MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence
Memo. No. 160

May 1968

Focusing

Berthold Horn

This memo describes a method of automatically focusing the new vidisector (TVC). The same method can be used for distance measuring. Included are instructions describing the use of a special LISP and the required LISP-functions. The use of the vidisectors, as well as estimates of their physical characteristics is also included, since a collection of such data has not previously been available.

CONTENTS

Introduction: This memo describes the application of a fourier transform method to the focusing problem. It is assumed that the reader has some familiarity with various modes of fourier transform. In particular, use will be made of certain similarities between the transform of a function of a continuous variable and that of a function of a discrete variable, since the discrete transforms usually have much more complicated analytic expressions, yet have much the same behaviour. It will be seen that the function $2*J_1(x)/x$ plays the same role in the two-dimensional transform as $\sin(x)/x$ does in the one-dimensional transform. It may be found that some of design has been spelled out in too much detail - if so forgive those of us who have forgotten their optics and would like it spelled out. The ordering of topics may also be found to be unusual - it is an attempt to write it such that forward references were not required. This work is part of the work done towards a Masters Thesis on the application of fourier transforms to image processing, and comments would be appreciated. Finally, no guarantee is given on the accuracy of the vidisector constants reported herein - when more accurate values become available they will be reported in the log-book(under the monitor).

Purpose: Focusing is one aspect of camera operation that is not usually automated (unlike exposure time setting). It was thus of interest to show that it is possible to focus an optical device automatically with the same degree of accuracy achieved by a human watching the picture on a ground-glass screen (or in our case the monitor). The method described shows one of a number of applications of the fourier transform in image processing. Another goal is distance measuring without utilizing a stereo effect (thus avoiding the stereo match problem).

Variations on the fourier theme: The standard fourier transform pair of a function of a continuous variable can be written as:

$$g(\omega) = \underset{R\to\infty}{\mathrm{l.i.m}} \frac{1}{\sqrt{2\pi}} \int_{-R}^{R} f(x)\, e^{ix\omega}\, dx$$

$$f(x) = \underset{R\to\infty}{\mathrm{l.i.m}} \frac{1}{\sqrt{2\pi}} \int_{-R}^{R} g(\omega)\, e^{-ix\omega}\, d\omega$$

This can be extended to functions of more than one dimension, by allowing the transform to be a function of as many frequency-variables as the original function is of space-variables.

$$g(U) = \underset{R\to\infty}{\mathrm{l.i.m}} \frac{1}{(2\pi)^{n/2}} \int_R f(X)\, e^{i(X\cdot U)}\, dV_X$$

$$f(X) = \underset{R\to\infty}{\mathrm{l.i.m}} \frac{1}{(2\pi)^{n/2}} \int_R g(U)\, e^{-i(X\cdot U)}\, dV_U$$

X - space vector, U - frequency vector

R - volume of integration

X.U - dot product of X & U

When the function depends on discrete variables rather than continuous ones we may use the discrete fourier transform:

$$g(u\omega) = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} f(xT)\, W^{xu} \qquad u = 0,1, \dots N-1$$

$$f(xT) = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} g(u\omega)\, W^{-xu} \qquad x = 0,1, \dots N-1$$

$$W = e^{-\frac{2\pi i}{N}} \;;\; \omega T = \frac{2\pi}{N}\;;$$

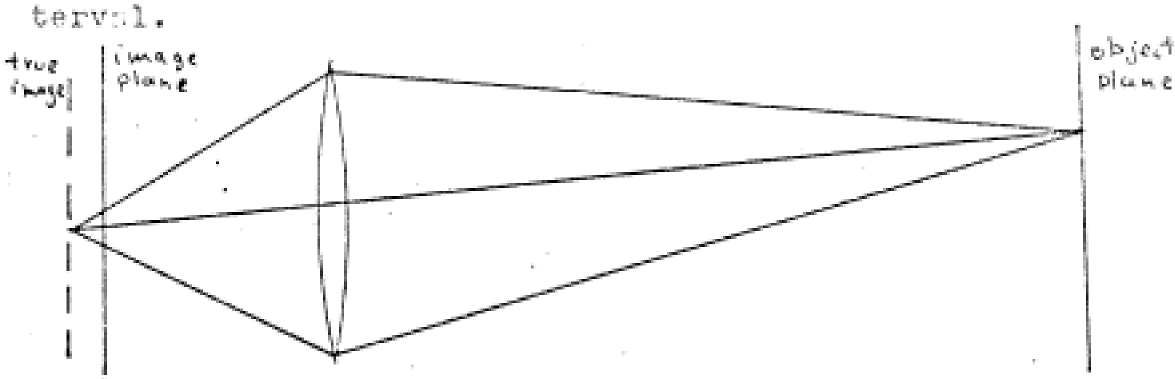
T - spacing of points ; N - number of points

This again may be simply extended to more than one dimensions. In certain cases it is possible to calculate discrete fourier transforms (DFT's) very rapidly by the method of fast fourier transforms (FFT) described in an appendix.

Note that the DFT assumes a certain periodicity in the function:

$$g((N+u)\omega) = g(u\omega)$$
$$f((N+x)T) = f(xT)$$
$$g((-u\omega)) = g((N-u)\omega)$$
$$f((-xT)) = f((N-x)T)$$

ie. both the function and its transform are periodic functions of discrete variables. Another interpretation useful in certain cases is that one of the two represents samples of a periodic frequency-limited function, the other one then is discrete but nonzero only over an interval.



The effect of defocusing an image: Consider an image of a plane perpendicular to the optical axis. Each point of the object generates a point on the image with intensity proportional to the source point. When we insert a plane perpendicular to the optical axis between the image plane and the lens we cut each cone of light (corresponding to an image-point) in a circle. The intensity in each such circle is again proportional to that of the corresponding source point, uniform across the circle and decreasing with the radius of the circle such that the total light is constant. The new image is thus the convolution of the in-focus image and a little pillbox (whose height is inversly proportional to the square of its radius). This effect is easier to describe in the frequency domain, since convolution in the space domain corresponds to multiplication in the frequency domain. In other words the frequency spectrum of the defocused image is that of the in-focus image multiplied by the frequency spectrum of the pillbox. This transform is found in an appendix to be $2 * J_1(R\rho)/(R\rho)$ where ρ is the frequency (in radians per unit distance) in polar coordinates and R is the radius of the pillbox. (J_1 is the Bessel-function of order 1). Also shown in the appendix is a cross-section through this function (which is symmetrical about the origin).

It will be seen that this function contracts as R increases (ie. the first zero crossing approaches the origin) and the effect of defocusing can thus be seen to be a reduction of high frequency components in a certain way and such that lower and lower frequencies are affected as we defocus more and more (ie. increasing R). Taking the fourier transform thus gives us a function in which the effect of defocusing is easy to interpret. Various functions of this transform can now be used as functions to be maximised so as to obtain best focus (w.r.t lens position). A description of such a method is not complete however without an analysis of noise, since it is trivial to focus in the absence of noise (for example by maximising the difference between the light-intensity at two adjacent points on the retina corresponding to two points close together on a part of an object which has other than a uniform light intensity distribution). Noise appears in various ways in vidisector images, the most important appearing in an obvious way in the intensity values returned, caused by the statistical fluctuation in the number of photoelectrons generated at the area under consideration. The function of the fourier transform to be used has to be designed as a compromise between ones relatively free from noise and ones with a narrow maximum (in the absence of noise it is trivially possible to narrow the width of a maximum as much as one pleases by raising the function to a high enough power).

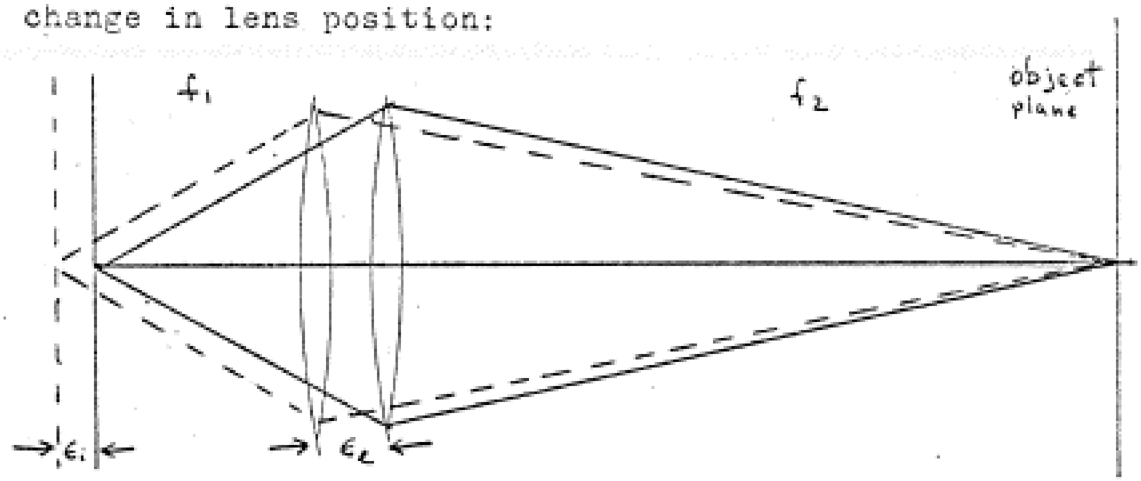
ANALYSIS

Sensitivity: We would like to know what error in distance measurement is incurred by two kinds of error:

a) Limitations on determining correct focus due to a combination of vidisector resolution limits and noise.

b) Error in lens positioning.

We have a fixed object plane and a fixed image plane. The lens is moved about, causing the true image to form somewhere behind or before the image plane. First let us find the change in true image position due to a small change in lens position:



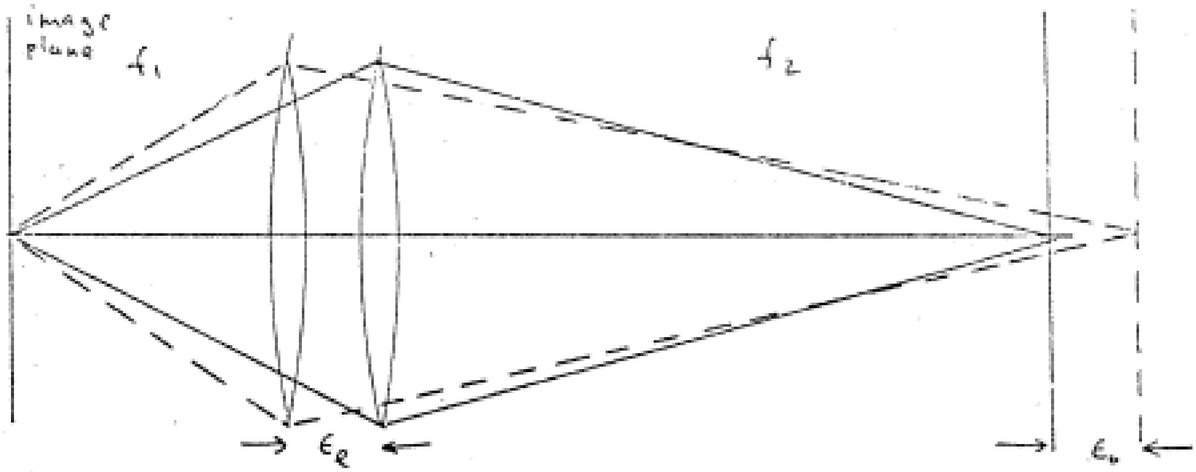
Let $f_2 > f_1$, $d = f_1 + f_2$; $\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f}$ where f is the focal length

$d = f_1 + f_2 = \frac{f_2^2}{f_2 - f}$, $f_2 = \frac{d + \sqrt{d^2 - 4df}}{2} = f\left[\left(\frac{d-2f}{2f}\right) + 1 + \sqrt{\left(\frac{d-2f}{2f}\right)^2 - 1}\,\right]$

Let $x = \left(\frac{d-2f}{2f}\right)$

Differentiating w.r.t d : $\frac{\epsilon_e}{\epsilon_i} \doteq \frac{d f_2}{d d} = \frac{1}{2} \frac{x + \sqrt{x^2 - 1}}{\sqrt{x^2 - 1}}$

Next we would like to know what change in image distance this corresponds to:



$d = f_1 + f_2 = \frac{f_1^2}{f_1 - f}$; $\therefore\ f_1 = \frac{d - \sqrt{d^2 - 4df}}{2} = f\left[\left(\frac{d-2f}{2f}\right)^2 + 1 - \sqrt{\left(\frac{d-2f}{2f}\right)^2 - 1}\right]$

Differentiating w.r.t. d we get: $\frac{\epsilon_e}{\epsilon_o} \doteq \frac{df_1}{dd} = \frac{1}{2}\,\frac{x - \sqrt{x^2-1}}{\sqrt{x^2-1}}$

It also follows that $\frac{\epsilon_o}{\epsilon_i} = \frac{x + \sqrt{x^2-1}}{x - \sqrt{x^2-1}}$

for $x > 1.25$ (ie. $d > 4.5 f$) $\frac{\epsilon_o}{\epsilon_i} \doteq 4x^2 - 2$ (within 5%)

$$4x^2 - 2 = \left(\left(\frac{d}{f} - 2\right)^2 - 2\right)$$

$\left(\frac{d}{f}\right)^2$ is unfortunatly not a good approximation to this unless $\left(\frac{d}{f}\right) > 100$

$\frac{d}{f}$	$\frac{\epsilon_o}{\epsilon_i}$
4.0	1.0
4.5	3.99
5.0	6.88
6.0	13.9
7.0	23.0
8.0	34.0
10.0	62.0
15.0	168.0
20.0	322.0

<u>Focusing function</u>: Before continuing to an analysis of noise effects it will be necessary to describe the function used in more detail. This is done with some hesitancy since it is rather ad hoc, and many other simple-minded procedures would have been satisfactory. The point is that the information is present in the fourier transform in a 'convenient' form and the practical details of the particular function used might detract from this fact. To get some idea of how the number of points used, their spacing, noise and resolution limits effect the accuracy one has to be specific however (and also make some drastic simplifying assumptions in the arithmetic).

For each setting of lens position investigated, intensity values are read for N points spaced uniformly T cms apart along the circumference of a circle (the reason for choosing a circle will be apparent later). We now perform a one dimensional FFT on these points to obtain N frequency components (the reason for using the less powerful one-dimensional approach will also be discussed later). Next we form the power spectrum by adding the square of the imaginary part to the square of the real part at each frequency. One now sums all components starting at some minimum frequency $N_1\omega$ up to the centre frequency $(\frac{N}{2}-1)\omega$. The top half of the spectrum is not used since it is merely a reflection of the lower half (the transform of any real function has this symmetry property).

To obtain some independence of changes in overall level we have to normalise the result. Dividing by a single value (eg. the DC or zero-frequency component) causes relative noise in the result of equal magnitude as the relative error in this single value and is therefore not advisable. Empirically it was found that 'best' results could be obtained by dividing by the sum of all terms except the DC term (zero frequency term).

The program now proceeds to carry out this operation for some range of lens-settings, thus forming a function F of lens-setting which has a noisy peak near optimum focus.

Signal to noise ratio: The main noise contribution to the signal is caused by statistical fluctuation in the number of photoelectrons emitted by the photocathode. The video processor is so designed that the peak-signal to rms-noise ratio increase by a factor of 2 for each increase in the confidence level, by counting 4 times as many electrons.

Confidence level	Signal (Peak) / Noise (RMS)	Photoelectrons
0	8	$\approx 2^7$
1	16	$\approx 2^9$
2	32	$\approx 2^{11}$
3	64	$\approx 2^{13}$

The vidio-processor is so built that this relative error is the same for different levels of illumination, so that the absolute error in a low value will be smaller than that in a high value. It is convenient to assume however that the error in all points is the same, and equal to that in the point with the highest intensity. This quantisation error has a roughly uniform distribution. Adding many terms, as we do in forming the transform, gives us a roughly gaussian distribution. Let f be the pure signal and g a superimposed noise:

let $\frac{1}{N}\sum_{\nu=0}^{N-1} f^2(xT) = \sigma_s$ and $\frac{1}{N}\sum_{\nu=0}^{N-1} g^2(xT) = \sigma$

s_i (signal to noise power) $= \sigma_s / \sigma$

let $\bar{f} = DFT(f)$ and $\bar{g} = DFT(g)$

then $DFT(f+g) = DFT(f) + DFT(g)$

Each $g(u\omega)$ is the sum of N independent vectors each with variance σ. N/P of these point in one of p directions, (where p depends on the common factors of u and N). The sums in each of these directions, have mean zero and a variance of $p\sigma$. Decomposing these variations along the real and imaginary axis we have:

Real contributions	$\sum_{i=0}^{p-1} \frac{N}{p} \cos^2 \frac{i\pi}{p}$
Imaginary contribution	$\sum_{i=0}^{p-1} \frac{N}{p} \sin^2 \frac{i\pi}{p}$

The resultant vectors thus have zero mean and variance $N\sigma$. Since we divided by $\sqrt{N}$ in forming the DFT we have to divide by N to get the variance in each of the frequency components. The transform of such a noise signal thus has the same characteristics as the original noise signal (ie. zero mean and variance σ).

We next have to consider the power spectrum:

$|\bar{f}+\bar{g}|^2 = |\bar{f}|^2 + |\bar{g}|^2 + 2|\bar{f}\bar{g}_c|$

where $\bar{g}_c$ is the complex conjugate of $\bar{g}$

It can be assumed that the noise power $|\bar{g}|^2$ is small compared to the signal power $|\bar{f}|^2$ and the cross-term. This cross-term thus is the main source of noise in the power spectrum, it has variance $|\bar{f}|^2 \times \sigma \times \frac{2}{2}$. In forming the sum from N_1 to $(\frac{N}{2}-1)$ we find the ratio of <u>power</u> of the value so formed to its noise is:

$$S_f = \frac{\sum_{\upsilon=N_1}^{\frac{N}{2}-1} |\bar{f}|^4}{\sum_{\upsilon=N_1}^{\frac{N}{2}-1} |\bar{f}|^2} \times \frac{1}{\sigma} = \frac{\sum_{\upsilon=N_1}^{\frac{N}{2}-1} |\bar{f}|^4}{\left(\sum_{\upsilon=N_1}^{N/2-1} |\bar{f}|^2\right)^2} \times \frac{\sum_{\upsilon=N_1}^{N/2-1} |\bar{f}|^2}{\sum_{\upsilon=0}^{N/2-1} |\bar{f}|^2} \times \left(\frac{N}{2}-1\right) \times \frac{\sum_{\upsilon=0}^{N/2-1} |\bar{f}|^2}{\left(\frac{N}{2}-1\right)\sigma}$$

$$S_f = F_1 \times F_2 \times \left(\frac{N}{2}-1\right) \times S_i$$

where F_1, F_2 are the ratios in the above expansion.

<u>Note</u>: 1) The result is only approximate and merely presented to give a clue as to what factors make for a good signal to noise ratio; the variation with N is particularly important.

2) The form-factor F_1 depends on the image and is approximately one for an image having a more or less flat spectrum, and is inversely proportional to N_1 for one having a hyperbolic $(1/\upsilon)$ spectrum. Most images lie somewhere between these two extremes; an image with a lot of texture coming closer to the first form, one with a single edge coming closer to the second. F_1 describes how 'wobbly' $|\bar{f}|$ is above the minimum frequency $N_1\omega$. The other form-factor F_2 describes how much of the signal power is in the high-frequensy part (above $N_1\omega$) of the spectrum. It has a similar variation as F_1, but is not as sensitive to small changes in an image.

3) It is clear that one should try for a high signal to noise ratio in the image and besides using high values of the confidence level one can help to achieve this by averaging light intensities at each point over more than one reading. This was not found to be necessary (unlike such local operations as homogeneity determination where this is necessary to avoid missing small changes in level between regions differing only slightly in brightness).

4) The last term in the product is in fact S_i since the sum of squares in the transform is equal to the sum of squares in the original function and since the transform is symmetric.

Width of the maximum in the focusing function: Let the width be the distance between two points on the curve at which the function reaches $1/\sqrt{2}$ of its maximum value. If $f(xT)$ is the intensity function when the true image falls on the image-plane, $g(xT,R)$ the Bessel-defocusing function with R the radius of the circles corresponding to each in-focus image point. We have to find two values of R s.t:

$$\sum_{\nu=N_1}^{N/2-1} \bar{f}(xT) \times g(xT,R) \Big/ \sum_{\nu=N_1}^{N/2-1} \bar{f}(xT) = 1/\sqrt{2}$$

Since its such a pain to do arithmetic on the Incompatable Time Sharing System (ITS) and since we only need approximations, we only have estimates for R:

$$\frac{3.832}{R} \cdot \frac{1}{\omega} \doteq N_1 + \frac{N}{2} \qquad \text{for a 'flat' spectrum}$$

$$\frac{3.832}{R} \cdot \frac{1}{\omega} \doteq 1.3\, N_1 + .3\, \frac{N}{2} \qquad \text{hyperbolic spectrum}$$

For N_1 large (i.e. near $\frac{N}{2}$): $R \doteq \frac{3.832}{\omega} \times \frac{1}{N} = 1.22 \left(\frac{T}{2}\right)$

In practice however we keep N_1 small so as to give us high form-factors in the signal to noise expression and we do not have 'flat' spectra. Emperically it is found that one can achieve a width 1 to 2 times the above ideal case. It is clear that the choice of N_1 is one place where one can trade off signal to noise ratio versus width of the maximum. Clearly also the width depends on the relative high-frequency content of the intensity function: it is easier to focus on a line (two edges) than on a single edge and it is very easy to focus on say newsprint or coarsely textured material. It would seem at first sight that one could improve matters without limit by decreasing T. Intuitively however we know that one cannot expect to get much improvement once T is smaller than the resolution limit of the instrument. So we need next to concern ourselves with this kind of limitation in the vidisector.

<u>Distortions</u>: All optical instruments have certain limitations (if only to preserve our belief that one cannot observe nature exactly - whatever that means). Some of the distortions are global in character (such as pin - cushion distortion and spherical distortion) and do not concern us here. Other distortions are more easily described in terms of local effects (eg. diffraction). These distortions act in a similar way to defocusing in that they take each point of the image and 'smudge' it over its neighbours. The smudged image then is the true image convolved with some function which is large only in a small area near the origin. The shape of this little 'hill' depends on the distortion under discussion, for diffraction it is $2 * J_1(Rx)/(Rx)$. The scale factor R is $(\pi/\lambda) * (a/f)$ (λ- wavelength of light, f-focal length of lens, a - diameter of lens). By the well known duality of transforms we can immediatly conclude that the effect of diffraction is to multiply the image spectrum by a pillbox of radius $(1/\lambda)(a/f)$. Diffraction thus has the effect of low-pass filtering the signal. The maximum frequency passed with $\lambda = 5000 \mathring{A}$ and a numerical aperture of 4 is 2500 cycles/cm. The well known phenomena of resolution limits in telescopes and microscopes are due to exactly this.

A larger effect in our case is due to the vidisector. It operates by allowing photoelectrons from the photocathode to fall on a pinhole (approximate diameter 0.005 cm).

Those emerging on the other side are counted. The magnetic field in the tube selects the particular circular area of the photocathode from which these photoelectrons come. The effect is thus seen to be identical to that of defocusing. In practice however various other effects set in, causing electrons from the outside area to be counted and the loss of some of the electrons from inside this area. In other words we have 'smudged' the image not by a pillbox but a smoother, larger 'hill'. The size and shape of this hill depend critically on the adjustment of the vidisector. To get some idea of the effect this new 'smudging' has, we will assume that the 'hill' has a gaussian shape, ie.:

$$f(r,\theta) = k.\, e^{-\frac{1}{2}\left(\frac{r}{R}\right)^2}$$

$$f(x,y) = k.\, e^{-\frac{1}{2}\left(\frac{x}{R}\right)^2} \times e^{-\frac{1}{2}\left(\frac{y}{R}\right)^2}$$

transforming we get:

$$g(u,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{i(ux+vy)}\, dx\, dy$$

$$= \frac{k}{2\pi}\left[\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x}{R}\right)^2 + iux}\, dx\right]\left[\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y}{R}\right)^2 + ivy}\, dy\right]$$

$$= \frac{k}{2\pi}\left[k_1\, e^{-\frac{1}{2}(uR)^2}\right]\left[k_1\, e^{-\frac{1}{2}(vR)^2}\right]$$

$$= c\, e^{-\frac{1}{2}(R\rho)^2}$$

The radius of the gaussian 'blob' before transformation, (out to where it has $1/\sqrt{2}$ of its maximum value) is $.833\,R$ the radius of the gaussian 'blob' in the frequency domain is $.833/R$. Experimentally it is found that the vidisector had a halfpower frequency f_o of 26 cycles/cm (when adjusted - even slightly out of adjustment the cut-off frequency drops considerably). This corresponds to a 'blob' width of 0.009 cm, which indicates that points have to be at least 0.015 cm apart to be reasonably independent.

A knowledge of f_0 allows us to choose reasonable values for T. The maximum frequency appearing in the transform is $\frac{\omega}{2\pi} \times \frac{N}{2} = \frac{1}{2T}$. It is apparent that there is no point in making T so small that $\frac{1}{2T} \gg f_0$ (or else many of the high-frequency terms in the transform will always be near zero). Empirically one finds $1/2T = 2f_0$ satisfactory. ($T = 1/4f_0 = .01$ cm) A practically achievable width for the maximum in the focusing curve is thus $\doteq 1.22 \times T$ (.012 cm) (ie. .01 cms). To find the corresponding allowed variation in the image plane position we have to multiply by the numerical arperture r. Using our sensitivity result we finally obtain:

$$d_r \text{ (distance range)} = 1.22 \times \frac{1}{4f_0} \times \left[\left(\left(\frac{d}{f}\right)-2\right)^2 - 2\right]$$
$$= .012 \quad \left(\left(\frac{d}{f}-2\right)^2-2\right)$$

eg. $r = 4$, $(d/f) = 6$, $d_r \doteq .6$ cm

<u>Servo inaccuracy</u>: The main obstacle preventing one from achieving such high accuracy is the servo controlling the lens position. This servo operates in units of about 0.005 cms. The error in positioning is made up of three components:

a) Systematic error causing the servo to settle 4 units higher than requested.

b) Backlash of about 2 units.

c) Unpredictable variation of about 4 units in both directions. This is actually a time variation;

ie. the servo 'oscillates' by this amount about the steady position at a very low frequency (approximately 0.4 cycles per second)

In addition the lens holder can move the equivalent of 2 units w.r.t. the servo positioning mechanism by tilting. Errors a) and b) can be accounted for, leaving us with a total error of about .03 cm in either direction. (When badly adjusted the servo mechanism can be much worse than this, as much as 20 units either way where observed when the coupling between motor and cam was too stiff or too loose) This error now has to be multiplied by one of our sensitivity terms:

$$\frac{2\sqrt{x^2-1}}{x-\sqrt{x^2-1}}$$

which for $x > 1.5$ is also about $4x^2 - 2$

d_s (distance range) = $.06 \times ((\frac{d}{4} - 2)^2 - 2)$

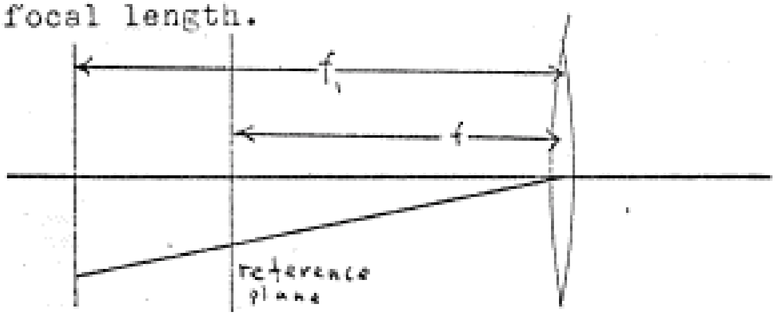
The servo error thus is dominant unless the numerical arperture is more than about 6. (the cross-over point of which of the two errors is more important depends very much on the adjustment of the two systems).

Practical Foibles: The motion of the lens also causes two side effects:

a) Change in magnification.

b) Change in numerical aperture (ratio of (lens to image plane distance) to diameter of iris)

The change in magnification can be taken care of by refering all coordinates to a standard system fixed w.r.t the lens. In the program the distance was arbitrarily chosen equal to the focal length.



The transformation of coordinates simply multiplies distances from the optical axis by f_1/f. This introduces two new effects:

a) Some points of the image may move outside the useful image area of the vidisector as the lens is moved further away.

b) The seperation T, of points at which the image intensity is measured, increases as the lens moves forward. This causes a decrease in sensitivity - unless T is chosen small enough so that accuracy will be limited by f_0 rather than 1/T.

Changes in numerical aperture can be corrected by operating the iris servo - since it doesn't work, one has to account for changes introduced by two effects:

a) Image dimming.

b) Change in accuracy of the focusing function.

Image dimming is not a direct problem since the calculations are normalised. It is however a practical problem since one has to adjust the iris when the lens is close to the vidi-sector (ie. when the numerical aperture is smallest, giving the highest light-intensity). Since intensity depends on the square of numerical aperture, quite a range of intensities is involved and some significant areas may be 'dim-cut-off' when the lens is in the fully extended position.

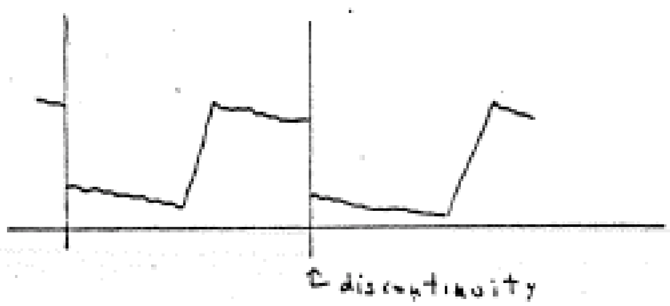
The change in accuracy is due to the fact that the defocusing circles are smaller for smaller numerical apertures. Most of the above effects are of no concern when one has a rough idea of where the lens should be, and are not important problems normally.

Another problem, possibly more of historic interest, was caused by occasional bad points. At times an intensity would be returned by the system in a garbled form - the problem was highly irregular and most prevelant at high intensities. An individual bad point acts like a pulse and thus adds very high values to the transform, resulting in completly useless results for that particular setting of the focus servo. Smoothing the resulting focus function F was found to be quite useless and the disturbance required a non-linear method for removal. It was found that most of the bad spots in F could be located by comparing points to the sum of their immediate neighbours - if high the value at that point was replaced by the average of its neighbours. Since the problem is now well hidden it was never determined wether it was caused by the video-processor or the time-sharing system(or wether it still excists).

The function F presents a last place where one may trade off signal to noise ratio against width of the peak. The trade-off achieved in designing the focusing function appears here in a somewhat garbled and highly variable condition, so it is advisable to smooth F a bit.

<u>One-dimensional versus two-dimensional</u>: The choice of which to use involves two factors, namely speed and area. When using the two-dimensional transform (and doing much the same to it as we did to the one-dimensional one) one can obtain similar results to the ones we obtained. For the same accuracy one requires approximatly the same number of points - except that they are now arranged on a rectangular area. The time required for the transform is still roughly the same, but a much smaller area of the image is used. The result is thus highly dependend on small movements of the camera (the camera moves around quite a bit on the raised floor when somebody walks a few yards away). Also of course, one has to know a bit more accuratly where the area or feature is that one wants to focus on. The alternative is to work with a larger area with a corresponding increase in time. It was found experimentally that the two-dimensional transform allows one to measure the distances to very small areas of the image, provided these areas contained some detail (edge or texture of some kind) and the vidisector was supported to stabilise it w.r.t the object. The one-dimensional approach was used in the general-use program since it is more 'robust'.

<u>Choice of curve in image plane</u>: The DFT assumes a certain periodicity in the function to be transformed. When reading light-intensities this periodicity is not necessarily present and this causes certain problems. In effect one is introducing a sharp edge (with its attendant high-frequency components) at the junction between two cycles.



One can match up the two ends by subtracting a linear function. Theoretically one should match up all the derivatives as well, but in practice this is not required since the components introduced by not matching them are smaller than the noise present. Another, possibly more elegant approach is to measure light-intensities along a closed curve, rather than a straight line, thus assuring the periodicity requirement. This works very well despite the fact that it is difficult to interpret what is meant by a fourier transform along a curved line (The corresponding problem for two-dimensions has not been solved)

THE PROGRAM:

Operation: The user calls the focusing function and gives it two values representing the range of servo positions over which the search is to take place. The program then evaluates the focusing function at about 13 uniformly spaced positions in this range. At each position the light intensities corresponding to two intersecting circles are measured and used as the real and imaginary parts of a function to be transformed. (as is explained in the appendix, one can obtain the transform of two real functions as cheeply as that of one). There are 128 points on each circle. After taking the transform and seperating out the two transforms, the power spectra are formed by adding the square of the real part to the square of the imaginary part. Now all but about the first 5 components are added and divided by the sum of all but the DC component. The value used is the geometric mean of the two values so formed.

The set of values F, so obtained is then 'clipped' to remove noise, smoothed and normalised to have a minimum of zero and a maximum of one. The new range in which the search is to be continued is then found by finding where the function exceeds about .8 . The range is actually made slightly larger 'just in case' (Noise may have caused one to select a range not containing the maximum.)

This new range is then treated similar to the first one except that fewer intervals are used, the peak is sharpened by using a slightly smaller T and slightly larger R_1 and in addition the cut-off for finding the next range is set slighly higher. If necessary a third range is found and searched with similarly modified parameters. At each stage the range is checked to see if it is less than about 40. focus servo units. (0.2 cm). If so the midpoint is chosen, the lens positioned and various useful values printed (such as distance focused at, numerical aperture etc.). The whole process takes about 20 to 30 seconds. The limiting factor is the servo settling time and a more sophisticated search procedure should be able to operate much faster.

The area focused on is on the optical axis, unless the user changes this to a point of more interest to him. Functions are provided for relating coordinates to the standard coordinates and also for obtaining a print-out of intensity values in a selected rectangle on the image. This last item alleviates the one serious practical problem in using the focusing program(and in fact in using the vidi-sector at all) namely finding where the areas of interest lie.

<u>Calling the program</u>: The arithmetic part of the program is written in MIDAS and was loaded together with a relocatable LISP, using STINK. This modified version of LISP was then dumped onto tape RSZ as a file named FOCUS BIN. The required LISP-functions are on the same tape as FOCUS TEXT. The calling sequence is:

Mount tape RSZ on tape n.
In DDT type LISP$J, then
$LFOCUS BIN UTn 2
When loaded, type LISPGO$$G
Answer Y to ALLOC , and allocate
25 blocks of core and 5000_8 Binary Program Space.
Then type (UREAD FOCUS TEXT n) ⌐Q⌐W
The LISP will print (V) when all the functions

have been read in.

The video-processor has to be on, the multiplexor must be in computer-input/computer-output mode and the focus motor switch in the servo position (see Appendix). The focusing function can be called by

    (RSCH R1 R2)

where R1 and R2 are the limits of the range to be investigated (in servo units). The maximum ranges are:

    for the 164mm lens   300. to 2800.
    for the 254mm lens   300. to 1700.

The program will print each new range it is investigating and a number of useful values when it has focused successfully. If it is desired to 'display' a part of the image on a GE-console, [illegible]

    [illegible]

[illegible]

    (LOOK XLOW XUP YLOW YUP)

Where (XLOW,YLOW) is the lower left-hand corner,(XUP,YUP) the upper right-hand corner of the rectangle of interest. The intensity values displayed are multiplied by a scalefactor LSCL, which can be modified to bring the numbers into a convenient range. The coordinates are for RES = 2000_8. Once the coordinates of an area of interest have been determined, one can scale them to the reference plane by executing:

    (SCALE)(SETQ X(XRL x))(SETQ Y(YRL y))

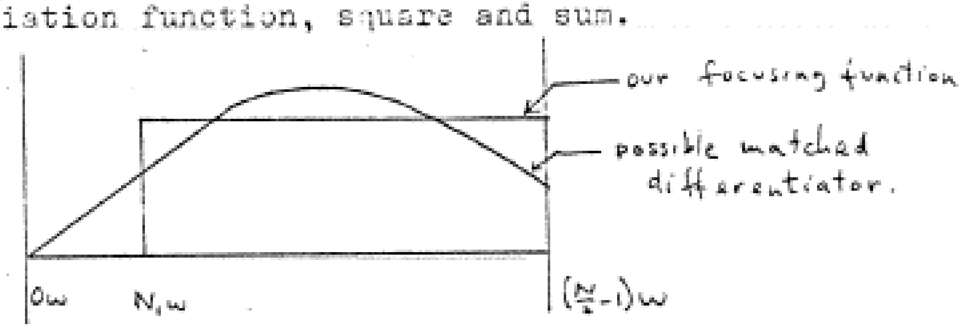
where x and y are the centre of the area one would like to focus on.

Performance: The program came up to theoretical expectations, and was found to perform as well as a human operator observing the monitor screen (where the more noisy signal on the screen is improved somewhat by averaging in the eye). In addition a comparison was made with two methods of finding optimum focus used previously.

1) SUM-OF-SQUARES: The special case that $N_1 = 0$ is of interest since the sum of squares of the transform is equal to the sum of squares of the function. We can expect then that the sum of squares of the intensity function also shows a maximum near best focus. We also know that it will be much 'wider' than the one that has been achieved by throwing out the low frequency components. Small movements of the camera also have a large effect on this measure because of the large resultant change in the low frequency components. Also it is not easy to normalise the result (in fact we divided by this number to normalise, illustrating our faith in just how flat a maximum it has !) All of these problems were born out in practice, the last one being most serious, often the function would indeed have a hill, but it was difficult to find because of the sloping ground it stood on.

2) SUM-OF-SQUARES-OF-DIFFERENCES: This function (suggested by J.L.White) is also more easily analysed in terms of frequency spectra.'Differentiating' in the discrete case is equivalent to multiplying the transform by $(w^{-v} - 1)$. That is de-emphasising low-frequencies and emphasising high frequencies. The first of these two effects is shared with our method, the second gives the hint as to why this function is inferior. The high frequencies consist largely of noise (unless T was much larger than the resolution limit, in which case we have lost sensitivity) and the function then is excessively noisy. This noisyness was confirmed by J.L.White, and also using this program in a modified form.

so numerical analyst would use the straight difference as an estimate of the derivative, but rather use some weighting procedure taking into account more than just the two points. In this way he would be less sensitive to noise but produce distortions on rapid changes of the derivative. He may in fact design a matched differentiation filter to give him the desired trade-off between these two features in the best possible way for the particular class of functions and the particular noise he has to put up with. Such a differentiation approximator is more easily understood in terms of the frequency domain. One has to design a function whose transform matches as closely as possible $(w^{-u} - 1)$ over the range where the signal has high components, yet make it small whereever the noise has high components (things are not just done by aiming over the thumb; there is a respectable theory to this). Typically the function will be a close approximation to $(w^{-u} - 1)$ at low frequencies and fall off at higher frequencies. Using this kind of differentiation would probably be the one way of improving on the focusing function presented in this memo. It will be seen that the rapid way of using this function would again be to first find the transform, then apply the transform of the matched differentiation function, square and sum.



The above discussion should have made it clear that the fourier transform is a tool, rather than an end in itself. It is very useful in thinking about certain aspects of images, such as distortions of various kinds.

Results: For a numerical aperture of 4, typical ranges for the lens position when in focus where found to be (in servo units)

± 10 focusing on newsprint
± 15 focusing on a single edge

Maladjustment of either vidisector or servo can cause ranges to be as bad as 40. The relative accuracy is thus:

$$d_r = .10 \times \left(\left(\frac{d}{f} - 2\right)^2 - 2\right) \quad \text{cm}$$

eg. $(d/f)=6$ $d_r=1.4$ cm

The absolute accuracy is not so good since similar errors are encountered in calibration.

APPENDIX 1: Vidisectors:

Old functions: (for the old vidisector only)

The two arguments to these functions specify the location of the point whose intensity is to be read; they range from 0 to 1777_8.

(VIDI X Y) returns value from 0 to 400_8, proportional to inverse of intensity.

(VIDLIN X Y) returns value from 0 to 400_8, proportional to intensity.

(VIDLOG X Y) returns value from 0 to 400_8, proportional to log of intensity.

New functions: (these use the video-processor and apply to both the old vidisector(TVB) and the new vidisector(TVC))

Before using the new functions one has to initialise the system by calling NVSET. NVSET can be called again if required to change parameters.

(NVSET FIL CONF RES DIM XYZ)

where:	FIL	0-7	Filter selection, any combination of the three filters can be select- by the appropriate bit combinat- ions.(not operative as yet).
	CONF	0-3	Signal to noise ratio; 3 gives highest ratio, but takes longest per point.
	RES	to 40000_8	number of lines in raster (usually 1000_8 to 4000_8)
	DIM	0-7	Dim-cut-off; 0 is most patient ie. darkest cut-off.
	XYZ	0	indicates new vidisector.
		1,2,3	indicates old vidisector.

The two arguments to the following functions again indicate the location of the point whose intensity is to be read; they range from 0 to RES-1.

(NVFIX X Y) returns a value proportional to the log of of intensity; the values range from x to ≈1777_8,

	77		0
where x is	277	when CONF is	1
	477		2
	677		3

(NVID X Y) returns a value proportional to the inverse of intensity; the values range from ≈1.0 to y,

	≈16000.0		0
where y is	≈ 4000.0	when CONF is	1
	≈ 1000.0		2
	≈ 250.0		3

Note: 1. NVID is the only function that returns floating point values.
2. For points that are not cut off by dim-cut-off the values returned are independent of CONF.

SWITCHES: Old vidisector:

1. Small switch on the back of the camera selects signal to noise ratio; usually set in up(high S/N) position.
2. Second small switch on the back of the camera, below 1. selects old or new interface(set according to whether one is using the old or the new functions.
3. On the trolley on which the old vidisector is mounted, is a large black power switch that controls power to the receptacles into which the two sunguns are usually plugged.
4. Above it is a smaller power switch for the camera itself.
5. On the other side of the trolley is a switch selecting the origin of the deflection signals, either the 340 display or the Video-processor (set according to whether one is using the old or the new functions.

Note: To set up the old vidisector it is convenient to set switches 2. & 5. as though one was using the new functions (even if one is going to use the old ones), so as to connect it to the video-processor and allow display on the monitor. They then have to be reset to their correct positions before commencing read-in.

New vidisector:

1. On the right of the main panel of the video-processor is the power switch.
2. To the left is the LIN/LOG switch set according to whether one is using NVID or NVFIX (although only low order bits are affected if it is set incorrectly)
3. The left-most of the row of small switches controls the choice of vidisector connected to the processor; up for the new camera, down for the old.
4. Not far to the right are three small switches which select wether the monitor display is controlled by three pots mounted approximatly below them(switches in up position) or by three remote helipots (presently mounted lower down on the processor).
5. Between 2. & 3. is another small switch which controls the deflection of the monitor; in the up position the monitor behaves normal, ie. when the magnification pot is turned left so as to display a small area, the deflection is increased so as to always fill the monitor screen. When in the down position, the monitor is deflected in step with the vidisector.
6. Small switch on extreme right has to be set in out position or you loose.

<u>Potentiometers</u>:

1. Mentioned in 4. above are three pots just to the right of the meters on the main panel, which control the position of the area which the monitor is displaying and its size.
2. Near the power switch (1. above) is the contrast control for the monitor(marked video-gain), usually turned all the way right.
3. Mounted low down is the helipot controlling monitor focus - do not adjust unless necessary - recommended setting is marked down near it.
4. Gain controls on the two channels of the monitor have standard positions, the left one(y-deflection) should be turned right when using the monitor as a focusing aid(by switching a marked switch behind it into the focus position).When used in this way the vertical deflection is proportional to intensity and using the three pots in 1. to select a small area of the scene, one can adjust the focus so as to give as sharp as possible a rise on some edge.

5. The two raster-rate potentiometers may be set to give a flicker free and line free display, but do not affect the operation in any other way.

6. The two helipots marked 'Manual reference' and 'Reference gain' control two hardware parameters normally of little interest to the user and are usually left in a more or less optimal setting. If somebody has adjusted them and one has no reason to set them to some specific value, settings of 600 for both seem to be reasonable.

Pushbuttons and Meters:

1. 3 meters display various values which vary with the illumination. The video processor trips out the camera if any of them reach a 100 deflection. The right-most one is usually observed when setting the iris. Starting with the iris closed, it is opened slowly until this meter shows between 10 & 50. Above 50 small changes are likely to trip out the camera, below 10 one has to start worrying about dark-cut-off's.

2. When the camera has tripped out, the Anode-warn push-button lights up. After reducing the light (or closing the iris) it can be reset by pressing it.

3. The High-voltage may be off under other conditions than such a trip, and it may then be reset by pressing the High-voltage button.

4. The Raster pushbutton selects manual control of the raster(white light) - allowing the use of the potentiometers mentioned in 1. above. In the red state the processor is ready for the PDP-6 to request input; while busy the pushbutton is not lighted and the monitor displays the position and intensity of the point requested by the CPU. When not busy for more than about a second, the video processor goes into an automatic raster state.

Note: For more details, current problems and cures, consult the log-book under the monitor-scope.

Optics:

Old Vidisector: 55mm f1.2 lens, raster is .001 inch. Resolution is approximatly .002 inch. Distance settings on the lens are not to be trusted, rather use the monitor as an aid in focusing.

New Vidisector: 164mm f3.2 lens, raster is .003 inch when RES is 2000_8. Resolution is approximatly .006 inch. Some drum distortion is evident and not all of the surface is available (ie. some is blocked off). The area of the image is a circle placed not quite centrally with a radius of $\approx 500_8$ when RES is 2000_8.

also: 254mm f5 lens.

APPENDIX 2: Analog Multiplexor:

The analog multiplexor allows one to read settings of potentiometers and to cause servo-controlled motors to operate. In LISP one first has to open the multiplexor by executing:

(MPX T)

More than one person can have the mu tiplexor open at the same time. It can be closed by executing:

(MPX NIL)

To read a value from channel n:

(IMPX n)

To cause the servo controlled motor n to assume position x:

(OMPX n x)

n ranges from 0 to 255; x and the values returned by IMPX are usually some subrange of 0 to 7777_8. Be sure to the limits of the servo you are controlling.

Switches: these are on the multiplexor and the positions for computer operation are underlined.

1. Computer output/test
2. Computer input/clock

When not in use, these switches should be in the un-underlined position. Certain servos are slaved to certain pots in this position. This for example allows fine setting of the focus servo. Channels of interest to the vidisector user:

	Servo	Read-in	Slaved to
Iris	32_8	33_8	132_8
Focus	33_8	34_8	133_8

Thus in the test - clock position one may adjust focus by using potentiometer 132_8. In the computer input - computer output position one can cause the servo to go to a certain position by executing:

(OMPX 33 3000.)

Control is returned immediatly to the user program and it is his responsibility to check on the servo's position by executing: (IMPX 34)

Rather than loop on this test it is fair to other users to use the function:

(SLEEP n)

which returns control to the program after n/30. seconds.

n should be about 10. plus(units to travel / 25.)

Both Iris and focus motors can be operated manually and the switches which select manual or servoed(ie. computer output or slaved) are in a small box near the vidisector. These switches have a third position (straight out) which is the one they should be left in when not being used. Above these are the two switches for manually operating the motors.

Various useful (?) constants:

Distance from front of main body of vidisector to surface of vidisector: 93mm

Distance of equivalent lens of 164mm focal lens from vidisector:

306. - 0.0459 * CH34 in mm's
330. - 0.0459 * CH34 (for the 254 mm lens)

Diameter of iris:

0.0532 * CH33 - 75. in mm's

where CH34 & CH33 are the values returned by analog channels 34_8 & 33_8 respectivly.

The limits on the servos are:

Focus: 240. (all the way out) to 2860.

Iris : 1500.('closed') to 2300.

The servos can be expected to settle within 10. (better near centre of range - more like 5.) of the value requested.

Problems: 1. Iris servo inoperative.

2. For some unknown reason servos may suddenly depart from their position, hunt around for a second or two and return to their position.

APPENDIX 5: References

Fourier transforms: N.Wiener: "GENERALIZED HARMONIC ANALYSIS AND TAUBERIAN THEOREMS".

Fast Fourier trans: Rader & Gold: "WHAT IS THE FAST FOURIER TRANSFORM ?" June 67, Proceedings of the IEEE.

Application of fourier techniques to optics: P.M.Duffieux: "L'INTEGRAL DE FOURIER ET SES APPLICATIONS A L'OPTIQUE" 46, Besancon

F.D.Smith: "OPTICAL IMAGE EVALUATION AND THE TRANSFER FUNCTION" April 63, Applied Optics

Application to Pattern Recog.: J.D.ARMITAGE & A.W.Lohmann: "CHARACTER RECOGNITION BY INCOHERENT SPATIAL FILTERING" April 65, Applied Optics.

Also interesting: L.J.Cutrona: "OPTICAL AND ELECTRO-OPTICAL INFORMATION PROCESSING" 65, M.I.T Press.

T.G.Stockham Jr.: "HIGH SPEED CONVOLUTION AND CORRELATION" 1966, Spring Joint C.

R.Nathan: "DIGITAL VIDEO-DATA HANDLING" NASA Technical Report 32-877, JPL, Jan 66 Also NASA CR-71399, N66-21646.

A.V.Oppenheim, R.W.Schafer, T,G.Stockham Jr.,
"THE NON LINEAR FILTERING OF MULTIPLIED
AND CONVOLVED SIGNALS" (to be published)
IEEE Proceedings 68.