OBJECT ENHANCEMENT AND EXTRACTION

Judith M. S. Prewitt
University of Pennsylvania
Philadelphia, Pennsylvania

INTRODUCTION

Object enhancement, extraction, characterization and recognition are closely linked aspects of visual perception, and cognizance of their interplay as well as their individual roles is important for the construction of successful image processing automata. The last two decades have produced a proliferation of models and computer realizations of sensory and cognitive processes in which both the degree of separation of these functions, and the manner of their integration, vary greatly. Past designs have been influenced heavily by practical considerations, and have tended to be simplistic. Only recently has technology advanced enough to permit unified consideration of all of these operations.

In early work in automated character recognition and aerial reconnaissance, object detection and recognition were often performed simultaneously. Images were either binary or high contrast, and consisted of standardized shapes or well defined curvilinear contours. Techniques for eliminating spurious noise in digitized images and for mending broken contours were either embedded in the recognition scheme (usually a variant of template matching), or else handled with a modicum of auxiliary processing. Objects of interest were identified with little or no explicit articulation. Surveys of this early work may be found in [127, 131, 202]. In approaches using representations such as Fourier transforms and Haar-Walsh expansions, the distinction between detection and characterization is also obscured, but these approaches have the advantage that no information is lost in the transformation from the original image domain to the domain of the representation, and objects can be resynthesized by inversion of the process [e.g., 3, 31, 49, 60, 227].

As attention turned toward less stylized scenes with greater tonal and topological information content, the distinction between sensory and
cognitive processes was formalized. Feature extraction and object detection were considered precursors to description and identification. This division of function was harmonious with the modular viewpoint imposed by machine construction and programming. At the time, it was also the only practical approach to complex material [e.g., PIP, OCR, 139].

When the potential of image-processing automatons in scientific applications became evident, the relationship between reliable interpretation of visual material and image fidelity, long an acknowledged consideration in commercial communication systems, again became a compelling factor in design. Image restoration and enhancement research was revived principally because of interest in two areas: (1) video systems for terrestrial and extraterrestrial exploration, in which spatial and tonal degradations are unavoidable but almost completely reversible, and (2) systems for biomedical image processing, including quantitative scanning microscopy and interpretive medical radiography, in which object size is often at or near the limit of resolution, and anomalous imaging may result. In both cases, the resultant images are strongly dependent on the transfer function of the total electro-optical configuration, and proper image interpretation depends on correcting or compensating for degradations. Furthermore, the amount of data to be transmitted or stored is so great that optimal spatial and tonal quantization and efficient encoding became important considerations in data acquisition, and materially affected subsequent image manipulation [e.g., CIEEIP, PPR, 222]. Image processors which were designed to handle these additional requirements still had a linear structure. They generally consisted of four types of independent units: transducer, preprocessor, feature extractor, and classifier; they performed signal detection, enhancement, object description, and identification, in that order.

Most recently, serious thought has been given to less trivial pictorial material and scenes. Realistic complications include the following:

1. fragmentation, coalescence, and disappearance of objects and object parts because of high noise levels and limited resolution;
2. pseudo-resolution: the appearance of spurious objects and parts due to a singular relationship between the object and the transfer function of the transducer;
3. differences in illumination, contrast, and signal to noise ratio over the field of view;

4. variations in the amount of detail (high spatial frequency content) over the field of view;
5. distortions due to the projective mapping from three to two dimensions;
6. occlusion or overlapping of objects by others which intervene in the line of sight;
7. elision of objects by shadows;
8. attenuation of high spatial frequencies, hence obliteration of fine detail and blurring of boundaries, because of phenomena such as glare, atmospheric turbulence, or fog;
9. motion of object or sensor;
10. ambiguity resulting from the use of a single function such as monochromatic reflectivity or transmissivity, which may be resolved by use of additional functions such as color and distance;
11. intrinsic ambiguity due to uncertainty principle considerations.

Linear models cannot cope realistically with such complexities, and are being supplanted by models with cyclical or recursive structure, which can interleave detection and enhancement with parameterization and description, vary resolution, and change the focus or type of effort, so that tentative decisions and descriptions can be refined. This approach is implicit in many programs, albeit in rudimentary form. Such semi-formal heuristic models, as well as syntactic models, have been used in the analysis of noisy photomicrographs [20] and bubble chamber tracks [79, 189, 229, 331], in robot design [42, 59], and in the analysis of hand printing, handwriting and I/I-formed printed text [57, 64, 122]. In these higher level models, the organized collection of information by primitive feature extractors, the assemblage of pictorial elements into objects, and the description of their relationships are all in accordance with explicit formulae for well formed scenes [98, 123, 129, 194].

Even with these complex models, the central and most difficult problem is characterization, i.e., extracting sets of significant features, properties, and relationships from a background of irrelevant detail. In most existing systems, a finite set of presumably robust and adequate features is preprogrammed. Insensitivity of individual feature extractors to noise and to acceptable variations in object size, registration, orientation, contrast, texture, and topological deformation is obviously desirable. On the other hand, redundancy in ensembles of feature extractors and object characterizers is wasteful and can actually dilute measurement of overall
discriminatory achievement [165, 170]. Criteria exist for the deletion of irrelevant or uninformative features and the optimal utilization of informative features [170, 211, 228], and there are some results on the theoretical limitations of certain popular discrimination techniques [130]. Little guidance is available, however, on the generation of suitable new features [for exceptions, see 217]. Thus the success of current systems depends basically on the intuition and ingenuity of the programmer and designer.

This paper is concerned with accomplishments in these difficult tasks of feature and object extraction and enhancement. An enormous amount of research has been done in the last 25 years, but reports are scattered throughout the literatures of several disciplines. Although collected readings, conference proceedings, and surveys have appeared from time to time, little of this diverse material has been unified (until the recent appearance of [183]) and evaluated. Rather than aim for complete and definitive coverage, we have set the more modest goal of presenting basic principles with illustrative implementations.

Image formation may be described mathematically in terms of degradation operators or spread functions which characterize the optical communication channel. Image processing in turn can be thought of as modification or transformation of the image by (1) restoration operators, (2) enhancement operators, (3) structural operators or feature extractors, and (4) syntactic schemata. The basis of this classification is function; a given operator may be used in several capacities.

Restoration operators compensate for the system spread function and other noise and perform the inverse operation of the object signal. Degradation and restoration are thus inverse operations. Enhancement operators change the image further rather than restore it—directly, by masking or convolution, or indirectly, by emphasizing or deemphasizing the spatial frequency and temporal content. Both restoration and enhancement may be used as aids to either human or automatic photointerpretation.

Structural operators detect picture elements such as contours, features and connected regions. Syntactic schemata direct the concatenation or assemblage of picture elements into figures and more complex configurations; in our sense they are higher level, organizational procedures.

ENHANCEMENT AND EXTRACTION

In the body of this paper, we shall discuss the use of such operators for object recognition and detection. A paper on object recognition and on syntactic schemata is in preparation.

IMAGE FORMATION

Image formation involves a spatial reorganization of the irradiance pattern of an object by a process that is not one to one, so that the recorded image is at best a degraded copy of the original. The information content of the recorded image is limited by the resolution of the imaging system and by the presence of both random and systematic noise. Whether restoration and other enhancement processes are warranted or even possible in any particular case depends on the transfer characteristics of the imaging system and also on the purpose for which the image was recorded.

The mathematical theory of integral transformations is a suitable framework for describing many aspects of image formation, restoration, enhancement, and object detection. The integral transform $F(u)$ of a function $f(r)$ is defined by an equation of the form

$$F(u) = \int_0^\infty f(r)K(u,r)\,dr$$

where $K(u,r)$ is a specific function of $r$ and $\omega$ called the kernel of the transform. The variables $r$ and $\omega$ may be real or complex numbers or vectors, and the limits of integration $\alpha$ and $\beta$ may be finite or infinite. If the kernel $K(u,r)$ and the transform $F(u)$ are known, then the process of determining a solution $f(r)$ of the integral equation is called inversion. The properties of the original function $f$ and the kernel $K$ over the region of integration determine the existence of the integral transform $F$ and the existence of (unique) inversion formulae or algorithms giving $f$ as an integral transform of $F$:

$$f(r) = \int_0^\infty F(u)K^{-1}(u,r)\,du$$

Mathematical conditions are detailed elsewhere; they are presumably satisfied for the cases under discussion here, but should be verified in specific applications [31, 205, 214, 237].

In using this formalism to describe two-dimensional image formation, the function itself represents the object pattern, the kernel rep-
JUDITH M. S. PREWITT

represents the action of the imaging system, and the transform gives the resulting image. The kernel may be thought of as a blurring operator or spread function $s$, which modulates and combines intensity values at all luminous points on the original object $o$, and thereby forms the image:

$$i(x, y) = \int -\infty \rightarrow \infty \int -\infty \rightarrow \infty o(\xi, \eta)s(x, y, \xi, \eta)d\eta d\xi$$

Restoration is mathematically equivalent to finding the appropriate inversion formula

$$o(x, y) = \int -\infty \rightarrow \infty \int -\infty \rightarrow \infty i(x, y)s^{-1}(x, y, \xi, \eta)d\eta d\xi$$

The imaging process is schematized in Fig. 1. See [108, 121, 153].

**Figure 1.** The imaging process. The amplitude at any point in the image is built up from contributions associated with all luminous points in the object. This is described mathematically by the convolution of an object $o(x, y)$ with a spread function $s(x, y)$. After Pernin [153].

Image processing operations such as enhancement can also be represented within this framework. If $o$ is the original pattern and $\Omega$ is the processing operator, then $p = o * \Omega$ or $p(x, y) = [o * \Omega](x, y)$ represents the processed pattern. Shifting the original pattern is described by translation operators

$$[T_xo](x, y) = o(x - \xi, y - \eta)$$

and uniform amplification is described by scaling operators

$$[c o](x, y) = c \cdot o(x, y)$$

Most operators in current use satisfy the following criteria:

(a) $\Omega$ is homogeneous or position invariant, i.e., it commutes with all translation operators

$$\Omega T_{\xi, \eta} = T_{\xi, \eta} \Omega$$

(b) $\Omega$ is linear, i.e., it commutes with all scaling operators and satisfies a superposition theorem

$$\Omega(a_1 o_1 + a_2 o_2) = a_1 \Omega o_1 + a_2 \Omega o_2$$

(c) $\Omega$ is local, i.e., $[\Omega o](x, y)$ depends only on the values of $o$ in a specified neighborhood of the point $(x, y)$ where $N(x, y)$ is obtained by translating a neighborhood $N$ of the origin

$$N(x, y) = (x', y') \mid (x' - x, y' - y) \in \Omega$$

Some operators also satisfy

(d) $\Omega$ is isotropic, i.e., rotation and reflection invariant.

In these cases a large body of mathematical theory becomes applicable to image processing [48, 149]. It is well known that any homogeneous linear operator $\Omega$ is equivalent to a convolution integral, in which the kernel is a function of two displacements rather than four spatial coordinates, and conversely:

$$s(x, y, \xi, \eta) = \delta(x - x', y - y')$$

Thus, to each such operator $\Omega$ there corresponds a kernel function $s(x, y, \xi, \eta)$ which allows us to describe the action of $\Omega$ on any function $f$ in its domain as a convolution integral:

$$\Omega f = s(x, y, \xi, \eta) f(x, y, \xi, \eta)$$

or

$$[\Omega f](x, y) = \int \int f(x, y, \xi, \eta) s(x, y, \xi, \eta) d\eta d\xi$$

80

ENHANCEMENT AND EXTRACTION

81
Conversely, any such convolution integral is a homogeneous linear operator.

In systems where degradation is virtually independent of location, image formation itself can be described using a convolution integral:

\[ i(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x-u, y-v) f(u, v) \, du \, dv \]

or \( i(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x-u, y-v) f(u, v) \, du \, dv \)

The kernel of this integral is called the point spread function. Furthermore, the result of applying any homogeneous linear operator \( \Omega \) to the blurred picture \( i \) can be expressed in terms of a modified spread function \( s = \Omega h \) as follows:

\[ i(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x-u, y-v) f(u, v) \, du \, dv \]

In practice, the spread function is essentially a local operator and the limits of integration become finite. In some cases, the degradation is also isotropic or direction independent, the corresponding circularly symmetric spread function has the form:

\[ s(x, y) = s(x, y) = s(r) \]

and additional simplification of the equations is possible through the use of polar coordinates:

\[ x = r \cos \theta \]
\[ y = r \sin \theta \]

Restoration is the inverse convolution

\[ x^{-1}(x, y, z) = z^{-1}(x, y, z) \]

Given a processing kernel \( \Phi \), it is sufficient to know the effect of \( \Phi \) on the inverse spread function, since

\[ \Phi \Omega = \Phi (s_{-1}^{-1}) \]

Most useful in image processing are those linear integral transforms for which there are "convolution-product" and "inversion" theorems. For such a distinguished kernel, if \( f \) and \( g \) are arbitrary functions whose integral transforms are

\[ T[f] = F \quad \text{and} \quad T[g] = G \]

respectively, then the integral transform of the convolution of \( f \) and \( g \) is found by multiplication:

\[ T[f \ast g] = T[f]T[g] \]

Furthermore,

\[ T^{-1}[FG] = f \ast g \]

Also in extensive use are two operations derived from convolution, the cross correlation and autocorrelation, which measure goodness of match or fit of two patterns. Let \( I \) be the result of reflecting \( f \) in the origin:

\[ I(x, y) = f(-x, -y) \]

The cross correlation \( f \times g \) is defined by

\[ f \times g = f \ast g = \int_{-\infty}^{\infty} f(x, y) g(x, y) \, dx \, dy \]

and the autocorrelation \( f \times f \) is defined by

\[ f \times f = f \ast f \]

Symmetry and antisymmetry properties of these correlations make them useful in image analysis:

\[ f \ast g = f \times g \]

Convolution and inversion theorems are known for several kernels, such as the Laplace and Mellin transforms, but the Fourier and Hankel transforms have been the most useful in object detection and enhancement for several reasons: first, they can readily be interpreted in terms of spatial frequency; second, they can be implemented efficiently by digital as well as electro-optical and optical means [27, 28]. Two-dimensional moments [1, 30] and transforms involving Haar-Walsh functions [51, 118] and Chebyshev polynomials [18] have also been used.

The complex Fourier transform \( T[f] \) of a univariate function is defined by

\[ \hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ix\omega} \, dx \]
and the corresponding inverse transform $T^{-1}[F]$ is given by

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega.$$ 

$F(\omega)$ is generally complex valued with real and imaginary parts comprising the cosine and sine transforms

$$\text{Re } F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \cos(\omega t) dt$$

$$\text{Im } F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \sin(\omega t) dt$$

Related to these is the power spectrum $P(|f|)$ defined by

$$P(\omega) = |F(\omega)|^2 = \text{Re} F(\omega) \cdot \text{Re} F(\omega)^* + \text{Im} F(\omega) \cdot \text{Im} F(\omega)^*$$

where $^*$ denotes complex conjugation. For functions of two variables, the transform $T[\cdot]$ and the inverse transform $T^{-1}[\cdot]$ are defined by

$$F(x,y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x',y') e^{i(x'x+y'y)} dx' dy'$$

$$f(x,y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x',y') e^{-i(x'x+y'y)} dx' dy'$$

The Fourier transform satisfies the invariance theorem

$$T^{-1}[T[f]] = f$$

the linearity or superposition theorem

$$T[a f + bg] = aT[f] + bT[g]$$

the symmetry-antisymmetry theorem

$$T[f](t) = T[f](t)$$

$$T[f](t) = T[f]^*$$

the shift and rotation theorem

$$T[f(x + a, y + b)] = e^{-2\pi i (ax + by)} T[f](x,y)$$

$$T[f(x, y) e^{i\alpha x + i\beta y}] = T[f+(\alpha, \beta)]$$

the product-convolution theorem

$$T[g * h] = T[g] T[h]$$

$$T^{-1}[T[g] T[h]] = g * h$$

and the Wiener-Khinchin theorem on power spectra

$$T[f \ast f](\omega) = |T[f](\omega)|^2$$

The Hankel transform enters naturally in the analysis of systems with isotropic point spread functions and greatly facilitates restoration. The Hankel transform of order $n$ is defined by

$$H_n[f](r) = \int_0^{\infty} f(\tau) J_n(\tau r) d\tau$$

where $J_n(z)$, the $n$th order Bessel function, has the integral representation

$$J_n(z) = \frac{1}{\pi} \int_0^\pi \cos(z \sin \theta - n \theta) d\theta = \frac{1}{2\pi} \int_0^{\pi} e^{i(z \sin \theta - n \theta)} d\theta$$

In particular,

$$J_n(z) = \frac{1}{2\pi} \int_0^\pi e^{i(z \sin \theta - n \theta)} d\theta$$

(a arbitrary)

The Hankel transform has circular symmetry and satisfies scaling and convolution theorems, and has a particularly simple inversion formula:

Scaling: $HI[(\cdot) \omega] = \frac{1}{\omega} H[I(\cdot)]$  

Convolution: $H[\cdot \ast \cdot] = H[I(\cdot)]$  

Iversion: $HI[H[f]] = f$

By changing from rectangular to polar coordinates,

$$x = r \cos \theta$$

$$y = r \sin \theta$$

any function $f$ and its two-dimensional Fourier transform $F$ become

$$f(x,y) = F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\omega) e^{i\omega(x \cos \theta + y \sin \theta)} d\omega$$

$$F(r, \theta) = F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\omega) e^{i\omega(r \cos \theta + \phi)} d\omega$$

If $f$ has circular symmetry, $f(x,y) = f(r, \theta) = f(\sqrt{x^2 + y^2})$
then the direct and inverse Fourier transforms reduce to the simpler, univariate, Hankel transform

\[ F_0(\omega) = \int_0^\infty f(r)r \sin(\omega r) dr = \int_0^\infty f(r)J_0(\omega r) dr = \mathcal{H}[f] \]

and

\[ f(r) = \int_0^\infty F_0(\omega) J_0(\omega r) d\omega = \mathcal{H}^{-1}[F] \]

Analogous sine, cosine and Hankel transforms can be defined when the limits of integration are finite rather than infinite, and appropriate inversion formulas can be derived [200]. For example, the finite sine-cosine transforms are identical to the coefficients of the ordinary Fourier series; the inverse is the series itself.

\[ f(x) = \int_0^a f(x) \sin \omega x dx \quad (\omega \text{ an integer } > 0) \]

The finite Hankel transform is defined by bounding the range of integration:

\[ H(\omega) = \int_0^b f(r) J_\omega(\omega r) dr \]

where \( \omega \) is a positive root of \( J_\omega(\omega) = 0 \). By the theory of Fourier-Bessel series [31], \( f(r) \) can be represented by a Bessel series in the range \( 0 \leq r < 1 \):

\[ f(r) = \sum_n a_n J_\omega(\omega) \]

where the coefficients are

\[ a_n = \frac{2}{J^2_{\omega+1}(\omega)} \int_0^b f(r) J_\omega(\omega r) dr = 2H(\omega) / J^2_{\omega+1}(\omega) \]

Hence the inversion formula for the finite Hankel transform is

\[ f(r) = \sum_{\omega > 0} f(\omega) J_\omega(\omega r) / J^2_{\omega+1}(\omega) \]

Since real world images and spread functions have limited spatial extent, these finite transforms may be more relevant for picture processing.

ENHANCEMENT AND EXTRACTION

IMAGE RESTORATION

Efforts to extract features or objects can be in vain when there is severe distortion of the image. What is enhancement in one application may be a hindrance in others. Corrections needed in preparation for quantitative image analysis may be entirely superfluous for qualitative visual perception. If the nature of the degradation can be determined in advance, either from theory or by using known test objects to calibrate the imaging system, then corrective measures can be applied to new images as warranted.

Degradation may be caused by system dependent or object dependent factors, which can generally be classified into (1) deterministic factors, including (a) distortions of the spatial coordinate system and the photometric scale, and (b) blurring, which can be described as convolution with some point spread function; and (2) additive stochastic noise which is superimposed either before or after imaging.

Restorative procedures can also be classified according to the type of degradation to which they are addressed. Specific techniques include: (1) construction of tables of corrections and interpolating functions; (2) inversion of integral transforms; and (3) optimal filters for minimizing the mean square error of reconstruction or for maximizing the signal to noise ratio. Many of the concepts connected with restoration are also relevant to nondestructive image enhancement (e.g., sharpening) and feature detection.

Rectification of geometric distortion and photometric errors which are spatially and temporally stable (e.g., consistent slanting of lines; fixed but uneven illumination) involves changing coordinates and brightness values throughout the image. Corrections can be tabulated for selected points and suitable interpolation formulas used at intermediate points, or they can be summarized in the form of fitted functions.

Displacements for straightening the image coordinate raster can be obtained using a two-dimensional test grid. Photometric errors can be corrected by using the results for uniformly illuminated test fields, or the correction can be provided pointwise, on-line, in the form of a compensating reference signal. In most cases, the restoration involves nonlinear, position dependent operators [130, 143].

Most methods of restoration and enhancement involve homogeneous linear operators for which convolution-product or inversion theorems
hold. From a mathematical point of view, any invertible operator will suffice. Since imaging is a convolution, the object, image, and spread function transforms are related by


Formally, the restored picture is obtained by inverting,

$$O[F] = F[1/F[1]]$$

A technique that is often unsatisfactory because the reciprocal $1/F[1]$ is singular, the ratio $F[1]/F[1]$ is indeterminate or undefined, or the restored image will be heavily contaminated by restored noise.

The main steps in applying inversion methods to restorative enhancement are: selection of an appropriate integral transform (kernel), determination of the transform of the spread function, and inversion of its reciprocal. The transform of the spread function depends only on the relationship of the latter to the selected kernel, and in principle it can be determined by using any convenient object-image pair:

$$1/F[1] = F[1/F[1]]$$

Points, slits or lines, and especially edges, are the simplest and most widely used test objects.

If objects are thought of as mass distributions over the $x-y$ plane, then the response of the imaging system can be described [148] in terms of two-dimensional line impulses or generalized Dirac delta functions $\delta(x,y)$ representing a mass distribution of density

$$\lambda(x,y) = [(\omega_0/2\pi)^2 + (\omega_0/2\pi)^2]^{-1/2}$$

along the curve $s(x,y) = 0$.

Point masses located at the intersection of two curves $s(x,y) = 0$ and $s(x,y) = 0$ are described by

$$\delta(s(x,y)) \delta(s(x,y)) = \sum \delta(x-x_0) \delta(y-y_0)$$

in which the elementary impulses $\delta(x-x_0)$ and $\delta(y-y_0)$ represent impulse distributions on the lines $x-x_0$ and $y-y_0$. For a given degradation operator $D$, the value of the point spread function (PSF) $u(x,y)$ itself gives the response to a point mass located at $x=x_0$, $y=y_0$.

88

ENHANCEMENT AND EXTRACTION

$$PSF(u,v) = \iint f(x,y) \delta(x-u) \delta(y-v) dv$$

while the image corresponding to an object $o=c$ is also uniform:

$$k(x,y) = \iint f(x-u) \delta(y-v) dv$$

The response to an impulse situated on the $y$-axis is the vertical line spread function (LFS):

$$LFS(u,v) = \iint f(x-u) \delta(y-v) dv$$

and the result for a straight edge similarly oriented ($o=0$ in the left half plane and 1 in the right half plane) is called the vertical edge spread function (ESF) or edge trace

$$ESF(u,v) = \iint f(x-u) \delta(y-v) dv$$

Formally, the ESF is the derivative of the ESF:

$$ESF(u,v) = \iint f(x-u) \delta(y-v) dv$$

Degradation and restoration have customarily been analyzed using the Fourier transform. The output and image transform:

$$F[0](u,v) = O(u,v) = \iint f(x,y) e^{-j2\pi u x + j2\pi v y} dv$$

and

$$F[1](u,v) = I(u,v) = \iint f(x,y) e^{-j2\pi u x + j2\pi v y} dv$$

are related via the complex valued system transfer function

$$F[s] = F[1]/F[o] = \iint f(x,y) e^{-j2\pi u x + j2\pi v y} dv$$

In vector notation

$$a = x + iy$$

the complex transfer function is written as

$$M(a) = M(u,v) = F[s](u,v) = \iint f(x,y) e^{-j2\pi u x + j2\pi v y} dv$$

89
ENHANCEMENT AND EXTRACTION

The real component is caused by the cosine transform, and the imaginary component by the sine transform, in the Fourier transform. The transform itself is given by

\[ M(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') \cos(\pi x' x) \sin(\pi y' y) \mathrm{d} x' \mathrm{d} y' \]

and the phase transfer function as

\[ \mathcal{F}(\omega_x, \omega_y) = M(\omega_x, \omega_y) e^{-i \phi(\omega_x, \omega_y)} \]

where \( \phi(\omega_x, \omega_y) \) is the phase shift caused by the spatial frequency \( \omega_x, \omega_y \).

Monte Carlo [21] gives an extensive analysis, using Fourier techniques for the concept of shooting rays, and transmission. The distribution model for the object and the imaging system is shown to

\[ F_\text{sys}(\omega_x, \omega_y) = G_{\text{sys}}(\omega_x, \omega_y) \mathcal{F}(\omega_x, \omega_y) \]

where \( G_{\text{sys}}(\omega_x, \omega_y) \) represents the system transfer function.

Fig. 2 illustrates the concept of image formation for a two-dimensional spatial light modulator (SLM) with an intensity function.

The SLM is also shown as the image of a single-slit function.

These transfer functions are readily interpreted in terms of the line

\[ M(\omega_x, \omega_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') \cos(\pi x' x) \sin(\pi y' y) \mathrm{d} x' \mathrm{d} y' \]

and the phase transfer function as

\[ \mathcal{F}(\omega_x, \omega_y) = M(\omega_x, \omega_y) e^{-i \phi(\omega_x, \omega_y)} \]

where \( \phi(\omega_x, \omega_y) \) is the phase shift caused by the spatial frequency \( \omega_x, \omega_y \).

Monte Carlo [21] gives an extensive analysis, using Fourier techniques for the concept of shooting rays, and transmission. The distribution model for the object and the imaging system is shown to

\[ F_\text{sys}(\omega_x, \omega_y) = G_{\text{sys}}(\omega_x, \omega_y) \mathcal{F}(\omega_x, \omega_y) \]

where \( G_{\text{sys}}(\omega_x, \omega_y) \) represents the system transfer function.

Fig. 2 illustrates the concept of image formation for a two-dimensional spatial light modulator (SLM) with an intensity function.

The SLM is also shown as the image of a single-slit function.
Figure 2. Image formation for a sinusoidal input: (a) hypothetical PSF of width \( d \); (b) response to spatial frequency \( v \) (measured in units of \( 1/d \)); (c) image functions (dotted) corresponding to sinusoidal object functions (solid) of spatial frequencies \( 1/2d \), \( 1/d \), and \( 2/d \). After Perrin [153].

\[
M^*(u) = \int_{-\infty}^{\infty} LSF(x) e^{-2\pi i u x} dx = 2 \int_{0}^{\infty} LSF(x) \cos(\pi x u) dx = MTF(\sqrt{u^2 + v^2})
\]

Further, if the Hankel inversion can be performed, the spread function is recoverable as

Figure 3. Anomalous imaging effects showing the limitations of the MTF as a descriptor of resolution: (a) two hypothetical spread functions with different imaging properties; (b) edge traces for the spread functions in (a); (c) corresponding modulation transfer functions; (d) as measured by the MTF, the peaked function \( b \) compromises detail more than does the bell curve \( a \), but \( b \) is better than \( a \) in resolving two point sources. After Perrin [153].
$S^-(\omega) = \sum_{\xi} \frac{1}{2\pi} \int_{-\infty}^{\infty} M(\xi)I(\xi)e^{i\omega \xi} d\xi$

As an example, Rao and Jain [171] have derived MTF's for symmetric gaussian and exponential functions.

To make practical use of these ideas, numerical analysis can be used to estimate the spread function from regularly spaced photometric samples of an edge trace, based on the observation that the LSF is the derivative of the ESF. The ESF is estimated by least squares trigonometric approximation of the edge trace, and the LSF is estimated from this by direct differentiation [188]. The MTF and PTF for the isotropic case can then be approximated by taking sine and cosine transforms. Taking this one step further, the PSF can be gotten from the MTF by numerical integration. Errors can be ascertained for each computation (also see [163, 233]).

Given these pertinent quantities, operations can be implemented either directly, in the spatial $(x,y)$ domain, or indirectly, in the frequency $(u,v)$ domain. The indirect method consists of multiplying the Fourier transform of the picture with a corrective filter $1/MTF$ in the frequency domain, and then reinverting the resultant compensated spectrum back to the spatial domain. In the direct method, the picture is convolved in the original spatial domain with a mask gotten by inverting this corrective filter. The direct method has long been used for digital implementations. The indirect method (spatial frequency filtering) has been preferred in analog systems, but since the development of the fast Fourier transform [3, 29, 152] it is reasonable to use it for digital processing as well.

In terms of the Fourier transform, restoration

\[ o(x,y) = \int \left( \int E(u,v) e^{2\pi i (ux + vy)} dv \right) du \]

is achieved by applying the operator

\[ s^{-1}(x,y) = \int \left( \int F(s)(u,v) e^{2\pi i (ux + vy)} dv \right) du \]

where $F(s)$ is the transform of the spread function. As we have just seen, recovery from isotropic degradation involves just the ordinary MTF and the operator simplifies to a Hankel transform of $1/MTF$

\[ s^{-1}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi \gamma^2}} e^{-(u^2 + \gamma^2)} e^{2\pi i (ux + vy)} du dv \]

\[ = 2\pi \int_{0}^{\infty} \left( \frac{1}{\sqrt{\pi \gamma^2}} \right) I_s(\gamma) \gamma d\gamma \]
Practical and theoretical difficulties limit the applicability of this convolution-inversion approach. The picture may be complicated by superimposed stochastic noise, or the degradation may be anisotropic, and require more than the ordinary MTF. The method will also fail to give perfect restoration whenever the quotient \( F[I]/F[I] \) becomes either infinite or indeterminate. Modifications have been proposed to make this a feasible, albeit imperfect, method of restoration and several more realistic alternatives have been suggested as well.

When the spectrum of superimposed noise and the object spectrum are fairly distinct, as is the case with systematic noise (where the frequencies associated with noise cluster in the spatial frequency domain), or with high frequency noise and a low frequency object, a truncated MTF is useful. For noise centered around discrete frequencies, sharply peaked filters are appropriate. But if they are too broad they will also remove picture information, and if too sharp they will leave residual noise. Examples of such filters are square and triangular truncations of the cosine:

\[
sine(\omega_0)F(\omega) = \frac{\sin(\omega_0\omega)}{\omega_0}
\]

and

\[
sine(\omega_0)F(\omega)
\]

where \( \omega \) is a cutoff sharpness parameter. Periodic systematic noise can be detected by cross correlation with sinusoids, and subtractive correction can be applied to the image.

In cases where it is desirable to avoid overemphasis of high frequencies or to suppress singularities of a band limited MTF, a cutoff frequency \( \omega_0 \), determined by the characteristics of the object and the noise, is chosen so that noise information will not defeat image enhancement. A modified system MTF defined by

\[
\text{modified MTF}(\omega) = \text{MTF}(\omega) \text{ if } \omega < \omega_{\max} \\
= \text{constant} > 0 \text{ if } \omega > \omega_{\max}
\]

is inverted and convolved with the picture. Instead of the ideal flat frequency response, the actual response is level up to the cutoff fre-

quency and then tapers off. This correction can also be made contextual, by turning the filter off at points of high contrast detail where the local differences in brightness exceed a threshold [13, 116, 143].

If some extraneous information or residual degradation is tolerable in the recovered image, compromise processors can often be found. Let \( \alpha \) and \( \beta \) represent original object, degraded image, and processed picture, respectively. If the system spread function \( S_{00} \) is known, and if a final response spread function \( S_{\text{ret}} \) can be specified after considering a prototype image and an acceptable processed version, then the desired processing spread function \( \text{Desired} \) is given by

\[
S_{\text{Desired}} = S_{00} * S_{\text{ret}}
\]

and its Fourier transform by

\[
F[I_{\text{Desired}}] = F[I_{\text{Desired}}]/F[I_{\text{Original}}]
\]

The processed picture

\[
p = \text{Desired} \ast F[I_{\text{Original}}]
\]

\[
= F^{-1}(F[I]F[I_{\text{Desired}}])
\]

\[
= F^{-1}(F[I]F[I_{\text{Original}}]/F[I_{\text{Original}}])
\]

can be compared with a perfect restoration

\[
\alpha = F^{-1}(F[I]/F[I_{\text{Original}}])
\]

As Fig. 5 indicates, for any specific operation the existence of the desired compromise processor must be verified. For example, a square PSF acting on a line (delta function) object produces a square wave input image with corresponding input MTF equal to sine (\( \omega \)). If a line output is demanded, the response PSF must also be a delta function, with a flat MTF. The required processor has MTF equal to \( 1/\text{sine} (\omega) \), a function which has unrealizable infinities. On the other hand, if a centrally restored image with accompanying low-intensity ghosts is acceptable, the desired output MTF is \( \text{sine}^2 \) of the required processor spectrum of \( \text{sine} \) is realizable [68].

In principle, we can reconstruct the Fourier transform of an object wherever the spread function transform does not vanish. In practice, however, there is a cutoff frequency \( \omega_0 \) defining a forbidden region in which the transform does indeed vanish. The reconstruction based on the truncated MTF is band limited and, at best, imperfect. The use of
substitute processors by definition gives a deficient but nevertheless acceptable picture. When such methods of compensation fail, more sophisticated methods must be used.

It has been shown that perfect digital reconstruction of two-dimensional objects can be accomplished in the absence of noise, with two provisions: (1) that the object is limited in spatial extent, and (2) that experimental knowledge of the image and the PSF is available [12, 45, 196, 197]. Under these conditions, the integral imaging equation can be solved for the object in terms of the image. The solution is expressed as a series expansion in the eigenfunctions and eigenvalues of the system spread function. The eigenfunctions can be interpreted as objects that are only scaled and translated by the imaging system. The expansion can be performed in the object domain or in the frequency domain.

In both cases, certain higher transcendental functions are involved; these are the radial and angular prolate spheroidal wave functions. In the object domain

\[ o(x) = \sum_{n=0}^{\infty} \int_{-a}^{a} (w_n(x) v_n(\xi) / \lambda_n) \delta(\xi) \, d\xi \]

where

\[ \int_{-a}^{a} (w_n(x) v_n(x) / \lambda_n) \, dx = \delta(x-a) \]

and in the frequency domain

\[ O(w) = \sum_{n=0}^{\infty} a_n b_n(w) \]

\[ a_n = \int_{-a}^{a} O(w) \phi_n(w) \, dw \]

\[ \lambda_n = \frac{1}{\sqrt{a_n B_n}} \left[ a_{n+1} / (B_n 1/2) \right] \]

\[ \phi_n(w) = \sqrt{N_s N_w / (B_n a_1 / 2)} \]

\[ N_s = \int_{-\infty}^{\infty} \left( k_n (\omega / a_1) \right) ^2 \, d\omega \]

This approach has been used for restoration in diffraction limited imaging systems, with the sinc as the kernel, and also for resolving
double stars. In theory, it solves the reconstruction problem for any object whose image is described by a convolution.

Techniques based on the statistical theory of signal prediction have been developed to remedy degradation which consists of both blur and spatially stationary stochastic noise superimposed before and after imaging. The solution of the imaging equation is estimated statistically rather than determined analytically by applying an "optimal" linear filter to the picture. One kind of optimal filter minimizes the mean square error of reconstruction between the restored picture and the undistorted but unseen original [74, 105, 233]; another is a filter which maximizes the signal to noise ratio [46, 154]. One can also seek a Bayesian estimate of the object, i.e., the estimate which maximizes its posterior probability density. The PSF itself may also be considered as stochastic; it may be randomly chosen from a population of PSFs which are independent of the additive noise. Methods such as these are pertinent to restoration of quantum limited images such as isolate scans and images degraded by atmospheric turbulence (also see [226]).

Further discussion of resolving power and restoration will be found in [46, 66, 67, 115, 135, 174, 198, 235].

ENHANCEMENT

Whereas restorative enhancement is the art of making things seem what they are, non-restorative enhancement is the art of making things seem what they aren't. An optimally or even fully restored image may not be the most efficient form for visual data which are to undergo further processing and interpretation by man or machine. Information is retained in an unselective and unorganized way, independent of the semantic content and complexity of the scene. Furthermore, imperfect restoration may leave the image adulterated by residual noise and pseudo-objects—points, curves and even regions. Distorted or random noise usually takes the form of artifacts or deletions of signal over relatively small picture areas. With no obvious pattern. Organized noise displays a higher degree of spatial correlation, and takes the form of structured artifacts and systematic biases. Moreover, perceptual artifacts occur because of spatial and tonal quantization; these include artifacts of segmentation and coalescence ("false" contours and "false" regions).

Thus, whether people or machines do the final image processing, non-restorative preprocessing may be a worthwhile and even necessary intermediate step. No enhancement technique, however, can add to the amount of information in the original scene. On the contrary, by selective suppression and accentuation, enhancement re-presented, and information is inevitably lost. It is obviously important to make optimal use of prior information in deciding what to retain.

In general, it is desirable to deemphasize or discard irrelevant material and at the same time to emphasize or clarify features and objects of interest. Moreover, it is desirable to do this without having to identify or recognize specific objects, insofar as possible. Deemphasis is usually achieved by using smoothing or integrating operators, but these tend also to blur detail. Emphasis is usually accomplished by sharpening or differentiating operators, or by feature enhancers or contrast enhancers, but these also tend to accentuate noise. However, smoothing without obliteration of the relevant and sharpening without amplification of the irrelevant are the desired ends. Thus a judicious selection of techniques is required to balance these counter-tendencies and give optimal enhancement.

Enhancement operations can be classified as context-free or context-sensitive. Context-free enhancement consists in repeated and impartial use of a position invariant (and sometimes rotation invariant) procedure, usually determined a priori. Context-sensitive enhancement differs in that decision rules are used to determine the applicability of the various processors and to assign suitable values to processor parameters, in response to local image characteristics. (This is analogous to the use of priority controllers in mechanical languages, for deciding applicability and sequencing of grammatical transformations.) Context-sensitive enhancement is adaptive in the sense that the precise course of processing depends on accumulated experience with the given picture or similar pictures. Although simple, high-contrast or binary material can often be handled using context-free methods, pictures with high information content and noise generally require context-sensitivity.

Enhancement of regions and contours can be described in spatial or spectral terms, but the former predominates. The mathematics of enhancement differs little from the mathematics of restoration, and involves similar difficulties. While restoration deals with entire spectra and
The primitive enhancement is a smaller mask consistent with this sampling lattice.

\[
W_\alpha = (n_0 x m_0 n, n_0 = 0, 4, 8, \ldots), \quad \alpha = 1, 2, \ldots, N
\]

No explicit provision has been made here for transforming border effects by convolution with \( W \) with \( \alpha = 1, 2, \ldots, N \) or \( m_0, n_0 = -m_0, n_0 

\[
T_\alpha = \sum_{m, n = -\infty}^{\infty} W(m, n) \cdot B(m, n) = (m_0, n_0) \cdot (n_0, n_0 = 0, 4, 8, \ldots)
\]

Convolution can be interpreted in various ways, as discussed in the following sections.

For purposes of exposition it is convenient to organize the discussion as follows:

1. Spatial smoothing of regions.
2. Spatial smoothing and sharpening of edges.
3. Edge sharpening by spatial manipulation (frequency filtering).
4. Contrast enhancement by manipulation of the gray scale and brightness controls.

Smoothing of regions

Most primitive enhancement operations use convolution (spatial mask) as a rectangular sampling lattice and are represented by matrix notation.

Unless otherwise indicated, we assume digitalized image, defined by a rectangular sampling lattice and represented by matrix notation.
about continuity, whereas contour enhancement involves algorithms which are concerned with the topology of curves or outlines. Enhancement methods of both types are closely related to edge detection and which are related to image presentation. The most effective non-iterative methods use some type of spatial differentiation. The directional derivative of $p(x,y)$ is a point by coordinate system is defined by $\frac{\partial p}{\partial x}$ and $\frac{\partial p}{\partial y}$.

If the partial derivatives of $p$ at any two orthogonal directions are computed, the gradient of $p$, $\nabla p$, is defined as

$$ \nabla p = \left( \frac{\partial p}{\partial x}, \frac{\partial p}{\partial y} \right) $$

and the gradient magnitude is

$$ |\nabla p| = \sqrt{\left( \frac{\partial p}{\partial x} \right)^2 + \left( \frac{\partial p}{\partial y} \right)^2} $$

The gradient magnitude indicates the rate of change of magnitude to that rate. This distributed direction that direction contains

$$ \frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial y^2} $$

Other useful combinations of derivatives include the Laplacian operator, a scalar isotropic extension of the second derivative:

$$ \nabla^2 p = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} $$

and the bi-Laplacian

$$ \nabla^4 p = \frac{\partial^4 p}{\partial x^4} + 2 \frac{\partial^4 p}{\partial x^2 \partial y^2} + \frac{\partial^4 p}{\partial y^4} $$

The nth order generalized derivative is defined in terms of the partial derivative operators $\delta(x)$ and $\delta(y)$ as follows:

$$ \nabla^n p = \sum_{k=0}^{n} \binom{n}{k} (\frac{\partial}{\partial x})^k (\frac{\partial}{\partial y})^{n-k} p(x,y) $$

The gradient $\nabla p$ uses 2D masks, yielding

\[ \nabla p = \begin{bmatrix} P_{x} & P_{y} \end{bmatrix} \]

\[ P_{x} = P_{01}x + P_{11}y + P_{00} \]

\[ P_{y} = P_{01}x + P_{11}y + P_{00} \]

The gradient magnitude is used in preference to the more exact
The masks can be written symbolically as

\[
\begin{bmatrix}
-1 & 0 \\
1 & 0 \\
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
-1 & 1 \\
0 & 0 \\
\end{bmatrix}
\]

A more precise estimate is obtained by fitting a quadratic surface over a 3x3 neighborhood by least squares, and then computing the gradient for the fitted surface. The masks are

\[
\begin{bmatrix}
1 & 0 & -1 \\
0 & 0 & -1 \\
1 & 0 & -1
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
1 & 1 & 1 \\
0 & 0 & 0 \\
-1 & -1 & -1
\end{bmatrix}
\]

Fig. 6 has been spatially differentiated in this way, using these masks to approximate the true gradient and displaying the gradient magnitudes as gray values. The gradient direction at each point has direction cosines \((\hat{p}_x \mid \hat{v}_x), (\hat{p}_y \mid \hat{v}_y)\).

Information is lost in the preceding computation because the value at the center is not used. Better resolution is obtained by interpolating the image at points midway between nearest neighbors in the sampling lattice, and computing gradients at the inserted points. With a 4x4 neighborhood to fit a quadratic surface with six parameters, the following results are obtained:

<table>
<thead>
<tr>
<th>Sampling lattice</th>
<th>Interpolated picture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3 2 2 -3</td>
</tr>
<tr>
<td></td>
<td>2 7 7 2</td>
</tr>
<tr>
<td></td>
<td>2 7 7 2</td>
</tr>
<tr>
<td></td>
<td>-3 2 2 -3</td>
</tr>
</tbody>
</table>

* Sampled points
* Interpolated point

\[
\begin{array}{ccc}
\text{x-derivative} & \text{y-derivative} & \text{Laplacian} \\
-3 & -3 & -3 & -3 & -1 & 1 & 3 & 1 & 0 & 0 & 1 \\
-1 & -1 & -1 & -3 & -1 & 1 & 3 & 0 & -1 & -1 & 0 \\
1 & 1 & 1 & -3 & -1 & 1 & 3 & 0 & -1 & -1 & 0 \\
3 & 3 & 3 & -3 & -1 & 1 & 3 & 1 & 0 & 0 & 1 \\
\end{array}
\]

Gradients can also be found approximately by using a set of oriented edge detectors and searching sequentially at each point for the best match. Gradient magnitude is equated with the maximum response and
enhancement and extraction

direction is taken parallel to the orientation of the corresponding detector. Such a set is given by the eight masks
\[
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
-1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\
-1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 \\
-1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 \\
\end{array}
\]

Fig. 7a was obtained by applying this algorithm to a 256-level image of a white blood cell. Gradient values below 11 were suppressed for clarity. In Fig. 7b, the symbols \( - \), \( \), \( / \), and \( \) were used to encode isolevel directions at points with above threshold gradients.

For digital Laplacians, both cross and square shaped neighborhoods have been used, the weights also always summing to zero. Only simple convolution with the picture is needed. Examples are

\[
\begin{array}{cccccccc}
0 & -1 & 0 & -1 & -1 & -1 & -1 & 0 \\
-1 & 4 & -1 & -1 & 0 & -1 & -1 & 0 \\
0 & -1 & 0 & -1 & -1 & -1 & -1 & 0 \\
\end{array}
\]


The main weakness of the preceding technique is their treatment of signal and noise, resulting in overemphasis of accidental fluctuations. This can be controlled in two ways [56]. First, contour enhancement without noise enhancement can be realized by replacing the Laplacian in the crispening operation with the second derivative in the direction of the local gradient. The modified transformation is

\[
T(x,y) = p(x,y) - \frac{\partial^2 p}{\partial x \partial y} - \gamma \frac{\partial^2 p}{\partial x^2}
\]

Second, contour enhancement along the gradient and smoothing along the contour can be realized by replacing the Laplacian with a mixture of second derivatives normal (direction s) and tangential (direction t) to the contour:

\[
\frac{\partial^2 p}{\partial x \partial y} + \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}
\]
ENHANCEMENT AND EXTRACTION

\[ T_{p} = p^{-3} \left( \frac{2p^2}{2p^2 - 13p} \right) \]

\[ \frac{8p}{q} v_{p} = \frac{3p}{q} v_{p} + 2p_{q} + p_{q}^{2} \]

\[ p_{q} + p_{q}^{2} \]

At least nine-point interpolation is used. Using a 3 x 3 square neighborhood of half-width d or an octagonal neighborhood of radius d

\[ P_{-1-1} P_{-10} P_{-11} \]

\[ P_{-1} P_{0} P_{1} \]

\[ P_{1-1} P_{10} P_{11} \]

\[ P_{2} = \left( P_{1} - p_{10} \right) / 2d \]

\[ P_{3} = \left( P_{2} - p_{21} \right) / 2d \]

\[ P_{4} = \frac{2p_{2} - \left( P_{3} - p_{31} \right)}{2d} \]

\[ P_{5} = \frac{2p_{3} - \left( P_{4} - p_{41} \right)}{2d} \]

Similarly, lines and edges can be emphasized in selected directions and at the same time suppressed in others. Edge enhancers are masks partitioned by a line of like slope, with entries of opposite sign on opposite sides of the divide. The line enhancer consists of a like oriented ridge of 1's on a field of 0's. A typical set of 7 x 7 masks which can be used for line enhancement, and as we shall see, for detection also, is \{42, 70, 137, 148\}:

```
x x x x x x x
x x x x x x x
x x x x x x x
x x x x x x x
x x x x x x x
x x x x x x x
```
Features and shapes can be enhanced similarly, although several templates of different scale may be needed in order to span the size range. For example, corner, intersection and spot enhancers would take the following general forms:

\[ \begin{array}{cccc}
- & + & - & - \\
- & - & + & - \\
- & + & - & - \\
- & - & + & - \\
- & - & + & - \\
\end{array} \]

These enhancers are closely related to the optimum templates for certain patterns (see the next section).

Frequency filtering

Two-dimensional spectral analysis is the only systematic approach now available for describing the effects of spatial masks in terms of resolution and noise. Since the system MTF and PTF describe the rendition of detail, masks can be classified according to the action of the corresponding frequency filters. A qualitative classification into (1) low pass, (2) high pass, (3) low emphasis, (4) high emphasis, and (5) band pass filters is convenient. Masks should be symmetric about the central element if isotropy is desired, and asymmetric if directional information is to be emphasized. Most imaging systems preserve low frequencies well but respond increasingly poorly to high frequencies (Fig. 8a); thus, global information is usually preserved intact, but detail is compromised.

The main function of the low pass filter (Fig. 8b) is to retain low frequency information and reject high frequencies. Thus, low pass filtering is related to smoothing attributable to random noise. The challenge is to select a cutoff frequency \( \omega_c \) so that the filter is narrow enough to remove noise, but wide enough to preserve detail; or, in the spatial domain, to select a mask smaller than the smallest detail to be retained.

The purpose of the high pass filter is to remove low frequencies corresponding to background and to enhance contrast. Thus, high pass
filtering is related to sharpening. A spatial mask used for this purpose must be larger than the largest differential feature to be retained.

If \( V = (v_{ij}) \mid 0 \leq | i |, | j | \leq K \) is a low pass mask, then

\[ W = (w_{ij}) = \delta(k) \delta(l) - v_{ij} 0 \leq | k |, | l | \leq L \]

is a high pass filter. A band pass filter retains frequencies between two limits and is equivalent to a low pass, high pass pair.

High and low emphasis filters try to compensate for spectral losses and to stretch contrast, but not at the risk of disregarding useful information. The simplest low emphasis mask is an equal weight mask. Emphasis of high frequencies can be achieved by superimposing a scaled high pass transform \( v \) on the original image \( p \):

\[ t = p + rv \]

where \( v \) is adjusted according to some external criterion. The Laplacian mask can be regarded as a scaled equal weight mask superimposed on the original image; the result is a high emphasis transform.

It is straightforward to determine the effect of a given mask on the frequency spectrum. The image sampled at interval \( d \)

\[ p(x,y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} i(\text{md,nd}) \delta(x-\text{md}) \delta(y-\text{nd}) \]

has a Fourier series for its Fourier transforms:

\[ \hat{F}(u,v) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{i}(u,n) \delta(u-\text{md}) \delta(v-\text{nd}) \]

The Fourier transform of the sampled convolved image

\[ t(m,n) = \sum_{x} \sum_{y} v(x,y) \delta(m-x) \delta(n-y) \]

satisfies a convolution-product theorem

\[ \tilde{F}(m,n) = \hat{W}(m,n) \tilde{F}(m,n) \]

where the mask transform

\[ \tilde{W}(m,n) = \sum_{\text{md}} \sum_{\text{nd}} \hat{i}(\text{md},\text{nd}) e^{-j2\pi(m,d+n,d)} \]

completely characterizes the processing. For example, the equal weight mask corresponds to a low pass filter, since its transform consists of equal amplitude sinusoids which tend to interfere; the larger the mask, the more restrictive the filter.

The problem of finding a mask corresponding to a specified frequency filter is more difficult. If the filter is band pass (zero outside a finite frequency range), then the exact mask for \( \tilde{W}(u,v) \),

\[ \tilde{W}(u,v) = \int_{-\omega}^{\omega} e^{j\omega u} \delta(v) d\omega \]

is infinite and unrealizable, so a finite approximation \( \tilde{W}(u,v) \) which is optimal in some sense is desired. Possible figures of merit which measure the closeness of fit of \( W \) and \( W \) are minimal integrated mean square error and minimal maximum absolute error, both over the frequency domain [7, 8, 14, 15, 30, 38, 47, 91, 11, 113, 147, 151, 153, 216, 221, 234].

Manipulations of the gray scale and the sampling lattice

Selective contrast enhancement and prevention of false contours can often be obtained by direct manipulation of the gray scale, or by spatial and grayscale quantization. Most quantization schemes involve a fixed number of gray levels which are apportioned uniformly over the dynamic range of the system. In addition to the possibilities of creating spurious contours and obscuring genuine ones, this method of recording information is wasteful in an information theoretic sense. If the grayscale statistics of input images can be determined, it is possible to quantize optimally [57, 188, 223] by reconcentrating the levels in the most informative part of the scale. Failing in this, refinement of the scale (increased number of gray levels) may be necessary.

A similar distortion of the gray scale is effected by the type of contrast enhancement referred to as gamma correction. Although gamma correction originally referred to gray scale and contrast adjustments intended to compensate for uneven or nonlinear responses of the image sensor, the term can be generalized to include any continuous transformation of the gray scale. The selected function differentially stretches portions of the gray scale and compresses others; signals are suppressed in low-slope portions of the curve and expanded in high-slope portions. The function of Fig. 9a clips extreme dark and light values and enhances the renditions of the gray values. A variant of this uses continuous functions as in Fig. 9b. The same set of gray values is used to
Enhancement of contrast and curve

The enhancement of contrast and curve is achieved through a nonlinear transformation. The transformation is defined as:

\[ g(x) = a \cdot f(x)^b + c \]

where:
- \( g(x) \) is the transformed gray value
- \( f(x) \) is the original gray value
- \( a, b, c \) are adjustable parameters

Figure 9 shows the effect of this transformation on the gray scale.

(a) Original gray scale
(b) Enhanced gray scale

Conversely, the enhancement of the curve can be achieved by introducing a linear transformation:

\[ y = mx + c \]

where:
- \( y \) is the transformed value
- \( m, c \) are adjustable parameters

This transformation is useful when the curve needs to be stretched or compressed to fit a specific range.

Finally, linear effects may be generated by using dynamic-range compression, but these distortions are usually unacceptable, and therefore, a better approach is to use histogram equalization or adaptive thresholding techniques.
gramming [52] or non-linear programming [131, 132]. Fourier series can be used to analyze and synthesize contours; a contour can be smoothed by low pass frequency filtering.

Elongated forms or thick contours are usually extracted by clipping or thresholding the original picture or transformations of it. Contours can also be extracted using derivatives, as illustrated in Fig. 6. The resultant contours in either case are often discontinuous, thick, or jagged. The problem is to enhance them—to provide good continuation and closure by filling gaps and eliminating perturbations, and by thinning.

Thinning is the generic name for processes designed to reduce connected elongated objects to line-like representations which preserve connectivity. Many algorithms based on an erosive or stripping method have been suggested [90, 99, 114, 126, 140, 187]. The common procedure involves removing boundary points layer by layer, provided that their removal does not erode away ends of arcs and does not destroy connectivity, until the configuration stabilizes. The procedure is noise sensitive in binary pictures and depends critically on the method used to generate the original thick contour. This difficulty can be remedied to some degree by using skeletonization procedures [18, 21, 159, 185] for multi-level pictures which avoid the use of binary intermediates. In this connection, preprocessing with the gray-weighted distance transform [185, 187] gives additional smoothing.

Ensuring that outlines have good continuation clearly requires decision rules utilizing context and feedback. Filling of holes (removing "salt noise") and removing spurious connections (removing "pepper noise") are both necessary. Look-ahead procedures predict a position for the next outline point, and trigger a local search; the predicted position is an extrapolation based on assumptions or acquired information about local slope and curvature [62, 79, 83, 194, 195, 269, 353]. Sequential tracking and smoothing can be combined for filling gaps, and separating accidental joins as well. Let \((x_k, y_k), (x_{k+1}, y_{k+1})\), and \((x_{k+2}, y_{k+2})\) represent the coordinates of the \(k\)th sample on the unprocessed, smoothed and thinned tracks, respectively. A smoothed track can be obtained [58] by setting

\[
x_k = \frac{3}{2} x_{k-1} + \frac{1}{2} x_k \\
y_k = \frac{3}{2} y_{k-1} + \frac{1}{2} y_k
\]

and the \(k\)th point on the thinned track by

\[
x_k = x_k \quad \text{if } |x_k - x_{k-1}| < \text{some } \theta
\]

and similarly for \(y\).

A variant [128] uses

\[
x_k = x_k - \theta \quad \text{if } |x_k - x_{k-2}| > \text{some } \theta
\]

and similarly for \(y\).

Spurs can be removed by using digital "calipers" to detect and track sharp bends simultaneously [137]. The caliper tips move around the outer boundary in such a way that the arc length between them remains constant. A sharp decrease in the chord length indicates the presence of a spur.

Outlines can also be obtained from thick connected contours by tracking. A digital "probe" is laid across the thick line, and the midpoint of the intersection is determined. As the procedure is repeated at different positions along the line, the midline track is generated [105]. The procedure as reported in the literature is limited to horizontal and vertical probes. If gradients are computed at the boundary, however, orthogonal probes can be generated, but a probe orthogonal to one edge may not be orthogonal to the opposite edge.

OBJECT EXTRACTION

To borrow a term recently introduced in systems theory [230], a pictorial object is a fuzzy set which is specified by some membership function defined on all picture points. From this point of view, each image point participates in many memberships. Some of this uncertainty is due to degradation, but some of it is inherent. The role of object extraction in machine processing, like the role of figure/ground discrimination in visual perception, is uncertainty-reducing and organizational. In fuzzy set terminology, making figure/ground distinctions is equivalent to transforming from membership functions to characteristic functions.

We distinguish two types of processing that are commonly subsumed under the single heading of object extraction: detection and articulation.
Detection is concerned with making decisions about the presence or absence of specific object signals and estimating their position in the field of view. Detection in this sense involves finding optimum matches among a finite set of possibilities. Articulation is concerned with a more detailed structural analysis, in which the picture is punctuated by object delimiters or boundaries, much akin to the insertion of phrase markers in grammatical analysis of sentences. Potential objects, not presupposed, are sought. The parsing defines the final objects and their relationships.

For solving some imaging processing problems, it may be possible to combine detection and articulation in a multilevel extraction scheme. The detection algorithms would function as coarse scanners, locating potential areas of interest, and the articulation algorithms would function as fine scanners, delineating objects and preparing them for later analysis.

**Detection**

Detection refers to procedures for locating objects which are pre-specified in terms of either a local spatial distribution of gray values (local shape or template) or a transform of such a distribution (filter). The templates or filters can be selected intuitively [168], by self-design [82, 180, 181], by random generation and systematic evaluation [18, 217], by adaptive learning [17, 84, 78, 98, 97, 149, 178] from sets of training images, or by deliberate design according to the principles of statistical communication theory [107, 135, 133, 179]. Spatial correlation, optimum frequency filtering, and other transformations such as moment and orthogonal expansions are the main mathematical tools for object detection.

In a typical application, a picture or some transform is first compared with a library of standardized or representative features and objects or their appropriate transforms. For each possible feature, a figure of merit is computed at each picture point, yielding in effect a new picture transformation which indicates degree of match for that feature throughout the original picture. Using some criterion of significance, each of these transformations can be thresholded and converted to a binary map locating particular instances of the features in the picture. In addition, the various figures of merit at each picture point can be rank ordered.

**ENHANCEMENT AND EXTRACTION**

Cross correlation has been used as a figure of merit almost exclusively. The normalized cross correlation or correlogram of a pattern or template (figure) with a picture (ground) is

$$C(\xi, \eta) = \int \int f(u,v)g(u^\prime + \xi, v^\prime + \eta) \, du \, dv$$

$$\sqrt{\int \int f(u,v) \, du \, dv \int \int g(u,v) \, du \, dv}$$

which is 1 only when the picture is a scaled translation of the pattern. The correlogram values are high near positions of close match and are sharply attenuated elsewhere, the peak height being a measure of degree of match. However, numerical results are sensitive to changes in orientation, magnification, mean gray value and contrast. At the expense of making the correlogram multidimensional and unwieldy, invariance to changes in these factors can be introduced if desired. It is possible to eliminate gray value and contrast sensitivity by dealing with derivatives (see [137] for a generalization).

It is sometimes useful to detect significant change rather than significant match, without locating the site of this change. The mean absolute and mean squared differences of two pictures have been used as overall or gross measures of disparity [32, 63, 175]. Random generation of templates and filters which are subjected to subsequent evaluation has been attempted for binary pictures. While the same idea might be extended to grayscale pictures, the combinatorics involved and other intrinsic limitations make this approach unattractive.

Learning machines have perhaps been the most widely used methods of template construction [144, 189]. Template weights are adjusted as a sequence of typical features and their identifications are presented, until convergence or stability of the weights is observed. Alternatively, all the information in a sample of representative objects is used to determine optimum weights in one step. These methods do not directly take into account spatial correlations in the pattern, but templates invariably look like the objects they are supposed to represent. Although these approaches have been very successful in character recognition and other high contrast cases, they give little understanding of image structure.

More systematic methods of construction use optimality principles from statistical communication theory. The image is regarded as pattern plus pattern-independent noise with a known statistical structure:

$$l(x,y) = o(x,y) + n(x,y)$$
and a filter for the pattern is sought. The filter specifications depend
on the cross correlation of signal (i.e., pattern) with noise, and the
autocorrelations of signal and of noise. A common type of optimal filter
is based on minimizing mean square differences.

More widely used for picture processing is the matched filter, which
maximizes the signal to noise ratio. The classical matched filter re-
sponse maximizes signal energy to noise energy; it detects and locates
all translates of the pattern. If \( O^*(u,v) \) denotes the conjugate Fourier
transform of the pattern \( o(x,y) \) and \( N(u,v) \) denotes the power spectrum of
the noise, then the matched filter \( F \) is given by

\[
F(u,v) = \frac{O^*(u,v)}{N(u,v)}
\]

This concept can be generalized to gradient-matched filters which
detect on the basis of edges [4]. Such a filter of order \( p \) maximizes
energy in the \( p \)th order signal gradient to energy in the \( p \)th order
noise gradient:

\[
F_p(u,v) = (-1)^p \frac{(u^2+v^2)^p \Omega^*(u,v)}{N^p(u,v)}
\]

where \( N(u,v) \) is the power spectrum of the \( p \)th order gradient of the noise.
The generalized matched filters of the three lowest orders correspond
to three of the filters discussed in the previous section: \( p=0 \) to an
energy filter (a low pass filter); \( p=1 \) to the gradient, and \( p=2 \) to the
Laplacian (high pass filters). Generalized filters give stronger matches
than ordinary matched filters when edges rather than energy describe
the signal.

The spatial output of the filter is

\[
C_p(x,y) = \left( \nabla \psi_0(-x,-y) \right) * \left( \nabla \psi(x,y) \right) * (R_{-p}(\tau,T))^{-1}
\]

where \( R_{-p}(\tau,T) \) is the autocorrelation function of the \( p \)th order noise
gradient and \( \nabla \psi \) is the \( p \)th order gradient operator. The weights of the
corresponding mask resemble the object or object outline to be de-
tected. At exact registration, the filter output in the absence of noise is

\[
C_p(0,0) = \int \int (-1)^p |O(u,v)|^p (u^2+v^2)^p dudv
\]

124

ENHANCEMENT AND EXTRACTION

and in the presence of noise it is

\[
C_p(0,0) = \int \int (-1)^p (u^2+v^2)^p Q^*(u,v) t^p(u,v) dudv
\]

The correlation parameter \( C_0 \) thus measures the efficiency of
the match procedure.

The optimal spatial mask for a feature can be shown to be dependent
on not only its spatial pattern, but also on its internal spatial correlations
[5, 231]. This is true irrespective of the noise structure. For example,
consider a stochastic maxd signal pattern with expected value \( P^m(P_a) \)
and with covariance matrix \( \Sigma \) which consists of an mnx array of nnn
matrices \( S_0 \), where \( S_0 \) is the cross correlation matrix of the \( i \)th and \( j \)th
rows of \( P \):

\[
S_{i0} = \{ (P_{i,-1} - E(P_{i,-1})) \} \{ (P_{i,-1} - E(P_{i,-1})) \}^T
\]

The energy signal to noise ratio for a template \( W=(w_{ij}) \) is

\[
(S/N)^2 = \left( E(C(0,0)) \right)^2 / \text{var}(C(u,v))
\]

where \( C(u,v) \) is the discrete convolution

\[
W**P(u,v) = \Sigma_{i,j} w_{ij} (o(x,y))
\]

\( E(C(0,0)) \) denotes its expected value at perfect registration of \( W \) and
\( P \), and \( \text{var}(C(u,v)) \) is the variance of the cross correlation when the
template is shifted over the pattern. Since \( E(C(0,0)) = \langle W**P \rangle(0,0) \) and
\( \text{var}(C(u,v)) = \Sigma_{i,j} (w_{ij} P_{i,j}) \), the discrete optimal mask is \( F=\Sigma^{-1}P \).

Thus the optimal mask \( F \) for a random or uncorrelated pattern is
the pattern itself, but for any nontrivial covariance structure, clearly
it is not. For instance, if the correlation along a row tapers with dis-
cance according to the following autocorrelation matrix

\[
\Sigma_{ij} = \begin{cases} 1 & i=j \cr \rho^{|i-j|} & \text{if } i \neq j \end{cases}
\]

and if the cross correlation of the \( i \)th and \( j \)th row tapers according to

\[
\Sigma_{ij} = \begin{cases} 1 & i=j \cr \rho^{|i-j|} & \text{if } i \neq j \end{cases}
\]

then the optimal mask is

\[
F = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots \ \\
-\rho & 1 + \rho^2 & \cdots & \cdots \\
\rho^3 & -\rho(1+\rho^2) & 1 + \rho^2 & \cdots \\
\end{bmatrix}^{**P}
\]

125
"sloppy" binary pictures. There is no generally accepted criterion for assigning a "best" edge location. In grayscale pictures, traces or profiles of "true" edges can have a variety of shapes, depending on the spread functions involved (see Fig. 9). The position of the maximum gradient, which is usually, but not necessarily, found near the center of the edge trace, is the most attractive candidate for defining a unique edge location, but neither it nor the average gradient value over the trace determines how sharp the edge appears to an observer [41, 76, 252].

After restoration or enhancement, however, it would be expected that the maximal gradient should correlate better with the sensation of sharpness. In binary pictures, the issue is rather different: how best to segment a pattern of dots.

Some authors have grouped boundary assignment techniques into parallel vs. sequential. This categorization may not be as illuminating as the observation that some algorithms incorporate a good continuation or closure principle and others do not. Since the fundamental problem is to find connected regions of consistent flux which separate connected regions of constancy, relative to some surface function, a classification into homogeneity-emphasizing and heterogeneity-emphasizing techniques would seem to be more appropriate. Methods of the first type utilize the most redundant parts of the image, while methods of the second type utilize the most informative parts. But even this distinction is more a reflection on historical development than on underlying principles.

Thresholding is the oldest and most widely used method of extraction. It has been used on pictures p(x,y) and on picture transformations T(p).

In its most general form, thresholding is a quantization-coarsening transformation T of the form

\[ T[p(x,y)] = \begin{cases} 1, & \text{if} \, p(x,y) \geq t_1; \\ 0, & \text{otherwise} \end{cases} \]

where \( t_1 \) and \( t_2 \) are fixed constants. When thresholding is applied to the picture itself, the presumption is that regions of interest display different, fairly constant gray values. When thresholding is applied to a spatially differentiated picture, e.g., as a clipping operation \( (t = \infty) \), contouring occurs; i.e., the continuous image is replaced by a series of outlines or contours. Although thresholding is very effective on many types of pictorial material, it has noteworthy shortcomings: (1) smoothing and false contouring are byproducts; (2) there is no guarantee that the resultant contours are connected or closed; and (3) the results are very sensitive to selection of thresholds [119, 142].

Early use of thresholding involved human judgment. However, for moderate to high contrast pictures, machine determination can be based on inspection of a histogram of gray values. One semi-automatic solution used the gray value corresponding to the upper 95% of the histogram as threshold, where \( p \) is externally supplied [30]. Since suitable thresholds differ from picture to picture and even within the same picture, the procedure is of limited value.

By taking advantage of the multimodal nature of the gray value histogram, it is often possible to select thresholds automatically [164].

The method is suitable for pictures in which objects appear on a fairly uniform ground, and objects occupy a greater proportion of image area than edges. Background and objects will generate histogram modes, and edges, comprising a range of intermediate gray values over a smaller area, correspond to antinodes. Fig. 11 shows a histogram for a 256-level image of a white blood cell, the machine selection of two thresholds, and the partitioning into background, nucleus, and cytoplasm that resulted.

This method has been generalized [184] to make use of the second order probabilities that given gray levels occur as neighbors of other gray levels. Clusters of gray levels that have similar sets of neighbors are detected, and thresholds are selected at the boundaries of these clusters.

It is possible to readjust a tentative threshold until the isophote which best coincides with the average maximal gradient is found. In Fig. 12a, the mean gray value and mean gradient are plotted as functions of distance inward and outward from a postulated object boundary. Signed distances were determined by a bilateral extension of the distance transform, positive distances being propagated inside the boundary and negative distances being propagated outside the boundary. Fig. 12b gives a gray scale rendition of the results of the distance algorithm. If the postulated boundary corresponds to the actual boundary, the plotted gradient values should peak at distance 0 (the line spread function is the derivative of the edge spread function). Differentiation, statistical differencing, and image subtraction are heterogeneity-seeking methods. Image subtraction refers to the com-
parison of versions of a scene made at different times, with the aim
of detecting significant changes [173]. The overall difference between
two pictures $p_1$ and $p_2$ can be measured by

$$\left| \int |p_1 - p_2| \right|$$

or

$$\int (p_1 - p_2)^2$$

Algorithms for computing directional derivatives have been discussed
above.

Statistical neighborhood comparisons also yield information on change
[33, 81, 83]. The gray value at a point $p(x,y)$ is compared to the mean
$\bar{p}(x,y) \pm \sigma(x,y)$ over an annular neighborhood $A(x,y)$, where $\sigma(x,y)$ is
the standard deviation $\sigma^2(x,y) = \int_{A(x,y)} (p - \bar{p})^2 \, dx \, dy$ or mean absolute
difference $\sigma_A(x,y) = \int_{A(x,y)} |p - \bar{p}| \, dx \, dy$ over the annulus, $\bar{p}(x,y)$ the
mean, and $\sigma$ a prespecified factor, usually in the range 0.3 to 1.5. The
quantity $(p - \bar{p})/\sigma$ is relatively low at an edge surrounded by other
edges, and high at isolated edges or isolated small regions which con-
trast strongly with their background. Related to these are methods of
border definition based on statistical tests of alternative hypotheses
about the cumulative gray value distributions of neighboring areas (see
below).

Clipping or thresholding may be used conjointly with any of the
differencing methods to identify zones of high gradient. Likewise, use

Figure 11. Machine selection of two thresholds for a white blood cell and its
nucleus (a) image of the cell, with 32 gray levels, (b) density histogram,
indicating machine-determined thresholds, (c) image partitioned into back-
ground ($\lambda$), nucleus ($W$), and cytoplasm plus nucleolar granularity ($\gamma$). Reprinted
from Mendelsohn et al [120].
Figure 12a: Edge trace and its derivative at the boundary of the nucleus of the white blood cell of Fig. 11. Mean gray value (above) and mean gradient magnitude (on facing page) plotted as functions of distance from the boundary. Positive distances correspond to the interior, negative distances to the exterior of the nucleus. * denotes mean density or mean gradient, while - denotes the number of image points at the corresponding distance. The graph of density is actually an edge trace, while the graph of gradient is the corresponding line spread function.

of normalized cross correlation or matched filtering, followed by thresholding, can extract object locations.

Good continuation and closure are subject-dependent, and fortuitous in the preceding approaches, but they are built-in features of tracking and propagation algorithms. Tracking procedures may be thought of as elementary assemblage or concatenation approaches. The simplest tracking procedures trace around an already-delineated continuous boundary (see above).

It is not difficult to incorporate a look-ahead feature for bridging gaps and disrupting meaningless junctions [82]. Search for a next point can be continued past a gap of a certain length or less; the trace over the gap can proceed in the same direction as the most recent trace around the boundary, matching both slope and curvature, or a fixed angular deviation may be allowed. The computation of gradients and the sequential tracking of the locus of their maxima is a contrast-invariant method.
of delineation, the contour of a blood cell nucleus in Fig. 13 was constructed by the following algorithm: The picture is searched for a sharp change in gray value. The gradient threshold for this coarse scan is obtained mechanically from a histogram of gradient magnitudes for the picture. A fine scan mode is then initiated. Two types of tracking motions are involved: (1) local search along the direction of the gradient for a larger gradient magnitude, using a look-around procedure; and (2) movement orthogonal to the maximum gradient to the next candidate point. As the trace proceeds, lists of nodes corresponding to path bifurcations are constructed for later processing [101], also see [56, 219].

An adaptive line follower which has some principles in common with this has been constructed by Guzman-Arenas [59]. The line follower

Figure 12b. Distances from the nuclear boundary shown in Fig. 11. The darkness of each point is proportional to distance to the nearest boundary point.

Figure 13. Outlining by gradient-tracking: (a) photomicrograph of a white blood cell, (b) tracked locus of maximal gradient magnitudes obtained by the algorithm described in the text (tracking imposes continuity on the contours).
Figure 14. Partitioning the image field by means of accumulation. (a) Thresholded version of the image field of Fig. 12, containing a blood cell and a nearby dark granule in the same field of view. Starting from four-image, 1962. (b) Segmentation of the image field based on 200 levels of gray in the background. (c) Segmentation of the image field based on related distance values.
ACKNOWLEDGMENTS

The author wishes to express appreciation to the National Institutes of Health Research Grant 1-AT-00107, and the University of Pittsburgh Medical Computer Science Program, under the direction of Dr. M. S. Prevett, for financial assistance. The following collection of papers will be cited in the text:

REFERENCES

JUDITH M. S. PREWITT


207. Swokowski, W., Concepts, J. W., A System for Demonstrating the Effects of Changing Background on Automatic Target Recognition, in PFR, 33-44.


