus substituting Eq.(3.46) in Eq.(3.47) we get

$$J(X_i,X_i') = K^2\int_{-\infty}^{\infty} I(x_S)\exp\left(-\frac{2\pi i x_S}{\lambda f_S}(X_i'-X_i)\right)dx_S\,, \tag{3.48}$$

$J(X_i,X_i')$ is also proportional to the Fourier transform of the intensity function of the source. From Eq.(3.48) we can define a normalized mutual intensity function

$$\mu(X_i,X_i') = \frac{\displaystyle\int_{-\infty}^{\infty} I(x_S)\exp\left(-\frac{2\pi i x_S}{\lambda f_S}(X_i'-X_i)\right)dx_S}{\displaystyle\int_{-\infty}^{\infty} I(x_S)\,dx_S}\,. \tag{3.49}$$

$\mu(X_i,X_i')$ is usually called complex coherence factor. It can be demonstrated (see i.e. Ref.4) that the absolute value of the complex coherence factor is equal to 1 for a spatial coherent

illumination and 0 for a completely spatial incoherent illumination. For partially spatial coherent illumination we have the relation : $0< |\mu(X_i,X_i')| <1$. The physical meaning of this term can be seen be performing the Young experiment. We place in plane (x_i,y_i) a screen with two holes located at X_i and X_i', then the contrast of the fringes produced by the interference of the two spherical waves originating from the the two apertures is $|\mu(X_i,X_i')|$. In sections 3.5.2 we will discuss how the complex coherence factor can be used to describe the contrast of the interference fringes in the output plane of the discrete correlator.

3.5.1. Spatial coherence requirements

Spatial filtering in the Fourier plane of the correlator is only possible when the diffraction orders produced by the gratings do not overlap too much. It is clear that when we illuminate the input pattern with partially coherent light (e.g. using the arrangement of Fig.3.15), the spatial extension of each diffraction order in the Fourier plane is increased with respect to coherent illumination. The coherence degree of the illumination limits also the number of orders which do not overlap and thus the capability of the correlator. We explain now in more detail what happens in Fourier plane when the input mask is illuminated with partially coherent light coming from an arrangement like that described in Fig.3.15.

The complex amplitude of the field in front of the input mask is given by Eq.(3.46), it is also a superposition of plane waves which can be represented as a Fourier transform. We denote with $t(x_i)$ the transmission of the input mask, thus just behind the mask we will have the complex amplitude $u_i(x_i)= t(x_i)s(x_i)$. We consider at first only one diffraction order (for example the 0 order) and we determine its spatial extension in the filtering plane. The complex amplitude in the Fourier plane is given by a convolution between the Fourier transform of the input mask and the Fourier transform of the amplitude field in front of the input mask,

$$FT\{t(x_i)s(x_i)\} = T(x_f) * S(x_f). \tag{3.50}$$

$s(x_i)$ is given by Eq.(3.46) thus

$$S(x_f,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{I_S(x_S)}\, \exp\{i\phi(x_S)\}\exp\{-2\pi i x_i\left(\frac{x_S}{\lambda f_S} + \frac{x_f}{\lambda f}\right)\}dx_S dx_i\,, \tag{3.51}$$

and after some calculations we get

$$S(x_f,t) = \sqrt{I_S(-f_S x_f/f)}\, \exp\{i\phi(-f_S x_f/f)\}\,, \tag{3.52}$$

where the function I_S represents the intensity of the source. Thus the intensity $|S(x_f)|^2 = I_S(-f_S x_f/f)$ represents an inverted magnified image of the source.

We consider the case in which the input pattern is described by the one dimensional function composed by (1–dimensional) pixels spaced by Δ. Thus the transmission can be written as

$$t(x_i) = \sum_{k=-K}^{K} e_k \, \mathrm{rect}\left(\frac{x_i - k\Delta}{\Delta}\right), \tag{3.53}$$

where e_k denotes the encoding of the k-th pixel, if we use binary intensity encoding it will be a scalar having value 1 for transparent pixels and 0 for opaque pixels. The Fourier transform of the input mask is given by a superposition of sinc functions having different phases, viz.

$$T(x_f) = \Delta \mathrm{sinc}\left(\frac{\Delta x_f}{f\lambda}\right) \sum_{k=-K}^{K} e_k \exp\left(2\pi i k \frac{\Delta x_f}{f\lambda}\right). \tag{3.54}$$

Consider the case in which the input mask contains only one transparent pixel, then $T(x_f)$ is a single sinc function, and the discussion is easier. (The spatial extension of the Fourier transform will decrease if we consider a mask with more than one pixel.) According to Eq.(3.50), the amplitude in Fourier plane is thus given by $\mathrm{sinc}(\Delta x_f/(f\lambda)) * S(x_f)$. When we use a point source (coherent illumination), we have a convolution between a sinc function and a delta function and this gives as a result a sinc function. The intensity is thus a square sinc function.

Consider now an extended source having intensity $I_S(x_S)$. The radiation process of each point of the source is independent of each other, for this reason the intensity $|T(x_f) * S(x_f)|^2$ is given by an incoherent superposition of square sinc functions, viz.

$$|T(x_f) * S(x_f)|^2 = |T(x_f)|^2 * |S(x_f)|^2 = [\mathrm{sinc}(\Delta x_f/f\lambda)]^2 * I_S(-f_S x_f/f). \tag{3.55}$$

When the diffraction grating is introduced in the optical arrangement, the power spectrum can be described by a convolution between $|T(x_f) * S(x_f)|^2$ and a set of delta pulses spaced by $\lambda f/d$. Figure 3.16 shows the intensity of two diffraction orders in the Fourier plane for two different extensions of the source. In the first case (Fig.3.16a), the source is a point, the illumination is thus spatially coherent and the intensity of each order is a square sinc function. In the second case (Fig.3.16b)), the source has dimension X_S, and the intensity of each order is given by Eq.(3.55), it appears clearly that using an extended source each diffraction order has a larger extension compared with the case of a point source. The separation of two diffraction orders in the filter plane is given by $\lambda f/d$, thus we have to choose the period of the grating and the dimension of the pixels so that two nearest diffraction orders do not overlap. In fact when, for example, we like to suppress a diffraction order which overlaps partially with another, we

suppress at the same time a part of the other. In section 3.7 we will see that when the overlap is small, it will not have a dramatical effect on the result in the output plane of the correlator.

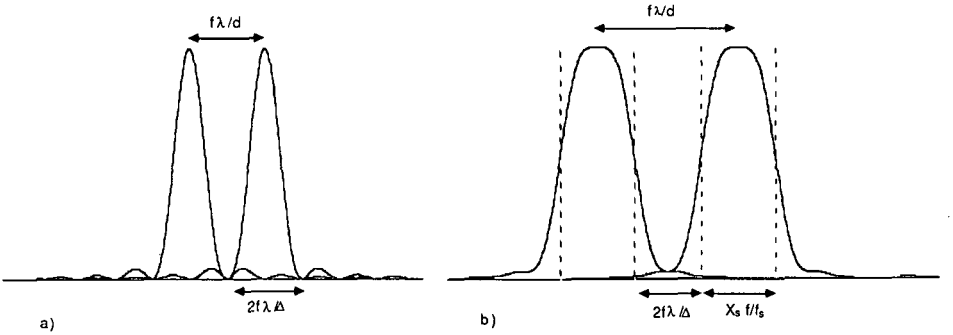


Fig.3.16. Intensity distribution of two diffraction orders in the Fourier plane for the case of spatial coherent illumination a), and partially spatial coherent illumination b).

The separation between two diffraction orders in the Fourier plane must satisfy the relation

$$\frac{\lambda f}{d} > \frac{2\lambda f}{\Delta} + \frac{fX_s}{f_s} . \qquad (3.56)$$

$2\lambda f/\Delta$ represent the width of the function $\mathrm{sinc}(\Delta x_f/f\lambda)$, or more precisely the distance between the two first zeros. fX_s/f_s is the extension of the image of the source. If $fX_s/f_s >> \lambda f/\Delta$ we can neglect the effect given by the diffraction of the pixels.

We consider now the case in which the input mask is itself a spatial incoherent source. It can be for example a CRT, an array of light emitting diodes (LED), etc. In this case the source cannot be introduced directly into the correlator, because it is completely incoherent, and for this reason the filtering in the Fourier plane is not possible. This kind of incoherent sources can be used when their emitting angle is limited. This can be done by spatial filtering as sketched in Fig.3.17. This process of spatial filtering is performed at the expense of optical power. Notice that the dimension W of the low pass filter must be at least $2\lambda f/\Delta x$, in order to allow good reconstruction of the pixels of the input mask in plane (x_i,y_i). After this filtering operation the image can be used as input for the discrete correlator. The width of each diffraction order in the Fourier plane (x_f,y_f), not shown in Fig.3.17, is equal to W. The separation between two diffraction orders in the Fourier plane must thus satisfy the relation

$$W < \frac{\lambda f}{d} . \qquad (3.57)$$

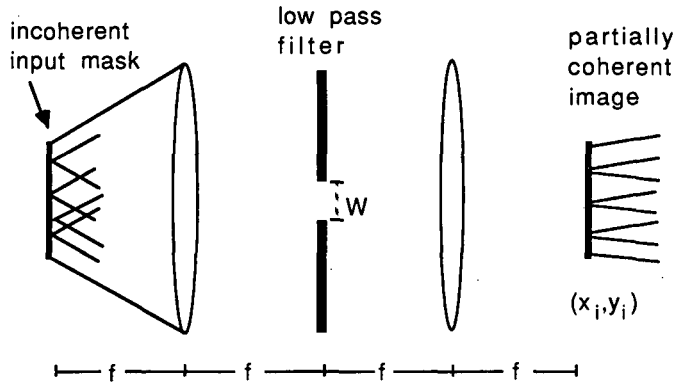


Fig.3.17. Spatial filtering arrangement to increase the spatial coherence of an input source array.

3.5.2. Visibility of the interference fringes in output plane

In section 3.5.1 we discussed the spatial coherence requirements which the illumination must satisfy to allow filtering in the Fourier plane. In this section we discuss the influence of the spatial coherence of the illumination on the output image, i.e. in plane (x_o,y_o). During the calculation reported in section 3.2 for the coherent case we noted that in the output plane of the correlator an interference term appears, this is produced by the coherent superposition of the different diffraction orders having different propagation directions. This term appears in Eq.(3.22) and was denoted by $I_I(x_o,y_o)$. We said that it prevents us from having a pure correlation and now we will see that using a partially spatial coherent illumination it is possible to suppress it.

A. Theoretical calculation of the visibility

Equation (3.21) describes the complex amplitude in the output plane of the correlator when the input mask is illuminated with a plane wave parallel to the z-axis. Consider now the case in which the input mask is illuminated with the arrangement depicted in Fig.3.15. A spherical wave originating from the point x_s of the source and transformed in a plane wave by the lens L_s, produces in the output plane of the correlator a complex amplitude

$$u_{op}(x_o) = \frac{\sqrt{I(x_s)}\exp\{\phi(x_s)\}}{N} \sum_{m=-K}^{K} \{ A_m\, u_i(x_o+\lambda am/d)\ \exp\{2\pi i(x_o m/d)\}\ \exp\{\pi i a\lambda (m/d)^2\}$$

$$\times \exp\left(2\pi i \left(x_o+\lambda a \frac{m}{d}\right)\frac{x_s}{\lambda f_s}\right)\}. \qquad (3.58)$$

A new phase term depending on $x_s/\lambda f_s$ appears in Eq.(3.58) (last term). This means that each point of the source produces a different (shifted) interference fringe pattern in the output plane of the correlator. In order to simplify the treatment, we suppose that in the output plane only two images are superimposed, i.e. the grating is supposed to have only two orders, p and q (p and q are integers and we suppose p>q). We will show later that the result obtained for the superposition of two orders can be used to explain the superposition of more orders. The fringes produced in output plane when only two orders are present can directly be calculated from Eq.(3.58) by putting $A_p = A_q = 1$ and all the other $A_m = 0$. The intensity in the output plane, after omission of some constant factors, is thus given by

$$I(x_o) = I(x_s)\,|u_i(x_o+\lambda ap/d)|^2 + I(x_s)|u_i(x_o+\lambda aq/d)|^2 +$$

$$+ 2I(x_s)|u_i(x_o+\lambda ap/d)||u_i(x_o+\lambda aq/d)|\cos\left[2\pi\frac{p-q}{d}\left(x_o+a\frac{x_s}{f_s}\right)\right]. \qquad (3.59)$$

The first term of this equation describes the intensity of the input image shifted by $\lambda ap/d$, the second term describes the intensity of the input image shifted by $\lambda aq/d$, the third term describes the interference between the two diffraction orders which we denoted with $I_t(x_o,y_o)$ in Eq.(3.22).

The wave originating from point x_s, causes a shift of the interference fringes by ax_s/f_s. Each point of the source produces a related shift of the fringe system, which has to be incoherently added in the output plane. In order to estimate the contrast of this superposition, we integrate the interference term of Eq.(3.59) over all the points of the source. We consider a rectangular source of uniform brightness, mathematically described by $\mathrm{rect}(x_s/X_s)$, thus after omission of some constant factors we get

$$I_t(x_o) = \frac{2|u_i(x_o+\lambda ap/d)||u_i(x_o+\lambda aq/d)|\lambda f_s}{X_s} \int_{-\infty}^{\infty} \mathrm{rect}(x_s/X_s)\cos[2\pi\frac{p-q}{d}(x_o+a\frac{x_s}{f_s})]\,dx_s\,. \qquad (3.60)$$

After integration and some straightforward trigonometrical transformations we obtain

$$I_t(x_o) = 2|u_i(x_o+\lambda ap/d)||u_i(x_o+\lambda aq/d)|\ \mathrm{sinc}\left(\frac{a(p-q)X_s}{df_s}\right)\cos(2\pi x_o(p-q)/d)\,. \qquad (3.61)$$

The absolute value of sinc[a(p–q)X_s/(df_s)] determines the contrast of the fringes in the output plane (0<|sinc[aqX_s/(df_s)]|<1). It is quite easy to find a relation between this contrast and |μ(X_i,X_i')|. We calculate Eq.(3.49) for the source rect(x_s/X_s); this gives the result

$$\mu(X_i,X_i') = \mathrm{sinc}\left(\frac{(X_i-X_i')X_s}{\lambda f_s}\right). \tag{3.62}$$

For (X_i–X_i') = a(p–q)λ/d we obtain the equivalence between μ(X_i,X_i') and the "sinc" term in Eq.(3.61). The contrast of the fringes produced by the superposition of two diffraction orders shifted by a(p–q)λ/d in the output plane of the correlator is equal to the contrast of the fringes obtained in a Young experiment in which a screen with two apertures spaced by a(p–q)λ/d is located in the illumination field of the input mask.

sinc[a(p–q)X_s/(df_s)] has value 1 when X_s tends to 0 (i.e. the illumination tends to be spatial coherent). The interference fringes disappear in the output plane of the correlator (zero contrast), when the extension of the source is

$$X_s = \frac{df_s}{(p-q)a} \tag{3.63}$$

Equation (3.63) shows that the width of the source needed to suppress the interference fringes between two adjacent diffraction orders [i.e.(p–q)=1], is larger than the width for the suppression of the fringes between any two other orders. We can also say, that when we have a partially spatial coherent illumination suppressing the interference fringes between two adjacent orders, all the interferences between the N diffraction orders represented in Eq.(3.58) vanish.

Equation (3.63) is not in contradiction with the filtering condition [Eq.(3.56)], this means that it is possible to choose the spatial coherence of the illumination so that no interference fringes appear in the output plane.

B. Experimental verification

We present the results of an experiment performed using partially spatial coherent light. Its purpose is to check that the contrast of the interference fringes, produced by superposition of two diffraction orders in the output plane of the correlator, is indeed given by the sinc–term of Eq.(3.61).

The arrangement used to obtain the result is sketched in Fig.3.18. The monochromatic (or more correctly quasi-monochromatic) light was produced by a mercury arc source (MS) provided with an interference filter which selects the peak wavelength of 546.1 nm. The

spatially incoherent light from the mercury lamp illuminates an aperture of width X_S, located in the front focal plane of a lens L_S, with focal length f_S, which collimates the light coming from the aperture. The degree of coherence of the illumination behind the lens L_S, can be increased or decreased by changing the dimension of this aperture, i.e. the degree of coherence increases for decreasing aperture width. The partially coherent beam illuminates a phase only grating (G), that in our experimental set-up has three diffraction orders with the same brightness. In the Fourier plane FP we can introduce a mask which allows us to choose the diffraction orders that we like to superimpose in the output plane. A magnified image of the output plane (OP) is observed by a CCD camera using the lens L_M.

If the grating is exactly located in the input plane of the 4f system, and if we allow to pass through the Fourier plane only two diffraction orders (the 0-th order and the 1-st order), we will find interference fringes in the output plane. This is independent of the degree of spatial coherence of the illumination. In fact, the two diffraction orders have the same optical path length and can also interfere. We calculated in section 3.5.1 (Eq.3.61), that for a=0, the contrast of the fringes is equal to 1. When the illumination is partially coherent and the grating is not located in the input plane of the 4f system ($a \neq 0$), the contrast of the fringes will decrease.

For the practical demonstration of this effect we chose the width of the aperture X_S=1mm, the focal length of the collimating lens f_S=380 mm, and a binary phase only grating having a period d=80 μm.

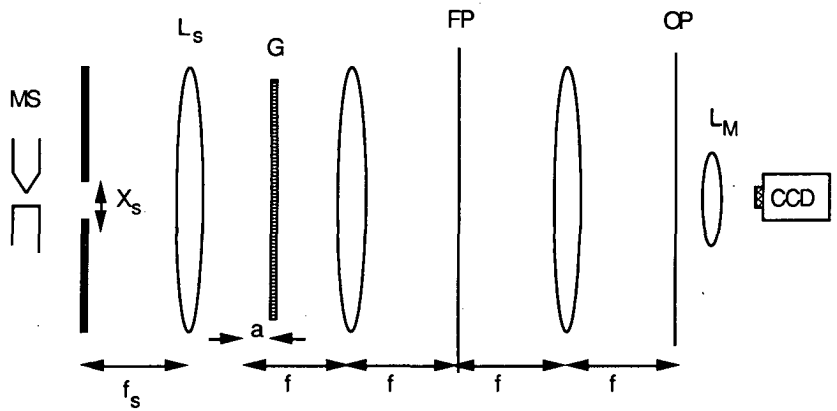
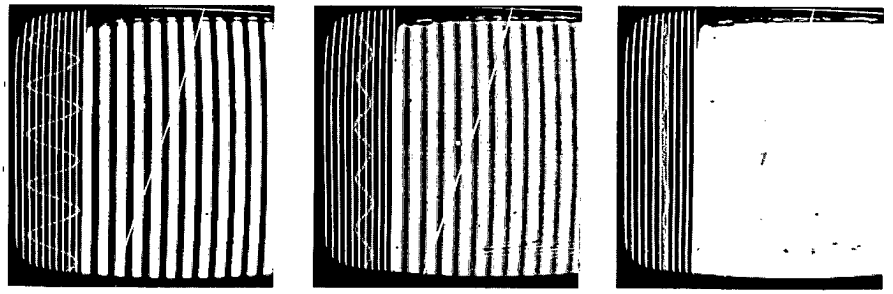


Fig. 3.18. Arrangement used to demonstrate that the visibility of the interference fringes in the output plane is dependent on the spatial coherence degree of the illumination.

Figure 3.19a) shows the magnified interference fringes (in reality the periodicity is equal to the period of the grating which is 80 μm), that can be observed when the grating is located in

the input plane. It should be noted that the contrast is very good, as shown from the display of the intensity profile along the indicated straight line.



a) b) c)

Fig.3.19. Visibility of the interference fringes for different locations of the grating when partially coherent illumination is used. a) a = 0 , b) a = 15 mm, c) a = 30 mm.

Figure 3.19b) shows that the contrast has decreased to about 1/2 when the grating is shifted by a=15 mm. For a=30 mm, (Fig.3.19c) the contrast vanishes. According to Eq.(3.63), we find that the interference fringes disappear for $a = df_S/X_S = (0.08 \times 380)/1 = 30.4$ mm. By choosing *a* larger than 30 mm, it was possible to observe that the contrast increases and decreases in accordance to the sinc function of Eq.(3.61). Negative values of the sinc function indicate a reversal of the contrast (bright fringes become dark and dark fringes become bright). This effect was observed during the experiment.

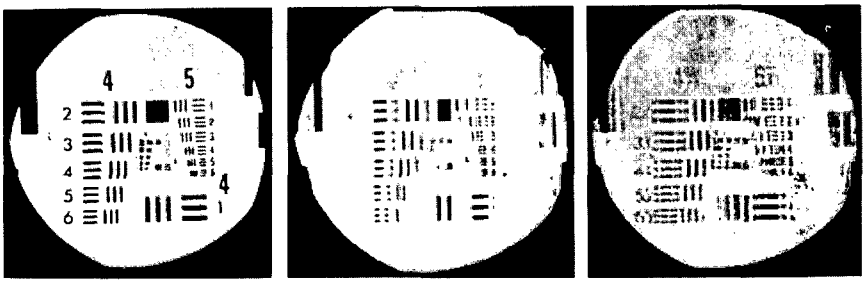
In section 3.5.1, we said that for more than two interfering diffraction orders the contrast of the fringes in the output plane produced by this multiple beam interference decreases faster when the grating is shifted out of the input plane than in the case of the interference of only two diffraction orders. Experimentally we demonstrated this for the case where three diffraction orders are superimposed. For the same period of grating and the same spatial coherence of the illumination as in the experiment described before, we observed that already for a shift a=15mm the interference fringes disappear.

C. Examples of correlation obtained with coherent and partially spatial coherent illumination

We present now an example in which the input mask is represented by an USAF test target. We consider at first the results obtained with partially coherent illumination. A mercury arc source and an arrangement like that described in Fig.3.18 were used to produce this illumination. The magnified image of the test target, observed at the image plane of the 4f lens system, is shown in Fig.3.20a). By inserting a diffraction grating behind the test target, it is

possible to produce many images of the test target that can be superimposed with a shift in the output plane. Figure 3.20b) shows the case of the superposition of only two images, shifted by about 50 μm one with respect to the other. The group element 4.3 of the test target corresponds in fact to 20.1 line pairs per millimeter, the centers of two dark lines are separated by about 50 μm. To obtain this result we used a grating having a period of about 20 μm and we choose a shift (50 μm) that is about 2.5 times this period.

Figure 3.20c) shows the superposition of three copies of the test target, we have a central copy (given by the 0 order) and the two other copies, shifted each by 50 μm (0.05 mm) with respect to the central copy. This can be seen as a correlation between the test target and three delta pulses located on an horizontal line and spaced by 50 μm.



a) b) c)

Fig.3.20. Examples of correlation by superposition of shifted copies of an USAF test target using partially spatial coherent light. a) input image, b) superposition of two copies shifted by 50 μm, c) superposition of three copies shifted by 50 μm.

To produce the images of Fig.3.20 we used an arrangement with X_S=5 mm and f_S=380 mm, the shift (Δ) was 50 μm, this means that $X_S\Delta/\lambda f_S = (5 \times 0.05)/(0.55\ 10^{-3} \times 380) = 1.2 > 1$. Thus we are beyond the first zero of the sinc function and we cannot observe interference fringes between the diffraction orders (the contrast of the fringes is low). By performing the experiment with a smaller aperture X_S=3 mm we observed interference fringes having a period of about 20 μm (period of the grating).

The images of Fig.3.20 are typically incoherent, in fact we do not observe the coherent noise which always affect the coherent images.

We describe now a result obtained by illuminating the test target with spatially coherent light produced by expansion of a beam coming from an He-Ne laser. Figure 3.21a) shows the

magnified image of the test target, obtained by inserting the target in the input plane of the 4f system (without the diffraction gratings). We see that this image is affected by coherent noise that is produced by blemishes, dust, scratches and reflections from the optical components. Even by carefully cleaning the optical surfaces we cannot obtain a much better image. The introduction of diffraction gratings in the system produces obviously an increase of the coherent noise. The degradation of the images deteriorates the quality of the superposition of the shifted copies of the test target. Good correlation could not be achieved because the images did not have a good quality. The result of superimposing two images, one shifted by about 50 μm with respect to the other, is shown on Fig.3.21b). The interference fringe pattern contributes dramatically to deteriorate the quality of the correlation. The periodicity of the interference fringes is equal to that of the grating used (about 20 μm).

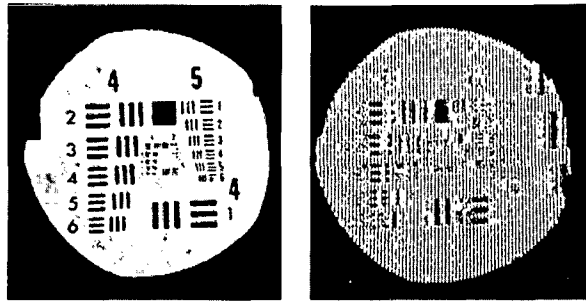


Fig.3.21. Example of correlation by superposition of shifted copies of an USAF test target, using spatial coherent light. a) input image, b) superposition of two copies shifted by 50 μm.

By comparing the results of Figs. 3.20 and 3.21 it is apparent that using partially coherent illumination we obtained better results than using coherent illumination.

3.5.3. Discussion of the spatial coherence using optical modes

Until now we calculated formally the contrast of the fringes in the output plane of our correlator when we use partially coherent illumination. We take now a more physical approach by using the concept of optical modes. The number of transverse optical modes emitted by a surface is in fact a measure of the spatial coherence of the source. The concept of optical mode [25] follows directly from the Heisenberg uncertainty relation, that says that it is not possible to know exactly the position and the momentum of a particle (in our case a photon). The uncertainty relation can be written as

$$\Delta p_x \Delta x > h\,, \qquad \Delta p_y \Delta y > h\,, \tag{3.64}$$

where Δp_x, Δp_x are the uncertainties of the momentum, Δx, Δy are the uncertainties of the position and h is the Planck constant. The angular uncertainty of a photon is given by

$$\Delta\vartheta_x = \frac{\Delta p_x}{p}\,, \qquad \Delta\vartheta_y = \frac{\Delta p_y}{p}\,, \tag{3.65}$$

where $p = h/\lambda$ is the total momentum of the photon. Thus using (3.64) we obtain

$$\Delta\vartheta_x = \frac{\lambda}{\Delta x_s}\,, \qquad \Delta\vartheta_y = \frac{\lambda}{\Delta y_s}\,. \tag{3.66}$$

If we denote the solid angle $\Delta\vartheta_x\Delta\vartheta_y$ with $\Delta\Omega$, and the surface $\Delta x_s\Delta y_s$ with ΔA, then we obtain the relation

$$\Delta\Omega\Delta A = \lambda^2 \tag{3.67}$$

for a transverse optical mode, representing the fact that the photons emitted by a surface ΔA into the solid angle $\Delta\Omega$ [satisfying Eq.(3.67)] are indiscernible. Thus, the number of transverse optical modes of a source having surface A and emitting into a solid angle Ω is given by

$$N_m = \frac{\Omega A}{\lambda^2}\,. \tag{3.68}$$

We can now calculate the number of transverse optical modes that illuminate the input mask of our correlator when we use the illumination arrangement of Fig 3.15. We consider a two dimensional source of square shape with a surface $A = X_S^2$. The solid angle Ω is given by $(D_S/f_S)^2$, where D_S is the diameter and f_S is the focal length of the lens. Thus

$$N_m = \frac{\Omega A}{\lambda^2} = \left(\frac{X_S D_S}{\lambda f_S}\right)^2\,. \tag{3.69}$$

We can also calculate the number of optical modes for which the contrast of the interference fringes in output plane vanishes. Equation (3.63) indicates the extension of the source for which the contrast vanishes. Considering the case in which (p–q=1) and using relation $\Delta=\lambda a/d$ [Eq.(3.25)], we obtain

$$N_{m0} = \left(\frac{D_s}{\Delta}\right)^2 , \tag{3.70}$$

where Δ represents the dimension of a pixel. If the diameter of the input mask is D_S (equal to the diameter of the lens L_S), the contrast of the fringes in the output plane will vanish if the number of optical modes is equal to the number of pixels in the input mask.

It is of course clear, that the input mask of our correlator can be itself a source having a certain extension and emitting in a certain solid angle. In this case too, it is possible to calculate the number of transverse modes of the source using Eq.(3.68). The filtering condition [Eq.(3.56)] can be written as

$$\frac{\lambda f\sqrt{N_m}}{D_S} + 2\frac{\lambda f}{\Delta} < \frac{\lambda f}{d} . \tag{3.71}$$

By taking the square, and after some straightforward manipulations, we obtain the two-dimensional condition

$$N_m < \left(\frac{D_s}{d}\right)^2 \left(1 - 2\frac{d}{\Delta}\right)^2 \tag{3.72}$$

for the number of modes. If the linear dimension of a pixel Δ is much larger than the period d of the grating, then the filtering condition is met if the number of optical modes is about equal to the number of grating cells having surface d^2, contained in the illuminated surface $(D_S)^2$.

3.6. Partially temporal coherent illumination

We consider now the case in which the illumination is spatially coherent and temporally partially coherent. This means that we illuminate the input mask of our correlator with plane waves having the same direction, but different wavelengths. In practice this can be approximated by collimating the light coming from a source having small dimensions (compared with the focal length of the collimating lens) and a spectral bandwidth $\Delta\lambda$. The source can also be for example a light emitting diode (LED) or a multi mode laser diode (LD).

3.6.1. Temporal coherence requirements for filtering

The use of diffraction gratings is the reason for the restrictions on the temporal coherence, i.e. the spectral width $\Delta\lambda$ of the light source. Thus we can say that, except for the zero-order, the higher orders diffractions are dispersed into a superposition of spectra having different wavelengths. The uncertainty of the position of the q-th diffraction order in Fourier plane due to non-monochromatic light is given by $q(\Delta\lambda)f/d$. (We use the notation $(\Delta\lambda)$ instead $\Delta\lambda$ to avoid confusion with the pixel size that we denoted with Δ). Again, the diffraction orders, which are now smeared, have to be separated. The spatial extension of each diffraction order when illuminated with light having wavelength λ is given by $2\lambda f/\Delta$ (distance between the two first zeros of the sinc function (see Fig.3.16a)). The extension of the q-th order is thus given by the extension due to the diffraction and the uncertainty due to the wavelength dispersion. This sum must be smaller than the separation between two diffraction orders. Mathematically we can write this condition as

$$q\frac{(\Delta\lambda)f}{d} + 2\frac{\lambda f}{\Delta} < \frac{\lambda f}{d} \quad . \tag{3.73}$$

The maximum dispersion is obtained for the highest diffraction order q_{max}. Thus by substituting q with q_{max} and after straightforward transformations we obtain the condition

$$\frac{(\Delta\lambda)}{\lambda} < \frac{1}{q_{max}}\left(1 - \frac{2d}{\Delta}\right). \tag{3.74}$$

for the spectral width of the illumination. If we like to use higher diffraction orders, we must increase the temporal coherence of the illumination.

3.6.2. Temporal coherence requirements to avoid undesired shifts of the images in the output plane

The complex amplitude in the output plane of the correlator, when the input mask is illuminated by a collimated beam having bandwidth ($\Delta\lambda$), is obtained by integration of Eq.(3.21) over the spectral bandwidth of the light source. In this case too, for explication purpose, we restrict us to the one dimensional case. We denote with $u_{i\lambda}(x_i)$ the complex amplitude in the input plane of the wave with wavelength λ. The distribution of the complex amplitude in plane (x_o), is given by

$$u_o(x_o)=\frac{1}{N}\int_{\lambda_0-\Delta\lambda/2}^{\lambda_0+\Delta\lambda/2}\sum_{m=-K}^{K} A_m u_{i\lambda}(x_o+\lambda am/d)\exp\{2\pi i(x_o m/d)\}\exp\{\pi i a\lambda(m/d)^2\}\exp\{i\phi(\lambda)\}\Big\}d\lambda \tag{3.75}$$

The phase term $\exp\{i\phi(\lambda)\}$ accounts for the fact, that plane waves with different wavelengths arrive in the output plane with different phases.

From Eq.(3.75) it appears clear that the spectral bandwidth $\Delta\lambda$ of the light source limits the performance of the correlator. In fact, images of different wavelengths are shifted in different manner in the output plane. In particular, images of the higher diffraction orders are more affected by this effect. The undesired shift of the q-th image is given by

$$\Delta\xi = \frac{q(\Delta\lambda)a}{d} . \tag{3.76}$$

Of course, $\Delta\xi$ must be much smaller than the pixel size Δ, otherwise the result of the correlation operation is canceled. The tolerable shift $\Delta\xi$ depends on the application for which the correlator is used. If we like to have well defined pixel at the output, we have to limit the source bandwidth. We denote the undesired shift $\Delta\xi$ as a fraction χ of the pixel dimension, viz.

$$\Delta\xi = \chi\Delta . \tag{3.77}$$

Using Eqs. (3.76),(3.77) and (3.25), we obtain the condition for the spectral bandwidth of the source as a function of the tolerance factor χ, the pixel dimension and the maximum diffraction order, namely

$$\frac{(\Delta\lambda)}{\lambda_0} = \frac{\chi}{q_{max}} . \tag{3.78}$$

If the grating that is used has, e.g., 21 diffraction orders (q_{max} =10), and we choose a tolerance factor $\chi = 0.1$, then $\Delta\lambda = \lambda_0/100$.

Eqs.(3.74) and (3.78) have a similar form. In spite of this their meaning is quite different. The first describes a filtering condition, and the second a condition for the superposition of the pixels in the output plane.

3.6.3. Fringe visibility in the output plane

The interference fringe patterns produced by each wavelength are incoherently superimposed. If the spectral bandwidth is large enough, the fringes disappear. We suppose now that only two orders are present. This allows us to simplify the study of the interference fringes. When we superimpose the 0 order and the q-th order, the intensity is given by

$$I(x_o) = \int_{\lambda_0-\Delta\lambda/2}^{\lambda_0+\Delta\lambda/2} \left\{ |u_{i\lambda}(x_o)|^2 + |u_{i\lambda}(x_o+\lambda aq/d)|^2 \right\} d\lambda +$$

$$+ 2 \int_{\lambda_0-\Delta\lambda/2}^{\lambda_0+\Delta\lambda/2} |u_{i\lambda}(x_o)|\,|u_{i\lambda}(x_o+\lambda aq/d)| \cos\left[2\pi i\left(\frac{x_o q}{d} + \frac{a\lambda q^2}{2d^2}\right)\right] d\lambda. \qquad (3.79)$$

Here we are only interested in the second integral, which gives rise to interference fringes. We suppose a source with a constant spectral intensity distribution in the wavelength range $(\lambda_0-\Delta\lambda/2, \lambda_0+\Delta\lambda/2)$, and since we are only interested in the visibility of the fringes, we suppose that $|u_{i\lambda}(x_o)|=1 \;\forall\; \lambda\in(\lambda_0-\Delta\lambda/2,\lambda_0+\Delta\lambda/2)$. For this case we can easily calculate the second integral of Eq.(3.79) and after some straightforward manipulations we arrive at the expression

$$I_t(x_o) = 2(\Delta\lambda)\cos(2\pi x_o q/d)\,\mathrm{sinc}\left(\frac{a(\Delta\lambda)q^2}{2d^2}\right). \qquad (3.80)$$

The contrast of the interference fringes is thus given by $V = |\mathrm{sinc}((\Delta\lambda/2)a(q/d)^2)|$. The fringes disappear when $(\Delta\lambda/2)a(q/d)^2=1$.

The phase term $\pi i a\lambda_0(q/d)^2$ in Eqs. (3.21) and (3.75) express the fact that the optical path lengths for the different orders are different. In particular between the 0-th and the q-th order the path difference is

$$l_q = \frac{q^2(\lambda_0)^2 a}{2d^2}. \qquad (3.81)$$

The contrast can thus be written as $V = |\mathrm{sinc}[l_q(\Delta\lambda)/(\lambda_0)^2]|$. The coherence length of a source having spectral width $\Delta\lambda$ is $\lambda^2/\Delta\lambda$. Thus, the coherence length, for which the fringes between the 0-th and the m-th order vanish, is $l_{q0} = (\lambda_0)^2/(\Delta\lambda)$.

The contrast of the fringes calculated above can be obtained in another way. We know (see e.g. Ref.4), that the contrast between two beams having spectral intensities I(ν) and time delay τ can be calculated using a Fourier transform. We substitute ν with $\nu_0+\nu'$, where ν_0 is the mean frequency ($\nu_0=c/\lambda_0$), and we define a spectral function $L(\nu-\nu_0)=I(\nu)$. The contrast is given by $|\gamma(\tau)|$, where

$$\gamma(\tau) = \frac{\int_{-\infty}^{\infty} L(\nu')\exp\{2\pi i\nu'\tau\}\, d\nu'}{\int_{-\infty}^{\infty} L(\nu')d\nu'} \qquad (3.82)$$

This expression is of the same form as the Eq.(3.49), which describes the complex coherence factor associated with a spatially extended source. In both cases a Fourier transform is involved. For a spectral intensity $I(\nu)=\mathrm{rect}[(\nu-\nu_0)/(\Delta\nu)]$ we have $L(\nu')=\mathrm{rect}[\nu'/(\Delta\nu)]$. Using Eq.(3.82) we calculate the contrast and we obtain

$$V = |\gamma(\tau)| = |\mathrm{sinc}[\tau(\Delta\nu)]|. \qquad (3.83)$$

The difference in path lengths between two orders can be expressed as $l_q = \tau c$ (where c is the light speed). Since $\lambda_0\nu_0 = c$ it follows that $(\Delta\lambda)/(\lambda_0)^2 = (\Delta\nu)/c$. By substituting in Eq.(3.83) τ with l_q/c and Δν with $c(\Delta\lambda)/(\lambda_0)^2$, we get $V = |\mathrm{sinc}[l_q(\Delta\lambda)/(\lambda_0)^2]|$. This is exactly the same expression for the contrast as obtained above.

3.7. Undesired effects related with filtering in the Fourier plane

In 3.5, we pointed out that the power spectrum of each order in Fourier plane has a certain extension, which is essentially given by the spatial extension of the source and the diffraction of the input pattern. Since the diffraction of the input mask contains high spatial frequencies, the various diffraction orders are partially superimposed, and by spatial filtering it happens that when we suppress one order, we perform an undesired cut of a part of the adjacent orders. This can have a dramatical effect, in fact by cutting some spatial frequencies of one order we affect the reconstruction of the corresponding image in the output plane. In this section, we investigate the effect of the undesired cutting of spatial frequencies for coherent and partially coherent illumination.

3.7.1. Spatially coherent illumination

We suppose, that the input mask is composed by square pixels (dimension $\Delta\times\Delta$). For a monochromatic spatial coherent illumination, we find for the zero order in the Fourier plane a complex amplitude described by Eq.(3.54) for the 1-dimensional case. For the other orders, we have similar complex amplitude distributions, multiplied by quadratic phase factors and shifted by $mf\lambda/d$.

In order to simplify the problem, we consider an input mask with only one pixel, and we calculate how it is reconstructed in the output plane if part of its spectrum is suppressed by filtering. Another simplification for this calculation is to consider only the zero order, but it is obvious that the other orders will produce a shift of the selected pixel in the output plane. We perform the calculation in one dimension, the x direction. The same calculation can be performed, of course, in the y direction. The input pattern is described by the transparency

$$t(x_i) = \mathrm{rect}\left(\frac{x_i}{\Delta}\right). \tag{3.84}$$

This transparency is illuminated by a monochromatic plane wave $s(x_i)=1$ (spatially and temporally coherent). Just behind the input mask we will have a complex amplitude $u(x_i)=s(x_i)t(x_i)=t(x_i)$. The complex amplitude in the Fourier plane of the pixel after the omission of the constant phase factors is given by

$$FT\{u(x_i)\}=U_i(x_f) = \Delta\ \mathrm{sinc}\left(\Delta\frac{x_f}{f\lambda}\right). \tag{3.85}$$

We can place a square filter having dimension $\mu\times\mu$ in the Fourier plane (see Fig.3.22). Just behind the filter the complex amplitude is

$$F(x_f) = \Delta \,\mathrm{sinc}\left(\Delta\frac{x_f}{f\lambda}\right)\mathrm{rect}\left(\frac{x_f}{\mu}\right). \tag{3.86}$$

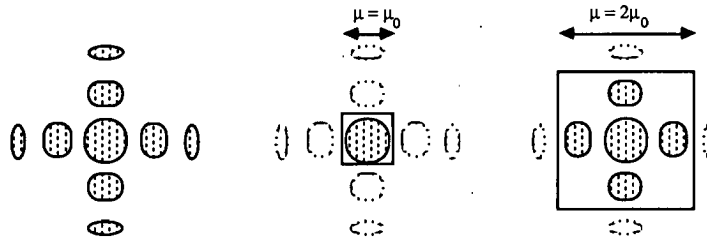


Fig.3.22. Filtering of a two dimensional sinc function in the Fourier plane. a) unfiltered sinc function, b) square filter having dimension $\mu_0 x \mu_0$, which cuts the sinc function at the first zero, c) square filter having dimension $2\mu_0 x 2\mu_0$, which cuts the sinc function at the second zero.

The complex amplitude in the output plane is given by the Fourier transform of $F(x_f)$. Since a product in the spatial frequency plane involves a convolution in the image plane, we obtain

$$u_0(x_0) = \mu \int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{x}{\Delta}\right)\mathrm{sinc}\left(\frac{(x_0 - x)\mu}{\lambda f}\right)dx\,. \tag{3.87}$$

The intensity in the image plane is $|u(x_0)|^2$. The integral of Eq.(3.87) was calculated by numerical integration. The results of the numerical estimation, for different dimensions μ of the filter, are presented in Fig.3.23. The result of Fig.3.23a) was obtained by choosing $\mu = \mu_0 = 2\lambda f/\Delta$, that corresponds to a filter having dimensions equal to the distance between the two first zeros of the sinc function, as shown schematically in Fig.3.22b). Notice that in this case, in the output plane we do not find an homogeneously illuminated pixel, but rather an intensity distribution that looks like a gaussian profile. The width of the pixel is larger than its original width Δ. Therefore it influences its neighbor, there is crosstalk between the pixels. By increasing the dimension of the filter, the reconstruction of the pixel intensity gets better and the crosstalk decreases, as shown in Fig.3.23b)...f). The crosstalk in the output plane depends on the quality of the filtering in the Fourier plane, and this depends on the distance between two adjacent diffraction orders. The required quality of the reconstruction depends of course on the application, in some cases it might be sufficient to have a reconstruction like that shown in Fig.3.23a).

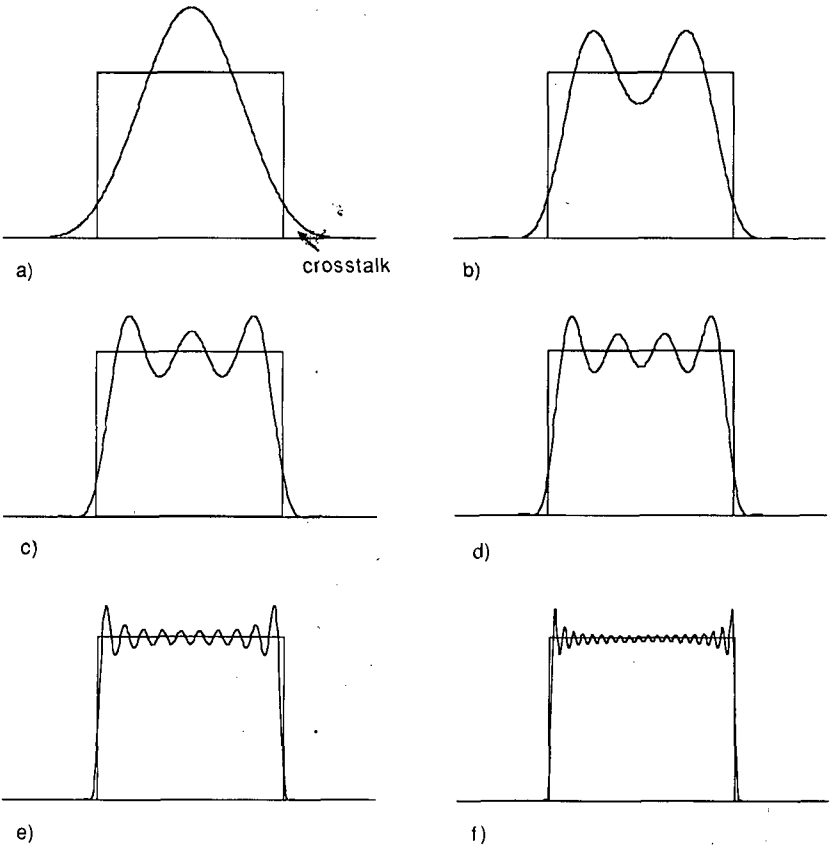


Fig.3.23. Intensity profiles in the output plane, obtained for different dimension of the filter, a) $\mu=\mu_0$, b) $\mu=2\mu_0$, c) $\mu=3\mu_0$, d) $\mu=4\mu_0$, e) $\mu=10\mu_0$ f) $\mu=20\mu_0$,

3.7.2. Partially spatial coherent illumination

We discuss now the formation of the image of a pixel, when the input is illuminated by monochromatic partial spatially coherent light. The optical arrangement which can be used to produce such an illumination is shown in Fig.3.15. The illumination can also be seen as superposition of uncorrelated plane waves having different directions of propagation, and can

be mathematically described by Eq.(3.46). In the Fourier plane we find an incoherent superposition of shifted sinc functions, described by the relation

$$U_i(x_f) = N_f \int_{-\infty}^{\infty} \mathrm{sinc}\left(\Delta\frac{x_f}{f\lambda} - \Delta\frac{x_s}{f_s\lambda}\right)\sqrt{I(x_s)}\exp\{i\phi(x_s)\}dx_s \ , \qquad (3.88)$$

where N_f is a normalization factor, $\exp\{i\phi(x_s)\}$ is a random phase factor, accounting for the fact that each point x_s of the source emits independently from the others, and $I(x_s)$ is the intensity at the point x_s of the source. In the following calculation we will put $I(x_s) = \mathrm{rect}(x_s/X_s)$, that is to say that the light source has dimension X_s and the intensity is uniform. If we place a filter of width μ in the Fourier plane, we obtain just behind the filter the complex amplitude $F(x_f)=\mathrm{rect}(x_f/\mu)U_i(x_f)$. The complex amplitude in the image plane is given by the Fourier transform of $F(x_f)$. Since a shift in the frequency plane involves the multiplication by a linear phase factor in the output plane, and a product in the spatial frequency domain involve a convolution in the output plane, we get

$$u_o(x_o) = \int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{x}{\Delta}\right)\mathrm{sinc}\left(\frac{(x_o-x)\mu}{\lambda f}\right)\int_{-\infty}^{\infty}\sqrt{I(x_s)}\exp\{i\phi(x_s)\}\exp\left(2\pi i\frac{xx_s}{\lambda f_s}\right)dx_s dx \ . \qquad (3.89)$$

The Intensity is also given by the square modules of $u_o(x_o)$, which yields

$$I_o(x_o)= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{x}{\Delta}\right)\mathrm{rect}\left(\frac{x'}{\Delta}\right)\mathrm{sinc}\left(\frac{(x_o-x)\mu}{\lambda f}\right)\mathrm{sinc}\left(\frac{(x_o-x')\mu}{\lambda f}\right)\int_{-\infty}^{\infty} I(x_s)\exp\left(2\pi i\frac{(x-x')x_s}{\lambda f_s}\right)dx_s dx dx' \qquad (3.90)$$

The third integral is the Fourier transform of the source and represents the mutual intensity function $J(x,x')$ of the illumination at points x and x'. Using a completely spatial coherent illumination, we obtain $J(x,x')$ = constant (Fourier transform of a delta function), and for a completely incoherent illumination $J(x,x') = \delta(x-x')$ (Fourier transform of a constant). This result inserted in Eq.(3.90) gives the well know formula for image formation with incoherent illumination (see e.g. Refs. 1 and 4). We are now interested in the partially coherent case, for which $I(x_s)= \mathrm{rect}(x_s/X_s)$. In this case we obtain

$$J(x,x') = \int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{x_s}{X_s}\right)\exp\left(2\pi i\frac{(x-x')x_s}{\lambda f_s}\right)dx_s = \frac{X_s}{\lambda f_s}\mathrm{sinc}\left((x-x')\frac{X_s}{\lambda f_s}\right). \qquad (3.91)$$

This result is inserted in Eq. (3.90), and some numerical integration were made, for different source extension and different widths m of the filter in the Fourier plane. Figure 3.24 shows

the result of the calculation for various filter widths (the values of m were chosen equal to those of Fig.3.23), and for a source width $X_S=\lambda f_S/\Delta$.

It appears clear from this results that, using the same filtering, the images are better reconstructed with partially coherent, than with completely coherent light. Notice that using a more incoherent illumination than in the cases represented in Fig.3.24, the fluctuations tend to vanish.

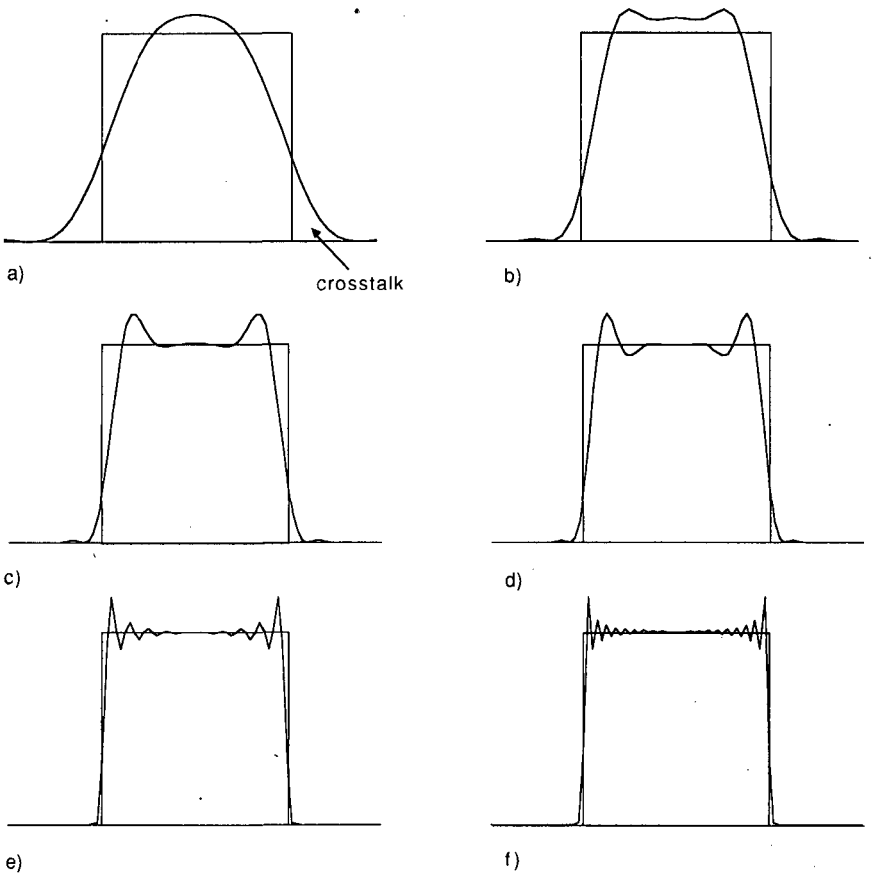


Fig.3.24. Intensity profiles in the output plane, obtained with partially coherent illumination produced by a source of width $X_S=\lambda f_S/\Delta$, for different dimensions of the filter, a) $\mu=\mu_0$, b) $\mu=2\mu_0$, c) $\mu=3\mu_0$, d) $\mu=4\mu_0$, e) $\mu=10\mu_0$ f) $\mu=20\mu_0$,

3.7.3. Undesired effects appearing when the spatial coherence of the input is increased by using spatial filtering

Consider the arrangement described in Fig.3.17, in which we used two 4f spatial filtering systems, the first to increase the spatial coherence of the input, and the second for the correlator. We suppose to have a completely spatially incoherent image and we increase the coherence degree using this arrangement. The pixels in the input plane of the correlator (after the first 4f system), are partially deformed by the first spatial filtering; in fact the filtering suppresses some spatial frequencies. The pixels are no more distorted in the second 4f system if the cut-off frequency of the filter of width μ, is larger than the cut off frequency of the filter of width W. To obtain the pixel reconstruction in the output plane of the correlator, it is sufficient to know it in the input plane after the first spatial filtering. This was done by numerically calculation of Eq.(3.90) for different width of the filter W. The results are represented in Fig.3.25. $W_0=2\lambda f/\Delta$ is the distance between the two first zeros of the sinc function, which represents the Fourier transform of the pixel. The results of Fig.3.25 are also directly comparable with the results of Figs.3.24 and 3.23, it is sufficient to substitute W_0 with μ_0.

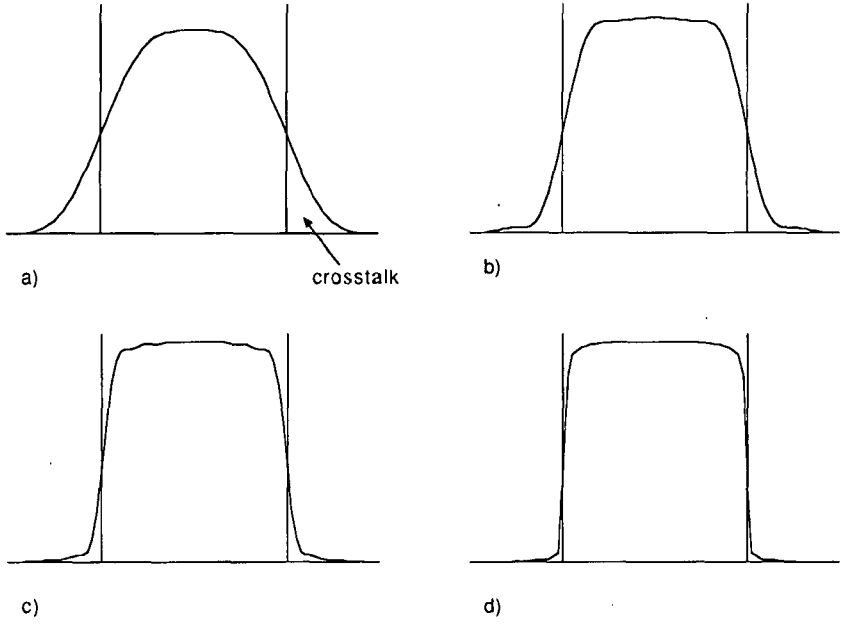


Fig.3.25. Intensity profiles, obtained after spatial filtering of a totally incoherent image of a pixel, for different dimension of the filter, a) $W=W_0$, b) $W=2W_0$, c) $W=4W_0$, d) $W=10W_0$.

3.8. Space-bandwidth product of the correlator

In characterizing and comparing different processing systems, the space-bandwidth-product (SBWP) is used to describe the computational capability of a system. For an input encoded using array cells (pixels), the SBWP is equal to the number of pixels that can be processed in parallel by the system. Thus in our system, the SBWP equals the number of processable pixels in the input plane. In Ref.20 is investigated the SBWP of the spatial filtering methods for performing binary logic, and is pointed out that the number of processable pixels is dependent on the number of orders which can be independently filtered in the Fourier plane. We calculate now the SBWP for our correlator and we will show that it is dependent on the complexity of the function which we like to correlate with the input, on the encoding and on the number of parallel channels.

3.8.1. SBWP of a single channel correlator

In the case of the single channel correlator, the input pattern is correlated with a single discrete function. In the calculation reported in section 3.2 we denoted this function with $g_d(x,y)$. The complexity of $g_d(x,y)$, i.e. the number of delta pulses contained in g_d, determines the number of orders in the Fourier plane. Consider the case in which g_d consists of NxN delta pulses. In the Fourier plane of the system, where the filtering of the various orders takes place, it is necessary to have an adequate separation of adjacent orders so that the overlap is negligible. Due to the size Δ of the input pixels, each order in the Fourier plane will have a finite dimension $\lambda f/\Delta$ (distance of the first zero of the sinc function from the center, see Fig.3.16a). To avoid overlap of adjacent orders, a separation of the orders in the Fourier plane of $\sigma\lambda f/\Delta$ is achieved by selecting gratings with σ times the spatial frequency $1/\Delta$ of the input mask. The value of the "security factor" σ determines the quality of the image in the output plane, in section section 3.7 we have shown that it must be at least 2.

It is assumed for this analysis that the distribution of the orders is that of a square, i.e. an equal number of orders in the x and y directions. The input mask is also considered to be a square with a side dimension of Y. In our 4f system the clear aperture of the lenses must be large enough to capture all of the diffracted beams (Fig. 3.26). Based on this criteria, the required clear aperture D is given by

$$D \geq \sqrt{2}Y + \sqrt{2}\frac{\lambda f}{\Delta} + \sqrt{2}\,(N-1)\sigma\frac{\lambda f}{\Delta}. \qquad (3.92)$$

The first term on the right hand side of the inequality is due to the size of the input mask. The second term is due to the diffraction resulting from the pixel spacing in the input mask. The

third term is dependent on the number N of diffraction orders to be operated on in the Fourier plane. This term reflects the increase in the lens diameter needed to capture all the orders resulting from the diffraction gratings placed after the input plane. In this equation the spacing *a* between the input plane and the gratings has been assumed small compared to the focal length f.

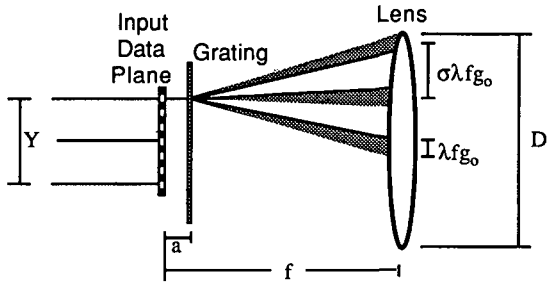


Fig.3.26. Required minimum lens diameter resulting from diffraction at the input plane and at the grating.

Based on the above requirement for the diameter, the minimum value for the pixel spacing Δx, can be calculated in terms of the other system parameters. The maximum SBWP is defined as $(Y/\Delta)^2$. The space bandwidth product for a correlation function of N^2 delta pulses follows readily from this analysis and is represented by

$$SBWP_N \leq \left\{ \frac{D}{f} \frac{Y}{\lambda} \frac{\left[1 - \sqrt{2}\frac{Y}{D}\right]}{\sqrt{2}\left[1+\sigma(N-1)\right]} \right\}^2 \tag{3.93}$$

As seen in the above equation, the SBWP is inversely proportional to the f-number of the system (f/# = f/D) and directly proportional to the scaling Y/λ, which represents the relative size of the system with respect to the wavelength of the illumination. By taking the derivative of the above equation with respect to the input dimension Y, it is easily seen that the maximum SBWP is achieved for $Y = D/(2\sqrt{2})$, which simply states that for a single channel system the input is of optimum size, if its diagonal is equal to one half the diameter of the lens used.

To get an approximate idea of the available SBWP and the effect of decreasing the system f/#, consider the following comparison between an f/5 and an f/2 system. In both cases the focal length is 200 mm, whereas the diameters are 40 mm and 100 mm, respectively. In the calculation it is assumed that the input mask has the optimal diameter and that the light source has wavelength λ = 500 nm. The calculations are performed for two cases N=21 and N=2. In

all cases the "security factor" σ was chosen equal to 4 to insure adequate separation of the orders within the Fourier plane. The results of this analysis are presented in the Table 3.1.

System f/#	N=21		N=2	
	SBWP	Pixel Spacing	SBWP	Pixel Spacing
f/5	625	736 μm	1.6×10^5	35 μm
f/2	2.3×10^4	222 μm	6.2×10^6	14 μm

Table 3.1. Computational capability of f/2 and f/5 systems, for a security factor $\sigma = 4$, search symbols with $N^2 = 21x21$ and $N^2 = 2x2$ pixels.

Suppose to use the system for pattern recognition, thus Table 3.1 shows that a f/5 system can recognize a symbol composed of 21x21 pixels in an input pattern composed of 625=25x25 pixels with size 736 μm. In this example the dependence on the system f/# is clearly illustrated, in fact if we like to recognize the same number of pixels (21x21) using a f/2 system we can have an input with 154x154 pixels with size 222 μm. The SBWP is dramatically increased if the search symbol is composed by few pixels, e.g. N=2. It should be noted that the resulting optimal pixel spacing is quite small for most cases.

3.8.2. SBWP of the correlator for various encodings of the input

For single rail, intensity and polarization encoding, the SBWP is given by Eq.(3.93).

For dual rail encoding, we need two input masks (a positive and a negative), and in Fourier plane we have twice the number of orders that we had for the single rail correlator. This decreases the SBWP by a factor 4.

For the theta modulation encoding correlator, we have a single input mask in which the pixels are encoded using gratings having different orientations. To calculate the SBWP, we have to distinguish between different kinds of gratings used for encoding. In fact when we use amplitude only gratings (for example Ronchi gratings), each orientation of the grating inside the pixels produce at least three diffraction orders, from which the zero order is common for all the encoding orientations, for this reason it cannot be used and has to be suppressed in the filter plane. If ν is the number of levels (grating orientations) in the input pattern and, if we consider the input mask composed by amplitude only gratings producing three diffraction orders, we obtain

$$SBWP_{theta} = \frac{SBWP_N}{2v + 1} . \tag{3.94}$$

Where the "+1" account for the zero orders, common for all the orientations of the gratings used to encode the pixels in the input mask. Using phase only gratings, it is possible to eliminate the zero order, and it is possible to obtain only one diffraction order, in this case the SBWP is equal to $SBWP_N / v$.

3.8.3. SBWP of a multi-channel correlator

In the case of a multi-channel system, more than one NxN pattern is being recognized in parallel. Therefore another grating is used to achieve the physical separation of the different channels both in the Fourier plane and in the output plane (see Fig.3.2). To achieve adequate separation of the MxM orders, representing the M^2 channels in the Fourier plane, the frequency of the grating must be at least $\sigma N/\Delta$. The distance a' between this additional grating and the input mask is chosen to achieve a separation of the individual channels at the output plane by at least the dimension Y of the input mask, to avoid overlap. The distance a' necessary to achieve a channel spacing of Y in the output plane with the grating frequency equal to $\sigma N/\Delta$ is found similarly to Eq.(3.25) as

$$a' = \frac{Y\Delta}{\lambda\sigma N} . \tag{3.95}$$

As a result of the above multi-channel configuration with M^2 channels, the SBWP is now limited by the diameter of the second lens of the 4f system. Analogous to Eq. (3.93) the resulting SBWP is

$$SBWP_{M,N} \le \left\{ \frac{D}{f} \frac{Y}{\lambda} \frac{\left[1 - \sqrt{2}\, M \frac{Y}{D}\right]}{\sqrt{2}\left[1+\sigma(N-1)+\sigma N(M-1)\right]} \right\}^2 . \tag{3.96}$$

It is apparent that the same dependence on the f/# and the system size exists also for the multi-channel system. From the above equation the optimum dimension of the input is found to be $Y = D/(2M\sqrt{2})$, M times smaller than that of the single channel system. A rough approximation of the effect of the multiple channels on the available SBWP is

$$SBWP_{M,N} \approx \frac{1}{M^4} SBWP_N . \tag{3.97}$$

Table 3.2 presents the result of an analysis of a system with M^2= 3x3 channels and N=2. The focal length f was chosen equal to 200 mm. The second grating, which produces the separate channels, was placed at a distance of 176 mm from the input mask.

System f/#	N=2 M=3	
	SBWP	Pixel Spacing
f/5	1.0×10^3	148 μm
f/2	3.9×10^4	59 μm

Table 3.2. Computational capability of f/2 and f/5 systems, for a security factor $\sigma = 4$, search symbols with N^2 = 2x2 pixels and M^2=3x3 channels.

4. IMPLEMENTATION OF THE CORRELATOR

Until now we have described the different features of the optical discrete correlator without to say anything about the elements which we need for its practical implementation. In this chapter we discuss how we can produce the diffraction gratings and which spatial light modulators (SLMs) can be used to implement the input mask and the filter.

4.1. Diffraction gratings

We have seen that to produce multiple copies of the input pattern, we need gratings with a certain number of diffraction orders of the same intensity. To calculate the transmission of a grating with a certain number of diffraction orders having the same brightness, one can take the inverse Fourier transform of a set of delta functions arranged on a regular grid. In general this will result in a complex periodical shape that will not be very efficient and too complicated to realize in practice. For this reason we have to make some restrictions to the grating structure. In order to have high efficiency, we restrict our self to phase only gratings, this allows in theory to have high efficiencies. We discuss now some types of phase-only gratings.

4.1.1. Linear gratings (1-D)

Linear gratings have a transmission function

$$t(x,y) = \exp\{-i\phi(x)\} \tag{4.1}$$

depending only on one dimension, where $\phi(x)$ is a function having period d_x. If a central symmetrical block of N equally bright diffraction orders is desired, then the relations between the diffraction orders are

$$|T(p_{x-K})|^2 = ... = T(p_{x0})|^2 = ... = |T(p_{xK})|^2, \tag{4.2}$$

where $T(p_x) = FT\{t(x)\}$ and $p_{x-K}...p_{x0}...p_{xK}$ are the abscissa of the different orders in the Fourier plane (K = (N–1)/2). There are of course many functions $\phi(x)$ satisfying this relation, but in our case we have to choose phase profiles that are easy to realize. For this purpose we consider only sinusoidal phase gratings and binary phase gratings (Dammann gratings [16]). A sinusoidal phase grating is defined by the transmission function

$$t(x,y) = \exp\{i\, A \sin(2\pi x/d)\}, \tag{4.3}$$

where A is the half of the peak to peak excursion of the phase delay. For this grating, the intensity of the n–th diffraction order is given by the square of the Bessel functions (see e.g. Ref.[1]). Thus

$$I_n = J_n^2(A) \tag{4.4}$$

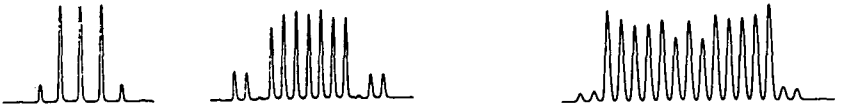
It is possible to choose A such that the three central diffraction orders have the same brightness

$$J_{-1}^2(A) = J_0^2(A) = J_1^2(A) = 0.30 \tag{4.5}$$

this means that about 90% of the energy arriving on the grating is diffracted in three orders. We define the efficiency

$$\eta = \frac{\text{radiant flux in the equally bright orders}}{\text{total radiant flux passing the grating}} \tag{4.6}$$

To realize sinusoidal phase gratings we used a holographic method. The interference of two plane waves is recorded on a silver halide photographic plate. After developing, the plate was bleached. Fig.4.1a) shows the result of the measured intensity through a phase only holographic grating.



a) b) c)

Fig.4.1. Measured intensity profiles through a) holographic grating, b) binary phase grating on bleached silver halide photographic plate, c) binary phase grating fabricated by PSI–laboratories.

If the fan-out does not exceed three, the ease of fabrication and the high obtainable space bandwidth product (SBWP) favor sinusoidal holographic gratings on bleached silver halide photographic plates. For higher order fan-outs, binary Dammann gratings [16] can be used. These gratings have a binary phase function $\phi(x)$. There exist many publications about Dammann gratings; in Ref.[17] the binary phase functions for 4 to 29 diffraction orders with high efficiency are reported. Of course, the complexity of the phase profile increases with the increasing number of diffraction orders.

To realize the phase only gratings we used two techniques. For fan-out up to 7 and small SBWP (< 2400), the gratings were drawn on a Laser printer, photographically reduced and bleached. Fig.4.1b) shows an intensity profile through a diffraction grating having seven orders of the same brightness and an efficiency of about 70%. These gratings, however, suffer from non-uniformity and diffusion. High quality binary transmission phase gratings with surface relief have been kindly fabricated for our purpose by the PSI-laboratories[18]. Master gratings were drawn with a laser scanner at 1 μm resolution and then copied onto glass substrates covered with a 510 nm thick layer (phase modulation π at $\lambda = 633$ nm) of the transparent photoresist AZ 1400-17. Fig.4.1c) shows an intensity profile through a diffraction grating having 13 orders of the same brightness and an efficiency of about 75%.

4.1.2. 2-D gratings

Dammann gratings are not restricted to one dimension, but can have a two dimensional structure. We did not realize 2–dimensional Dammann gratings. Instead, to produce NxN diffractions beams we simply placed two 1– D gratings in contact, and rotated one grating by an angle ψ with respect to the other. This is equivalent to a 2-D periodic transparency. The efficiency η is then given by the product of the single efficiencies of each grating ($\eta = \eta_1\eta_2$). In Fig.4.2a) is represented the power spectra of two identical holographic diffraction gratings with $\psi = 30^0$. In this case, each grating had three equal diffraction orders. Fig.4.2b) shows the power spectrum of two Dammann gratings, each having 13 orders, rotated by 90^0 one with respect to the other.

a) b)

Fig.4.2. a) Spectrum of two holographic gratings rotated by $\psi = 30^0$, b) spectrum of two Dammann gratings rotated by $\psi = 90^0$.

4.2. Implementation of the input mask and the filter

Since our correlator is working with different encodings, and with either coherent or partially coherent light, the possibilities to implement the input pattern and the filter using the existing techniques are numerous. It is clear that the described correlator becomes interesting only when the input pattern and the filter can be reconfigured in real time. For this reasons we need spatial light modulators (SLMs). In some applications, it is not necessary that both the transparencies (the input mask and the filter) are reconfigurable in real time, this allows the use of photographic plates, which have better resolution than the existing SLMs. Most of the well known SLMs can be used to implement the input mask and the filter of the described correlator. Some of them were available in our laboratory and were used to obtain the experimental results described in section 3.3. Other devices were not available and the possibility of their application was theoretically investigated. We discuss now some SLMs and we will point out how we can use them for the implementation of the correlator.

4.2.1. Liquid Crystal Displays (LCDs)

Liquid crystal displays (LCDs) are now a widely used commercial product, they are electronically addressable, inexpensive, and their quality is constantly increasing. Their spatial resolution (size of the pixels) can achieve one or two tenths of a millimeter. The response time of each pixel is typically 50 ms. Actually it is possible to buy ferroelectric LCD which have a typical pixel response time less than 100 μs. An LCD can be illuminated either with spatially coherent light or partially spatial coherent light, and can be used as intensity or polarization encoded mask. In our correlator they can be used as programmable input pattern or filters in the Fourier plane.

We had in our laboratory two LCDs available which could be used in transmission mode, both were obtained by transformation of commercially available LCDs [23].

The first is a 32x84 element graphic LCD [21]. It has square elements of 0.8x0.8 mm^2 area, which are 1/16 multiplexed. It has been modified from its original reflection mode for use in transmission. The polarizer sheets of the LCD have been removed. The display yields a difference in optical activity of about 60^0 between the pixels in the "off" and "on" state. For an appropriate polarization of the incoming plane wave (parallel to a principal axis of the birefringent LCD), the pixels are linearly polarized with an angle of 60^0 between "on" and "off" state.

The second is a 5x7 LCD [22]. It has square elements of 1x1 cm^2 area, which are statically driven. After removal of the diffusely reflective tape and the polarizer sheets, the display can be used as a shutter array with programmable optical activity. The supply voltage of the 5x7 LCD

was adjusted to yield 90^0 and 30^0 optical activity for the "off" and the "on" elements, respectively.

4.2.2. Liquid Crystal Light Valve (LCLV)

One of the most useful real time optical addressable SLMs is the liquid crystal light valve (LCLV). The LCLV is capable of converting an incoherent image into a coherent image. In our correlator we need a coherent or partially coherent input pattern, the LCLV can thus be used to increase the coherence and the intensity of an input. Another characteristic of the LCLV is the fact that it can convert a binary intensity encoded image into a polarization encoded image. The LCLV has a nonlinear response, it can thus be used to restore the binary form in the output plane of our correlator.

In our laboratory we have an LCLV which has been developed at ASULAB (Neuchâtel). The basic elements of the ASULAB LCLV are shown in the simplified schematic of Fig.4.3.

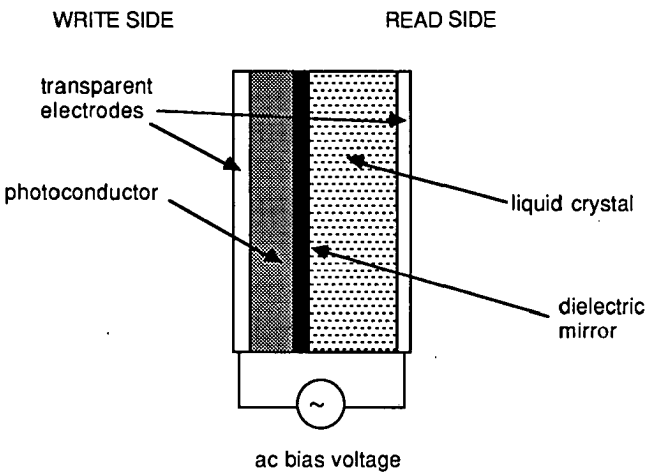


Fig.4.3. Simplified schematic representation of an LCLV.

The LC layer is placed between two transparent electrodes, where an ac bias voltage is applied to activate the device. The clear aperture of the LCLV is circular and has a diameter of 35 mm. The electric field in the LC layer can be optically modulated through a CdS photoconductor layer, which is placed between the two electrodes. The region of high spectral sensitivity of the CdS photoconductor is from 550 to 700 nm. If the write light (image projected on the back side of the valve) illuminates the photoconductor, it causes a decrease of its resistivity and thus an increase of the electrical field in the LC. The molecules of the liquid

crystal are strongly birefringent and carry a permanent dipole-moment, their orientation is thus changed by the electrical field. By this way, the birefringence of the light valve can be locally modulated by optical write in. The spatial modulation of the birefringence can be read out by reflection of a plane, polarized read out beam at the dielectric mirror, which is placed between the liquid crystal and the photoconductor. The state of polarization of the read out beam will be altered locally after the double pass through the LC layer.

A. Some important characteristics of the ASULAB LCLV

Various characteristics of the ASULAB LCLV, such as sensitivity, switching speed, spatial resolution, and wavefront distortion have been measured. Figure 4.4 shows the sensitivity measurements for three different supply voltages V_{eff} = 8.5 V,10 V and 12 V, and a supply frequency of 600 Hz. The sensitivity of the light valve increases with increasing supply voltage and varies for the above voltages between 300 $\mu W/cm^2$ and 30 $\mu W/cm^2$. It has further been observed, that the sensitivity decreases with increasing supply frequency. The frequency was chosen to be 600 Hz as a compromise between a good sensitivity and a minimum ac-component in the read out light.

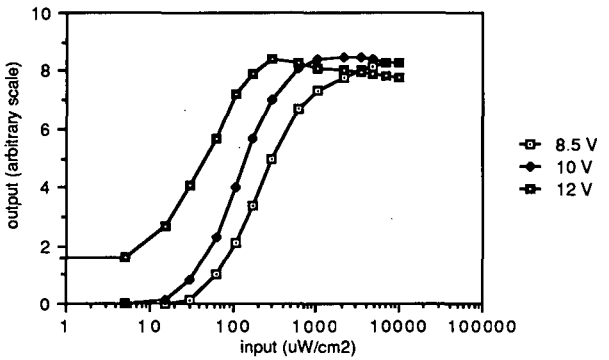


Fig.4.4. Response of the LCLV for different supply voltages V_{eff} = 8.5 V, 10 V and 12 V with a supply frequency of 600 Hz.

The spatial homogeneity of the optical sensitivity over the whole aperture was ± 15% for a write light intensity 400 $\mu W/cm^2$ (V_{eff} =10 V). The homogeneity improves with increasing write light intensity, i.e. when the light valve is saturated (± 10% for 1 mW/cm^2).

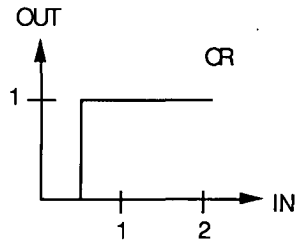
For the typical parameters (supply voltage V_{eff} = 8.5 V and write intensity 0.4 mW/cm^2), the rise and fall times (on and off switching) were found to be τ_{on} =170 ms and τ_{off} = 140 ms.

The switch-on time decreases with increasing intensity of the write light, whereas for the switch-off time no significant change was observed.

To determine the spatial resolution of the LCLV, an incoherent (white light) 1:1 image of an USAF test target has been projected onto the photoconductor. The spatial resolution was found to be 20 line pairs / mm.

B. Implementation of optical logic gates using a LCLV

Boolean logic can be realized using a nonlinear function. Figure 4.5 shows the ideal nonlinear function to implement two-input logic operations and the corresponding truth tables.



A	B	A OR B	A NOR B
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

Fig.4.5. Ideal nonlinear functions for the implementation of two-input logic OR and NOR (NOT) operations and corresponding truth tables.

Although no physical mechanism will achieve the ideal threshold switching characteristics of Fig.4.5, useful optical logic gates can be realized with physical devices having a nonlinear response.

The ASULAB-LCLV has a nonlinear response (Fig.4.4) which allows the implementation of both NOR and OR gates, depending on what polarization one takes in the output beam. In chapter 5 we will see how this nonlinear characteristic can be used to build a symbolic substitution processor composed of two discrete correlators.

4.2.3. Cathode Ray Tube (CRT)

We examine now, whether or not the use of a high performance cathode ray tube (CRT) would be possible for the described correlator. We do not have a CRT in our laboratory, for this reason experimental investigations were not possible. To get an approximate idea we report some numerical estimations obtained using the technical data of a "High Performance 1" CRT Assembly " produced by Westinghouse Electric Corporation [27]. This CRT has a screen area of 15.7x11.4 mm^2, an emission wavelength λ=525 nm with spectral bandwidth $\Delta\lambda$=25 nm. Its

luminance is of 3400 cdm^{-2}, which correspond to a radiance of 5.5 $Wm^{-2}sr^{-1}$ in radiometric units. Since the emission of light is completely spatially incoherent, the spatial coherence must be increased and this involves a waste of power. The purpose of our investigation is to determine if after spatial filtering there remains a detectable amount of light in the output plane of the correlator. As a detector we choose a relatively sensitive CCD-camera which needs for an acceptable image 0.05 lux [28]. We use the optical arrangement of Fig.3.17, with a square filter of surface WxW, to increase the spatial coherence of the image. After filtering, the image in (x_i,y_i)-plane (input plane of the correlator), can be seen as a source emitting into the solid angle $\Omega = (W/f)^2$. Consider an example with f = 100 mm and $W^2 = 4x4mm^2$, thus $\Omega = 1.6x10^{-3}$ sr. Using the data of the CRT we will have in plane (x_i,y_i) an image of the CRT screen with an intensity $I(x_i,y_i) = 1.6x10^{-3}x3400$ $lm/m^2 = 5.44$ lux, that correspond in radiometric units to $8.8mW/m^2 = 0.88\mu W/cm^2$! This image is divided by the diffraction gratings in NxN copies, and to allow the detection of the correlation operation each copy must have a detectable intensity >0.05 lux (sensitivity of the CCD camera). In our example the maximal number of copies is about 100 (5.44/0.05). It is possible to choose a larger filter (WxW>4x4 mm^2), in this case we will have more intensity but since the dimension of each diffraction order in the Fourier plane (x_f,y_f) is equal to WxW (image of the filter located in plane x_{fs},y_{fs}), and the orders must be spatially separated, their number is limited. In fact the total extension of the spectrum in the Fourier plane must have a reasonable dimension, typically equal to a few cm for a lens with focal length f = 100 mm. The other problem is given by the fact that to assure the separation of the spectra with large extension in Fourier plane we must choose gratings with high spatial frequency. This means that the diffraction angles are large and the aperture of the lenses must be large enough to capture all of the diffracted orders.

In conclusion we can say that the CRT can be used in the discrete correlator, but the complexity (number of delta pulses) of the function to correlate with the input is limited by the sensitivity of the detector or by the dimension W of the filter, and we can only recognize symbols composed by a few pixels (e.g. 10x10).

4.2.4. Implementation of a spatial frequency encoded input

The implementation of a theta encoded reconfigurable input transparency represents a problem. In fact it is possible to transform a multilevel intensity encoded image into a theta encoded image, by encoding each gray level of the input with a grating having different orientations [26], but for the moment this process cannot be performed in real time. The VGMLCD [29] is capable of performing in real time an intensity to spatial frequency conversion, where each gray level is transformed into a grating with a certain spatial frequency but the same orientation of the grating lines. This device could be used as input mask for the discrete correlator.

5. EXPERIMENTAL RESULTS AND APPLICATIONS

5.1. Discrete Binary correlator for pattern recognition

In chapter 3 we described the discrete correlator and we presented some simple experimental results. In this section we present some results which illustrate the correlator capability to recognize patterns composed of many pixels. The maximum size of the correlation function (in number of pixels) is limited to the number of diffraction orders of the gratings available. The appropriate combination of several gratings permits, in principle, to achieve an arbitrary number of diffraction orders. As discussed in section 3.8, the maximum SBWP of the input pattern is seriously limited if large correlation functions and thus a large number of diffraction orders are to be used.

The experimental results, that we will describe in this section, were obtained using coherent illumination. In particular the input pattern was illuminated by a collimated He-Ne laser beam. When the illumination is coherent, we always have interference fringes in the output, but in the presented results their period is very small and they are not resolved by the detection system. It is to emphasize that the same results could be obtained using partially coherent illumination. In this case we could suppress the interference by appropriately choosing the degree of coherence of the illumination, and we could at the same time suppress the coherent noise (see Figs. 3.20 and 3.21).

5.1.1. Single channel correlator

Figure 5.1 shows some experimental results of correlation. The input pattern was a 20x20 mm^2 photographic mask with 32x32 pixels. The 4f optical system was composed of two identical f/5 Fourier lenses with 38 cm focal length. Two pairs of appropriately positioned holographic phase gratings yield 9x9 equally spaced diffraction orders. The output image was projected onto a transparent screen and observed with a TV camera. Figure 5.1b) shows the correlation of the input image a) with the 9x9 correlation pattern representing a tree, as seen on the TV monitor. From the intensity profile, which is generated by a video-analyzer along the vertical line across the image, the good contrast of the dark correlation spots can be seen. Figure 5.1c) shows the correlation of the input image a) with the 9x9 correlation pattern representing a house. Two dark correlation spots in the correlation plane indicate the presence of two houses in the input image. Figure d) and e) show the results of correlation between the input image and a pattern representing the same house. In the two figures are taken the intensity profiles along two different vertical lines. Note that in Fig.5.1e) the correlation peak is less pronounced, this indicates that the symbol in the input image does not match exactly with the search symbol, in spite of this the presence of a peak indicates a resemblance between the two functions (two different views of the house).

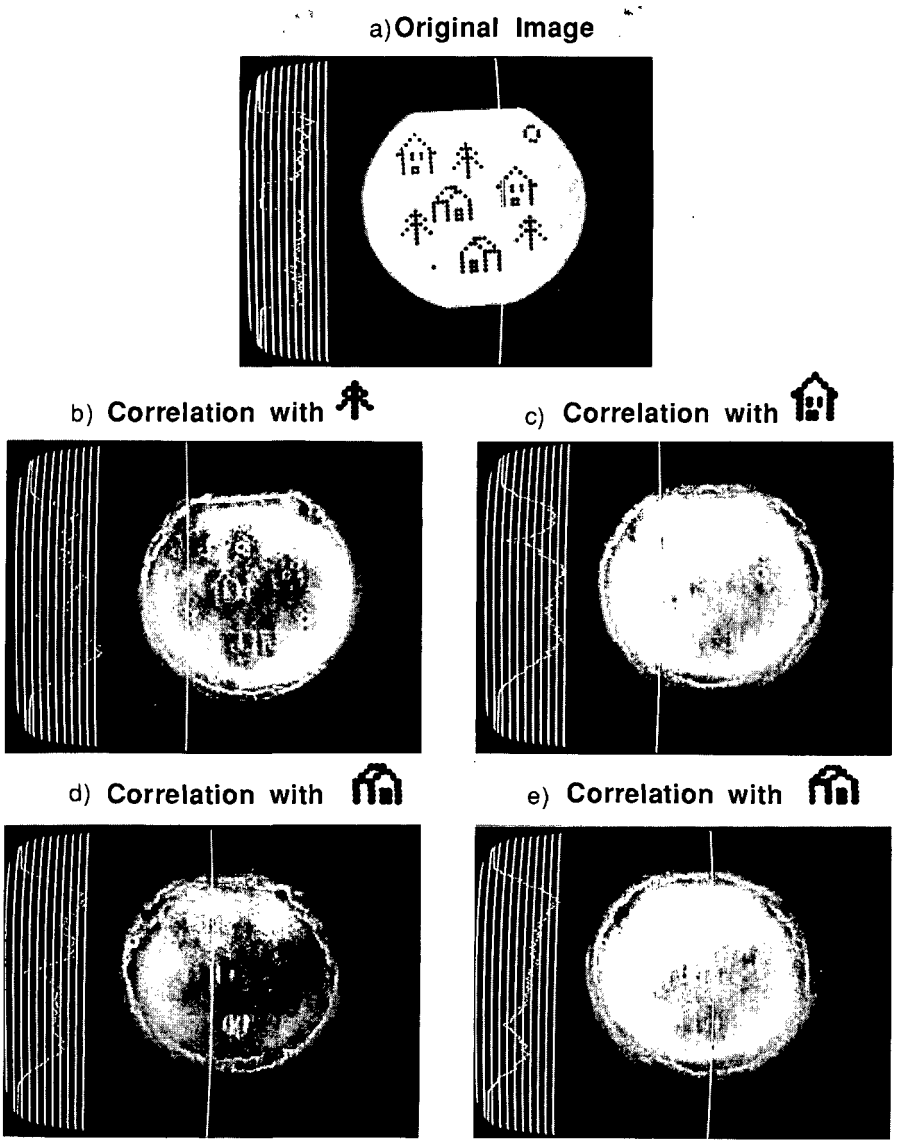
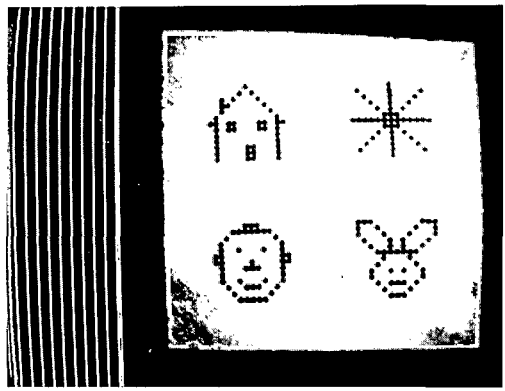
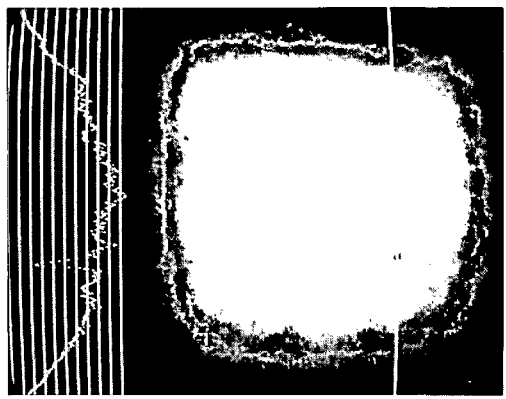


Fig.5.1. Example of recognition of different symbols contained in the input data pattern a). The intensity profile along the vertical white line within the image appears at the left and illustrates the contrast of the recognition.

Figure 5.2 shows another example of correlation. The input mask of Fig.5.2a) contains 50x50 pixels on an area of 16x16 mm^2. Two crossed binary gratings with 13 equal diffraction orders offer the possibility to program a correlation function with 13x13 pixels. The result of the correlation of the rabbit with the input pattern is presented in Fig.5.7.b.



a)



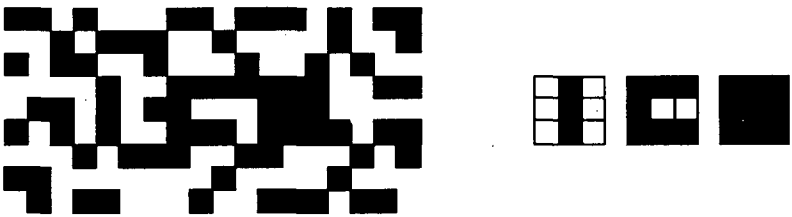
b)

Fig.5.2. The correlative search for the rabbit within the input pattern a) yields a black spot in the lower right corner of the image b).

5.1.1. Multi-channel correlator

The recognition of different symbols at the same time can be implemented with a correlator having multiple channels. In our set-up, multiple channel recognition can be easily realized by introducing an additional grating in order to produce multiple spatially separated images in the output plane (cf. section 3.3).

The following example illustrate at the same time multiple channel and dual-rail recognition. Figure.5.3a) shows the original binary mask, and Fig. 5.3b) shows the three symbols "I", "C" and "O", which are all to be recognized. Note than the symbols "I" and "C" are both contained within the "O" and thus dual rail recognition is necessary to distinguish them. Since we use a dual-rail system we have the representation of the positive and the negative input one beside the other in the input mask. In the experiment each of the input had dimensions 9.2 x 4.6 mm and were illuminated using spatially coherent light. Fig.5.4a) shows a photograph of the input with the positive and the negative pattern. The result of appropriate filtering for recognizing the three symbols is shown in Fig.5.4b). A dark pixel appears at the location of the recognized symbol.



a) b)

Fig.5.3. Original mask (a) and search symbols "I","C","O" (b), which will be recognized at the same time in three channels.

a) b)

Fig.5.4. Experimental results of multiple channel dual-rail recognition. a) input mask containing a positive and a negative pattern. b) Recognition of the three symbols.

5.2. Symbolic Substitution

Symbolic substitution is a concept for performing parallel and space-invariant logic operations on a binary input pattern [5], [6],[7],[8]. It allows the recognition of all occurrences of a set of search symbols (predefined spatial distributions of binary values) and substitutes them with other predefined symbols. The input and output data are usually spatially discrete binary patterns, coded either in intensity or in polarization. An example of symbolic substitution is shown in Fig.5.5. The pattern being searched is called the left hand side (LHS) of the transformation rule and the pattern that replaces it is called the right hand side (RHS). The pixels which do not contribute to a LHS disappear after the rule is applied.

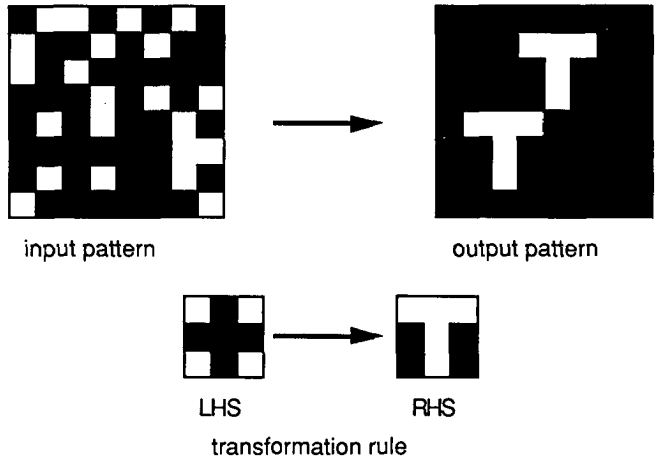


Fig.5.5. Symbolic substitution: a spatial configuration is recognized in the data plane (input) and substituted with another spatial configuration according to a transformation rule.

Many applications of symbolic substitution have been presented, in most of them more than one transformation rule are applied in parallel to the data input. In Ref.6 are presented the four rules that allow the addition of two binary numbers. Other rules were developed to implement Boolean logic, cellular logic and a Turing machine, and it was demonstrated that a computer based on symbolic substitution is capable of performing all operations performed by contemporary general purpose computers (cf. Refs 5, 6,7).

The first optical implementation of symbolic substitution was reported in Ref.6, the arrangement was composed by a recognition unit, a thresholding device and a substitution unit. The recognition unit was based on the superposition of shifted copies of the data plane which

were produced by a beamsplitter. In Ref.10 we presented another implementation of symbolic substitution based on the discrete correlator described in chapter 3.

The discrete correlator described in chapter 3, can be used as recognition unit of a symbolic substitution processor. The optical arrangement for substitution is identical to that for recognition. The complete system for symbolic substitution is thus composed of two 4f systems in series, as shown in Fig.5.6. A NOR gate array is inserted at the output plane of the 4f recognition system to perform the nonlinear thresholding and reestablish binary values. The NOR gate can be implemented using a LCLV. For our experimental demonstrations we used the ASULAB LCLV described in section 4.2.2. The bright pixels at the output of the NOR gate array, which indicate the presence of a search symbol, are split by a grating. The substitution symbol can again be chosen by appropriate filtering in the Fourier plane. Depending on the spatial filter, both, intensity and polarization coded output patterns can be obtained. The simple experimental patterns in Fig.5.6 illustrate the optical processing: a) input mask, as seen at the output, i.e. inverted by the NOR gate; c) output pattern of the recognition unit after spatial filtering for the search symbol b); e) output pattern of the substitution unit after spatial filtering for the substitution symbol d).

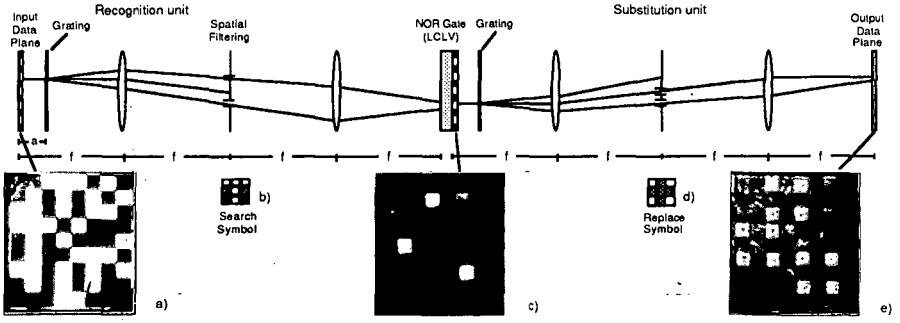


Fig.5.6. Optical arrangement for symbolic substitution and illustrative experimental results.

Attention must be paid to the fact that the output of the second 4f system does not yield a plane wave, but that it contains several spatial frequencies due to the diffraction grating. Thus it may be necessary to introduce another light valve into the system, if a further optical processor should be cascaded or the output be reintroduced into the same symbolic substitution processor for iterating applications.

Using two crossed gratings with several diffraction orders, nearly any arbitrary complex symbol can be recognized and substituted. In Figure 5.7, the recognition of a 3x3 symbols is illustrated. Figure 5.7a) shows the original input pattern (photographic mask), as seen at the output of the NOR gate array (LCLV). The 12 x 14 mm^2 mask contains about 49x56 square pixels of 0.3 mm length. Figures 5.7b) and c) show the recognition of "T" and cross ("x") symbols, respectively. In Figure 5.7d), a result of a substitution is displayed: the crosses are substituted by two points.

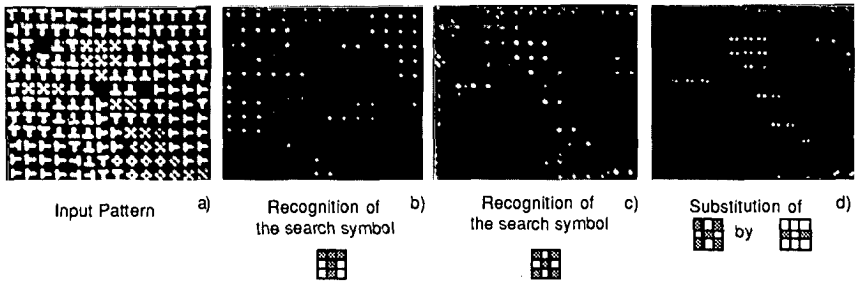


Fig.5.7. Symbolic substitution of intensity coded 3x3 search symbols.

The symbolic substitution processor described above is well suited for multi-channel architectures. We have realized an optical processor with eight channels which performs binary addition and subtraction in parallel [11]. Another application of this symbolic substitution processor is presented in the next section.

5.3. Image processing through mathematical morphology operations

5.3.1. General concepts

Mathematical morphology is a method for image analysis which has found increased importance in recent years. Using electronic image processing systems, it has been successfully applied, for example, for edge detection, noise suppression, or smoothing. The theoretical foundations of mathematical morphology were developed essentially by Serra and Matheron, [12], and are well established. Binary image algebra is a special case of morphological transformations, where input and output are binary discrete images represented by 2-D sets in plane $\mathbf{Z}^2$. Let X denote a set of image points (pixels having value "1"), X^c its complement and B a structuring element. The basic morphological transformations are mathematically described by:

Minkowsky addition: $X \oplus B = \{x+b : x \in X, b \in B\}$ (5.1.a)

Minkowsky subtraction: $X \ominus B = (X^c \oplus B)^c$ (5.1.b)

In Eq. (5.1.a) x and b are image points of the set X and B, respectively, described by their coordinates in the Z^2 plane. The brackets {...} denote the set of all points which fulfill the condition inside. If in Eq. (5.1), instead of B, its point-symmetric element $B^S=\{-b: b\in B\}$ is used, the well-known morphological transformations dilation and erosion, respectively, are obtained. For symmetrical structuring elements $B=B^S$, the dilation is identical to the Minkowsky addition, and the erosion is identical to the Minkowsky subtraction.

Figure 5.8 shows an example of dilation and erosion of a 2-D set X by a structuring element B. Since B is symmetric, these two operations are equal to Minkowsky addition and subtraction respectively. The black pixels in the figure represent the elements of the set, the white pixels are not element of the set.

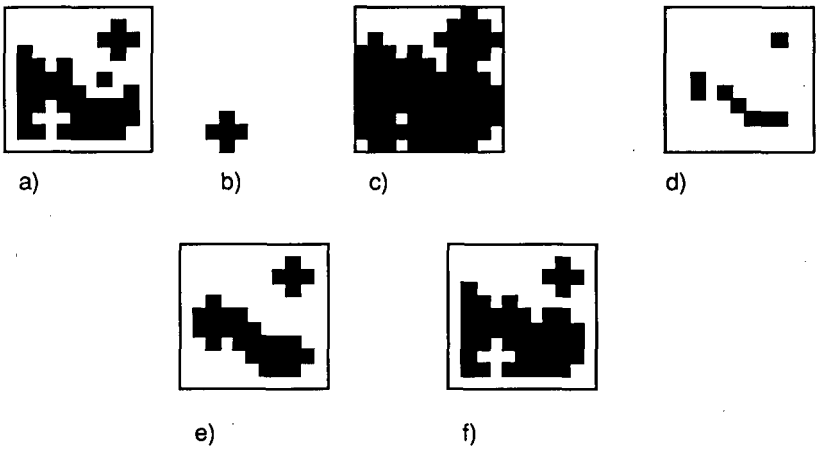


Fig.5.8. a) Original set; b) structuring element; c) dilated set; d) eroded set; e) opened set; f) closed set.

The dilation increases the number of elements of the set X and fills in the inner holes. The erosion shrinks the entire set, and the holes inside the set increase in size.

Opening and closing are the operations that follow directly from erosion and dilation, they are mathematically described by:

$$\text{Opening } X_B = (X \ominus B^S) \oplus B \qquad (5.2.a)$$

$$\text{Closing } X^B = (X \oplus B^S) \ominus B \qquad (5.2.b)$$

The opening operation is an erosion followed by a Minkowsky addition, the points of the open set coincide with the points of B which are completely contained in X. The closing operation is a dilation followed by a Minkowsky subtraction. Figure 5.8e) and f) show the opening and the closing of the set of Fig.5.8a) with the structuring element of Fig.5.8b).

The algebraic properties of the operations described above can be found in Ref.12. Our purpose is only to describe an application of the symbolic substitution processor and we will not go further into the mathematical description.

There are other operations that can be obtained by combination of the four basic operations, these include edge detection (EXCLUSIVE-OR between an eroded and a dilated image), noise suppression (open-closing), skeletonization, median filtering.

5.3.2. Mathematical morphology using symbolic substitution

In a recent paper [13], Goodman and Rhodes show how optical symbolic substitution can be used to perform the morphological operations. In particular, the four mathematical operations described above can be realized as one- or two-step application of symbolic substitution. The symbolic recognition performs directly a Minkowsky subtraction or erosion, using for the recognition symbol the structuring element B^S or B, respectively. The Minkowsky addition and the dilation is carried out by substitution of each "1" pixel by the symbol B or B^S, respectively. The opening of an input pattern can be accomplished in one recognition /substitution cycle, i.e. an erosion (recognition of B^S) followed by a Minkowsky addition (substitution of each pixel of the eroded image with the structuring element B). The closing operation is implemented in two steps. In the first step we recognize all the "1" pixels in the input pattern and we replace them with a structuring element B^S (dilation), in the second step we recognize the structuring element B in the dilated pattern and we replace it with a single "1" element (Minkowsky subtraction).

Optical implementation of the basic operations of mathematical morphology were described by Brenner [30] and Casasent [31] but it is to emphasize that no experimental results were reported in these papers.

The symbolic substitution processor, which was described in section 5.2, is well suited to implementing the basic binary image algebra operations, especially for complex structuring elements. Some experiments were performed and we show now some results. (The results presented are essentially described in Ref.24).

5.3.3. Experimental results

We begin with an example illustrating erosion and dilation. In Fig.5.9 the erosion and dilation of a binary discrete image with 30x30 pixels are shown. The structuring element is a 5 pixel cross, shown in the figure. Note that in this case the bright pixels are elements of the set.

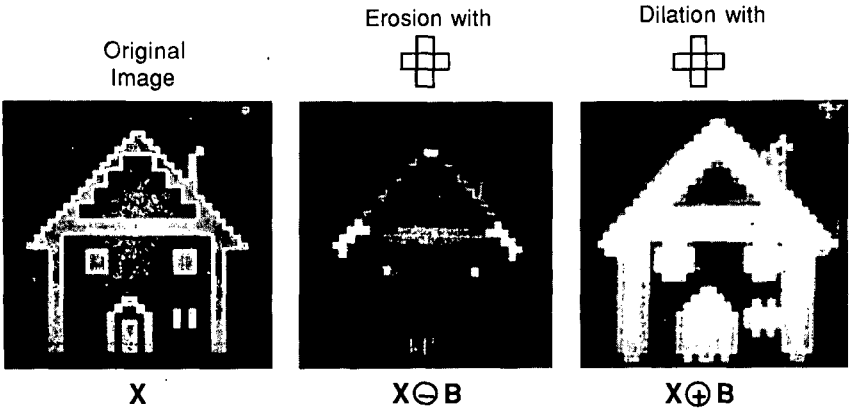


Fig.5.9. The effects of erosion and dilation of the original image of the house by a 5 pixel cross structuring element.

More complex morphological operations, like opening and closing, can be performed using a combination of erosions and dilations. Figure 5.10.b) shows the experimental result of an opening of the original image [Fig.5.10.a)] (60x80 pixel binary image). The structuring element B was composed in this case by two horizontally adjacent pixels. Notice that the isolated white vertical lines (composed by "1" pixels), which are present in the original image, are suppressed in the opened image.

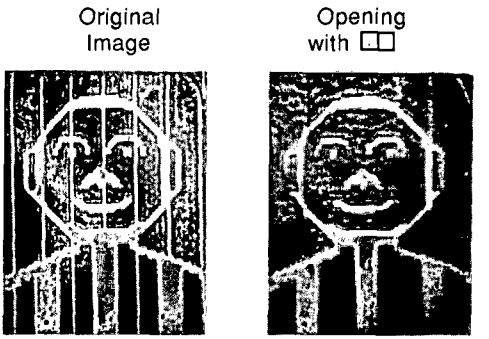


Fig.5.10. The effects of opening of the original image by a structuring element composed by two horizontal pixels.

Figure 5.11 shows an example of noise suppression. Figure 5.11a) shows the binary input pattern composed 50x64 pixels. Figure 5.11b) shows the experimental result of an opening by the structuring element B composed by two horizontally adjacent pixels. A subsequent closing operation would also remove the dark spots in the original image. For ease of experimental realization, the closing operation was replaced by a single dilation with the structuring element B upon the opened image. One obtains then the complete transformation

$$X_B \oplus B^S = ((X \ominus B^S) \oplus B) \oplus B^S \equiv (X \ominus B^S) \oplus (B \oplus B^S), \tag{5.3}$$

where the identity in the above equation follows from the algebraic properties of mathematical morphology [12]. The whole operation of Eq.(5.9) could again be performed in one recognition / substitution cycle. The input image was thus first eroded with the structuring element B (recognition), and then dilated with the structuring element $(B \oplus B^S)$ (substitution).

a) b) c)

Fig. 5.11. The opening b) and subsequent dilation c) of the original image a) results in removal of the noise.

The experimental result is shown in Fig.5.11c). Notice, that the isolated black and white pixels, which are present in the original image, are suppressed; this is an example of noise removal. The imperfections of the black spot suppression, exhibited in the experimental results, are probably caused by carrier migration in the photoconductor of the liquid crystal light valve and result in the expansion of dark image points (bright points at the input side of the inverting LCLV).

Figure 5.12 shows an experimental result of edge detection. A dual rail system performs in parallel the erosion and the complement of the dilation of the original image, Fig.5.12a) (50x80

pixels). The structuring element was a square of 9 pixels. The resulting image [Fig.5.12b)] was obtained by superimposing the dilated and the eroded images on a screen.

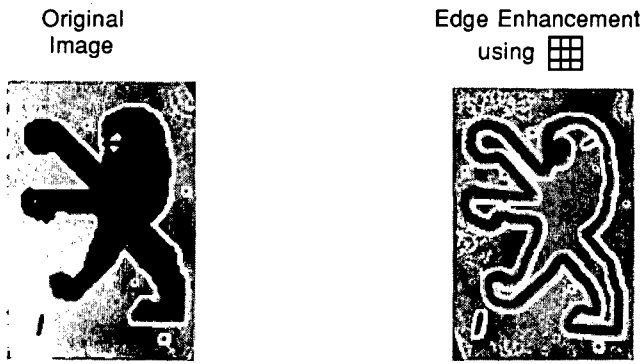


Fig.5.12. The edge enhanced result b) of the original image a) is achieved by the superposition of the dilated and eroded images.

5.4. Comparison with the existing systems

We will at first compare the correlator described in this work with the best known incoherent correlator, i.e. the shadow-casting correlator [32]. This correlator can process continuous or discrete functions. Here we will consider only the case in which discrete functions are processed.

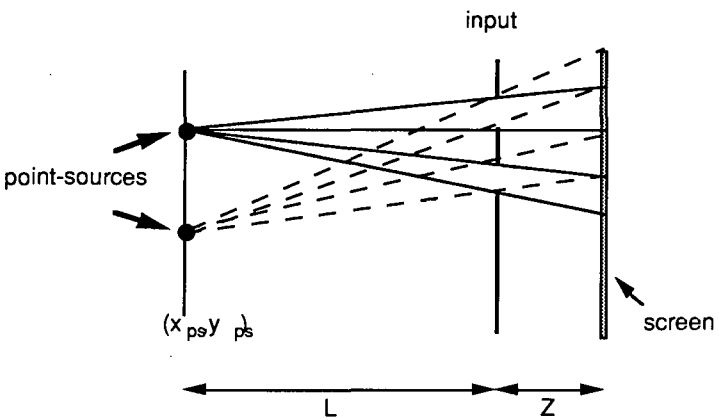


Fig.5.13. Geometrical configuration of the shadow-casting system.

Figure 5.13 shows a schematic diagram of the lensless shadow-casting system. The light beams radiating from the point sources in plane (x_{ps},y_{ps}), illuminate the spatially coded transparency (input data), and project multiple copies of the input onto the screen (in Fig.3.13 only two point–sources are shown. The correlation is performed by incoherent superimposition of shifted copies like in the correlator described in this work. The shadow-casting correlator is based on geometrical optics. An important limitations of this system is that diffraction effects may be significant when images composed by small pixels are to be processed. In the system described in this work, the quality of the images to be superposed is given by the resolution of the 4f filtering set–up. In Ref.33 are reported some considerations about the limitations of the shadow–casting system. In particular it is shown that the distance Z between the mask and the screen must satisfy the relation $Z<<\Delta^2/\lambda$, where Δ is the size of the pixels in the transparency. We choose e.g. $Z=\Delta^2/(10\lambda)$, and Δ= 50 μm thus Z=0.5 mm. This short distance limits the number of copies that can be superposed. This correlator is programmable, in fact the point-sources could be for example LEDs and the input transparency could be a LCD. It can process intensity and polarization encoded pattern, but it cannot process theta encoded patterns.

The shadow-casting correlator can be used as a recognition unit in a symbolic substitution processor [34].

The first optical implementation of symbolic substitution was reported in Ref. 6, it can recognize and substitute only symbols composed by two pixels. Multichannel processing is not possible and the programming of the search symbol is performed mechanically, by adjusting a mirror. Our system has the advantage to recognize symbols composed by more pixels (3x3, 5x5,..), multichannel processing is possible and the programming of the search symbol can be accomplished by using a SLM.

In Ref. 35 is described another symbolic substitution processor using two coherent correlators and a thresholding element. Some simple experimental results were presented. This system can perform multichannel processing. The search symbol is not programmable and to change the search symbol the hologram in Fourier plane has to be changed. This system is coherent, it needs optical components of high quality, and it is very sensitive to misalignment of the filter. Compared with the system described in this work, it is more expensive and more difficult to build.

6. CONCLUSIONS

An optical correlator based on diffraction gratings has been presented. The processor performs basically an incoherent, discrete correlation between an input data pattern and a predefined search symbol. Programmability of the search symbol is obtained by selecting the appropriate diffraction orders in the Fourier plane (spatial filtering). The described correlator can process input patterns of different encoding (intensity, polarization or spatial frequency encoding). Many of the existing SLMs (coherent or incoherent, optically or electrically addressable) can be used as input or filtering devices in our correlator, and real time programming of the input pattern and the search symbol is thus achieved. The possibility to use LCDs as input patterns and spatial filters has been experimentally demonstrated.

We demonstrated (theoretically and experimentally), that the correlator can work either with completely spatial coherent light or with partially spatial coherent light. In fact, the use of partially coherent illumination turns out to be advantageous to obtain images of better quality, at the expense, however, of a certain loss of SBWP.

The ability of the system to recognize multiple programmable symbols in parallel establishes its application in many areas. In particular, we have shown that two discrete correlators, separated by a nonlinear threshold element, can be used to build a symbolic substitution processor. Applications in binary optical data processing have been illustrated with experimental results in the areas of correlation, arithmetics and image processing.

During all this work we have emphasized the fact that we need a certain degree of spatial coherence because this allows the filtering (suppression or change of the polarization directions) of the diffraction orders in the Fourier plane. We can now imagine to have a phase grating which produces only the diffraction orders that we need for the recognition of a certain symbol. In this case, the function to correlate with the input pattern is defined by the diffraction grating itself and we do not need anymore the filtering in the Fourier plane. Such a system has the great advantage that it can work with spatially incoherent light, in fact we do not need a spatial separation of the orders in the Fourier plane, real images coming for example from a monochromatic TV screen could then be processed without increasing the degree of coherence (an increase of the degree of coherence involves a dramatical loss of power). We can now imagine a programmable phase only grating producing the diffraction orders which we need, this device will allow real time processing. Actually it does not exist such a device but the constant progress in technology of SLMs could produce in the future a programmable high definition phase mask. This would be the ideal device which could dramatically increase the performance of the discrete correlator presented in this work.

ACKNOWLEDGMENTS

I wish to thank Prof. R. Dändliker director of this thesis. His constant encouragement and his great knowledge have been of invaluable help to me.

Special thanks go to Dr. R. Thalmann for the fruitful collaboration during the last years and for his helpful advice along the way.

Furthermore, I express my gratitude to Prof. P. Chavel, Prof. F. Pellandini, Dr. N. Collings and Mr. K. J. Weible for many helpful suggestions.

Finally, I would like to thank all my colleagues, the technical assistants and the secretarial staff of the Institut de Microtechnique, who made it possible to realize this work in a pleasant atmosphere.

This research work was supported by the Swiss National Science Foundation, project number 2.255–0.86.

REFERENCES

1. J. W. Goodman,"Introduction to Fourier Optics", (McGraw-Hill, New York 1968).

2. F. T. S. Yu, "Optical information Processing", (John Wiley & Sons, New York 1983).

3. A. Vander Lugt,"Coherent Optical Processing", Proc. IEEE, **62**, 1300-1319 (1974).

4. M. Born, E. Wolf "Principles of Optics ", (Pergamon Press, Oxford, 1980).

5. A. Huang, "Parallel Algorithms for optical Digital Computers", in Technical Digest, IEEE Tenth International Optical Computing Conference, (1983).

6. K.-H. Brenner, A. Huang, and N. Streibl, "Digital optical computing with symbolic substitution", Appl. Opt., **25**, 3054-3060 (1986).

7. M. J. Murdocca, "Digital optical computing with one rule cellular automata", Appl. Opt., **26**, 682-688 (1987).

8. K.-H. Brenner, "Programmable optical processor based on symbolic substitution", Appl. Opt., **27**, 1687-1691 (1988).

9. K.-H. Brenner, "New implementation of symbolic substitution logic", Appl. Opt., **25**, 3061-3064 (1986).

10. R. Thalmann, G. Pedrini, B. Acklin, and R. Dändliker, "Optical symbolic substitution using diffraction gratings", ICO topical meeting on Optical Computing 88, Proc. Soc. Photo-Opt. Instrum. Eng. Vol. **963**, 635-641 (1989).

11. G. Pedrini, R. Thalmann, and K.J. Weible, "Parallel addition and subtraction in one computing cycle using optical symbolic substitution", Topical Meeting on Optical Computing, Salt Lake City, UT, USA, 1989.

12. J. Serra, "Image Analysis and Mathematical Morphology", (Academic Press, New York, 1982).

13. S.D. Goodman and W.T. Rhodes, "Symbolic substitution applications to image processing", Appl. Opt., **27**, 1708-1714 (1988).

14. M. J. Lighthill, " Einführung in die Theorie der Fourier Analysis und der verallgemeinerten Functionen", (B.I. Wissenschaft Verlag, Mannheim, 1966).

15 J. D. Armitage and A. W. Lohmann, "Theta modulation in optics", Appl. Opt., **4**, 399-406 (1965).

16. H. Dammann and K. Görtler, "High efficiency in-line multiple imaging by means of multiple phase holograms", Opt. Commun., **3**, 312-315 (1971).

17. U. Killat, G. Rabe, W. Rave, "Binary phase gratings for star couplers with high splitting ratio", Fiber and Integrated Optics, **4**, 159–167 (1982).

18. M.T. Gale at Paul-Scherrer-Institute PSI c/o Laboratories RCA Ltd, Zürich, Switzerland.

20. J. Weigelt, "Space-bandwidth product and crosstalk of spatial filtering methods for performing binary logic optically", Opt. Eng., **27**, 883-892 (1988).

21. 32x84 element graphic LCD, Epson EG-Y84320AT, 1/16 multiplexed driven 0.75x0.75 mm^2 square pixels with 0.1 mm spacing.

22. 5x7 element LCD, Epson LD-H9005AZ, static drive.

23. B. Acklin, R. Thalmann, " Liquid cristal displays as transmission spatial light modulators", IMT report N^0 245 EC 09/88.

24. R. Thalmann, G. Pedrini, K. J. Weible, "Optical symbolic substitution using diffraction gratings", Appl. Opt., **29**, 2126-2134 (1990).

25. R. Dändliker, "Optique cohérente", lecture at the University of Neuchâtel.

26. H. Bartelt, A.W. Lohmann, E. E. Sicre,"Optical logical processing in parallel with theta modulation", J. Opt. Soc. Am., **9**, 944-951 (1984).

27 High Performance 1" CRT Assembly , technical data WX-34250, Westinghouse Electric Corporation.

28 CCD camera, technical data Aqua HR 480

29 A. R. Tanguay, C. S. Wu, P. Chavel, T. C. Strand, A. A. Sawchuk, "Physical characterisation of the variable grating mode liquid crystal device", Opt. Eng., **22**, 687-694 (1983).

30 K.-H. Brenner, G. Stucke, "Architectures for digital optical image processing using morphological filters", ICO topical meeting on Optical Computing 88, Proc. Soc. Photo-Opt. Instrum. Eng. Vol. **963**, 657-662 (1989).

31 D. Casasent, E. Botha, "Optical symbolic substitution for morphological transformations", Appl. Opt., **27**, 3806-3810 (1988).

32 M. A. Monahan, K. Bromley, and R. P.Bocker, "Incoherent optical correlators", Proc. IEEE, **65**, 121–129 (1977).

33 J. Tanida, Y. Ichioka, "Optical logic array processor using shadowgrams", J. Opt. Soc. Am., **73**, 800–809 (1983).

34 Wei Xue, Li–Xue Chen, Chun–Fei Li, Qiang–Sheng Hu, "Symbolic substitution using shadow–casting", ICO topical meeting on Optical Computing 88, Proc. Soc. Photo-Opt. Instrum. Eng. Vol. **963**, 653-656 (1989).

35 D. Casasent, E. Botha, " A unified optical symbolic processor ", ICO topical meeting on Optical Computing 88, Proc. Soc. Photo-Opt. Instrum. Eng. Vol. **963**, 720-728 (1989).