Philips Research Laboratory Brussels
Av. E. Van Becelaere 2, Box 8
B-1170 Brussels, Belgium

Working Document WD54
(comments welcome)

Fourier analysis, mathematical morphology, and vision

C. Ronse

Philips Research Laboratory Brussels
Avenue Van Becelaere 2, Box 8, B-1170 Brussels, Belgium

June 1989 — corrections September 1989

Abstract: Two opposite orientations in image analysis are given on the one hand by linear filtering, spectrometry, and Fourier analysis, and on the other hand by mathematical morphology, which emphasizes order relations and set-theoretical properties. The former derives its appeal from its wide application in the processing of sound signals, while the latter has been sucessfully used in the analysis of materials or in cytology. We make a fundamental study of issues at hand in the choice of such methodologies in image analysis and vision. We start by outlining the difference in purpose of vision and audition and its physical basis, the scattering of waves. We criticize Serra's arguments on this matter. Then we consider the general limitations of linear filtering methodologies and the unsuitability of phase-independent spectrometry. We propose a paradigm of concurrent processing and of sorting of information rather than a single sequence of processing modules with a controlled loss of information. Finally we analyse the domain of applicability of mathematical morphology to the visual process and suggest that it is restricted to certain types of tasks.

Keywords: Vision, audition, wave scattering, diffraction, Fourier analysis, linear systems, spectrometry, mathematical morphology, set-theoretical models, uncertainty relations, concurrent processing, neurophysiology, psychophysics.

1. Introduction

In [38] Pavlidis notes that 15 years of recent research on image analysis have brought only small progresses in practical applications of this discipline. One of the problems faced by both theorists and practicing engineers is the existence of a multitude of methodologies and techniques, without a general framework indicating the scope of each. There is no coherent theory as in classical physics for example.

Although Pavlidis gives much attention to the lack of availability of general software tools, it is also important to look at the way researchers develop methods and techniques and to see if they bring a better understanding of the problem they deal with. As Marr indicated [29], a widespread attitude is empiricism: try something (or even anything), and see what happens. Although it tends to avoid reflection on fundamental issues, empiricism does not exclude the use of a theoretical and mathematical apparatus. For example: Theory X deals succesfully with Problem A, let us try to adapt it to Problem B. However the two problems might differ in a more fundamental way than in some apparent mathematical parameters, and we have to understand the mathematical and physical assumptions underlying Theory X, and its conceptual link with Problem A, to see whether it suits also Problem B.

In this work we deal with low-level vision, and we analyse methodologies for transforming images and their possible role in the analysis of these images. Indeed, as pointed out by Serra in the introduction of [46], perception involves a transformation of the signal, which makes explicit some features which are only implicit. In order to analyse an image we have to distinguish the meaningful information from the irrelevant one. Therefore any method for image analysis uses transformations.

Two important classes of image transformations, which are in some way opposite, are on the one hand linear filtering and Fourier analysis, and on the other hand mathematical morphology. The first one has been extensively applied in the processing of sound signals (music and speech), and it relies on tools such as the Fourier transform and the study of its amplitude spectrum. The second one originated in the Paris School of Mines and has been applied sucessfully to the analysis of material properties of samples (ores, porous objects, cytological preparations, etc.) from images made by planar sections; it emphasizes set-theoretical concepts, order relations, and lattice-theoretic operations.

How far can such methods be used in image analysis? If we want to apply techniques of speech processing to image analysis, we have first to see how they relate to the goals of audition and vision. In Section 2 we explain that these two senses have different purposes: audition analyses sound waves, while vision recognizes the form and surface properties of objects, as well as their position, orientation, motion and mutual spatial relations in a scene, from the pattern of light waves that they reflect. We consider this distinction in relation to psychophysics, neurology, but also physics: the scattering of waves and the relation between wavelength and resolution in imaging. This problem is dealt with by Serra (see the introduction of [47] or of [49], and [48]) in a justification of the non-linear and lattice-theoretic techniques of mathematical morphology; we criticize his arguments in this respect.

In Section 3 we recall Gabor's uncertainty relations for linear filtering (which generalize such relations encountered in quantum mechanics) and outline some of their consequences.

Then we explain the inadequacy of spectrometric methods in image analysis, in particular because of the greater importance of phase than amplitude in the Fourier transform of an image, and point out several facts which seem to limit the role of linear methods in vision, in particular in the perception of texture.

In vision and image analysis, a methodology developed for extracting features in some types of pictures seems certainly more suitable than one for speech processing, and indeed actually mathematical morphology is more fashionable than Fourier analysis, but it should not be considered as a panacea. In particular we need to know its scope and limitations. In Section 4 we criticize Serra's suggestion of a single sequence of transformations and of controlled loss of information; we suggest rather a concurrent processing and a sorting of information, based on actual neurological knowledge concerning primate vision. In Section 5 we analyse the physical assumptions underlying morphological operators and propose that such operators are suitable when local configurations of grey-levels in an image represent directly material properties of objects pictured; we show that it is generally the case for many typical applications of mathematical morphoplogy. However for many aspects of visual analysis, in particular the inference of three-dimensional features and spatial relations from two-dimensional images, this is not the case, and indeed mathematical morphology has not been applied for such tasks and perhaps will never be.

2. The Different Purposes of Vision and Audition

It is an obvious fact that when we hear a sound, what we perceive is that sound, not the objects which emit or reflect it. We cannot 'hear' the shape of objects around us. The only spatial information we obtain from hearing is a rough estimation of the direction of the sound source (this is why we have stereo audio equipment), or the existence of a surface reflecting sound in the case of echo.

On the other hand, when we see an image or a movie, we perceive objects in three-dimensional space, their shape, colour, position, orientation, movement, etc. (see [8], Figure 1, for a more complete list of visual attributes of objects). It is not the intensity and spectrum of light waves which interest us. Indeed, Land and McCann [20] showed that we perceive the brightness of a surface not as the intensity of light that it reflects towards us, but as a function of the change of that light intensity between neighbouring surfaces. This allows us to perceive accurately the reflectance of objects under varying illumination conditions (the light reflected by a black object under bright sunshine has greater energy than that reflected by a white object at twilight, but we are not confused about their intrinsic colours). Similarly, the spectral hue of a surface is perceived as a function of the change of spectrum of light reflected by neighbouring surfaces [3]. Indeed, the spectrum of sunlight varies a great deal between morning and evening, and we can still perceive the colour of objects and distinguish hues which differ by much less than the daily variation of the sunlight spectrum. Thus our visual system can separate the reflectance of objects from their illumination and orientation. In particular it can use context to discriminate between edges in image which are due to a change of reflectance and a change in illumination/orientation (shading or shadow): see for example [12] (we will discuss this further in Section 5).

Although nobody amalgamates audition and vision on the ground that they 'analyse waves', it is nonetheless customary to group them into 'signal processing': sounds are one-dimensional temporal signals, while images are two-dimensional spatial signals. However one must question this analogy, because it hides what physical reality a signal represents. An image may represent a two-dimensional object or a three-dimensional scene. Given a grey-level gradient edge in a picture, it can be due to a change of reflectance (between a darker and a lighter portion of a surface), or to a change of illumination/orientation (between two surfaces forming different angles w.r.t. the source of light); this is illustrated for example in the first figure of [12], where the same edge can be interpreted in one way or the other thanks to the context. Similarly, speech and music are both sound signals, but they are not analysed according to the same criteria.

In quite general terms, one can distinguish sound and image signals by the fact that the image represents an object, while a sound represents itself. Thus the information provided by an image is often indirect. While the primitive for sound analysis is generally a local frequency, it seems that the primitive for image analysis can be the point. This idea is supported by sensory physiology. In the cochlea of the inner ear, each acoustic nerve fiber responds to a narrow frequency band of sound waves [45], leading to a *tonotopic* coding of sound at the entrance of the brain, where each position corresponds to a certain frequency; portions of the cortex devoted to low-level auditory processing have their neurons sensitive to local intensity and phase features in the frequency spectrum (see [18], Section 4.4). On the other hand, photoreceptors on the retina respond to light coming from a certain direction, and the arrangement of neurons in the optic nerve up to the primary visual cortex is *retinotopic*, in other words to each elementary unit ('column') in the visual cortex corresponds a small portion of the retina, whose receptors are connected to the neurons of that unit, and these neurons respond to features in the corresponding local portion of the visual space [10].

2.1. Object location and navigation by ultrasound

One will object to the above that ultrasound echography is used in medical imaging, and that several animal species (bats, dolphins) detect obstacles and orient themselves in their environment thanks to some kind of sonar (like the one of submarines). We also mentioned above the human stereophonic ability of roughly locating the origin of sound.

In humans, mechanisms of spatial localisation of sound seem to be concentrated in primitive portions of the brain. For example in the mesencephalon (mid-brain), the inferior colliculus is devoted to sound processing, and it contains spatial maps whose neurons respond to stereophonic disparity of sounds, and some of them even detect changes in such a disparity, in other words spatial motion of sounds (see [18], Section 4.4). Above it, the superior colliculus contains maps of the visual space, and it controls unvoluntary eye movements. Correspondences between such visual and auditory representations of space allow us to automatically move our eyes in the direction of either a sound source, or a moving stimulus at the periphery of visual space.

At any rate, the tasks of navigation and object location performed by the auditory system of some animal species are small feats compared to the achievements of human vision. In particular, they require neither a precise description of the objects located nor a fine resolution in that location. They never approach the precision of our visual system, which detects a motion of less than one minute of arc per second of time, and perceives depth from stereopsis with binocular disparities of only a few seconds of arc.

In the next subsection, we will explain that such a precision cannot be achieved by an acoustical imaging process; it requires an optical one.

2.2. *The scattering and diffraction of waves*

The image of an object is obtained by measuring the interaction between it and a wave (light or sound). But the form of this interaction depends on the ratio between the object size and the wavelength. As it is explained at the beginning of Chapter 8 of [36]:

> "When a sound wave encounters an obstacle, some of the wave is deflected from its original course. It is usual to define the difference between the actual wave and the undisturbed wave, which would be present if the obstacle were not there, as the *scattered* wave. When a plane wave, for instance, strikes a body in its path, in addition to the undisturbed plane wave there is a scattered wave, spreading out from the obstacle in all directions, distorting and interfering with the plane wave. If the obstacle is very large compared with the wavelength (as it usually is for light waves and very seldom is for sound), half of this scattered wave spreads out more or less uniformly in all directions from the scatterer, and the other half is concentrated behind the obstacle in such a manner as to interfere destructively with the unchanged plane wave behind the obstacle, creating a sharp-edged shadow there. This is the case of geometrical optics; in this case the half of the scattered wave spreading out uniformly is called the *reflected* wave, and the half responsible for the shadow is called the *interfering* wave. If the obstacle is very small compared with the wavelength (as it often is for sound waves), all the scattered wave is propagated out in all directions, and there exists no sharp-edged shadow."

Note that this statement is true both for sound and light waves. (NB. Even if one takes into account the fact that light consists of particles (photons) rather than waves (see [9]), this remains true as long as the light is strong enough to behave as a classical wave.)

In optics, the scattering of light produces *diffraction*, that is an alternation of dark and bright bands near the edge of the shadow of the object. (NB. In acoustics, the scattering of sound waves also produces diffraction bands when the wavelength is very small compared with the size of the obstacle). The width of these bands depends on the wavelength. This puts a limit on the resolution of optical instruments (we follow [4], Chapter XI, Section IV). In such an instrument (telescope, microscope, or eye),the image of a luminous point is a small blob surrounded by diffraction rings which are generally unseen (since the energy in the bright bands is much smaller than in the central blob). When we have two neighbouring luminous points, they appear separated if the two central blobs are separated by at least one dark interval. This defines the *separating power of the instrument*, in other words its resolution: the smallest resolvable angular distance between two luminous points is such that the bright central blob of one of them coincides with the first dark band of the other (see [4], § 139). For both telescopes and microscopes (that is, viewing of objects at infinite or finite distance), this minimal resolvable distance is proportional to the wavelength (see [4], § 140-141).

This explains why acoustical imaging uses ultrasound rather than audible one (the greater the frequency, the smaller the wavelength), and why optical instruments have in general a finer resolution than acoustical ones, such as sonar (the wavelength range of visible light is about 400–700 nm, and for ultrasound in air, this would require a frequency around 500 MHz).

Note finally that the fact that wavelengths are very small compared to the size of obstacles is the physical basis of *geometrical optics*, which studies geometric properties of the propagation of light. Indeed, in this case waves can be considered as propagating along straight lines, contrarily to sound waves which 'turn around' obstacles. On the other hand *geometrical acoustics* is possible only for very high frequency ultrasound.

2.3. Serra's argument

The above discussion on physiological and physical aspects of acoustics and optics is relevant to an analysis of the methodological basis of mathematical morphology. Indeed in [47], p. 2, we read (translated by us): "To those who, later, will study the history of sciences in the XXth century, mathematical morphology will appear as a branch of optics, among others, as radiology and tomography". Serra [47,48] considers optics as the study of vision and light, and proposes to integrate the work of his team in the framework of what he calls 'morphological optics'.

One of his basic arguments deals with the concept of linearity and its applicability in signal processing (we follow Section 2 of [48], which is reproduced in [49], pp. 2-3). After stressing the wide applicability of linearity in geometrical optics, for example in the description of the behaviour of optical lenses, he also considers its relevance to accoustics:

> "Linearity also occurs in acoustics. Indeed, the intensity of sounds, when one leaves aside considerations of phase, combine arithmetically. When several sources emit sound at the same time, the hearing process accommodates all the vibrations, and, to a certain extent, isolates and compares them. If this were not the case, there would be no orchestras. Since preserving the ratios among the sound sources is necessary for an intelligent understanding of the sound scene, all amplifiers (or transmitters) are required to comply with the relative proportions of the source origins, i.e. in mathematical terms, they must be *linear*."

In brief, sound waves are linear w.r.t. sound sources. Next Serra considers linearity in vision:

> "However, visual signals combine differently. Objects in space generally have three dimensions, which are reduced to two dimensions in the photograph or on the retina. In this projection, the luminances of the points located along a line oriented directly away from the viewer are not summed, because most physical objects are not translucent to light rays, in the same way that they would be to X-rays, but are opaque. Consequently, any object which is seen hides those that

are placed beyond it with respect to the viewer: this self-evident property is a basic one."

This comparison between audition and vision shows a confusion between light waves and visual signals (that is, images), and between light sources and reflecting objects. A strict analogy between acoustics and optics would say: "The intensity of lights, when one leaves aside considerations of phase, combine arithmetically. When several sources emit light at the same time, the visual process accommodates all the vibrations, and, to a certain extent, isolates and compares them. If this were not the case, there would be no light shows." Indeed, televisions are subjected to linearity in the same way as sound amplifiers. As far as wave sources are concerned, the fact that the three-dimensional visual scene is projected onto the two-dimensional retina does not matter more than the fact that the three-dimensional auditory scene is projected onto the two-dimensional ear drum. However Serra does not speak of light waves, but of visual signals (that is, two-dimensional projections of these waves), and above all does not consider linearity w.r.t. light sources, but w.r.t. objects in the scene: this is completely different. Indeed, in acoustics too, the auditory signal is not a linear function of the set of objects in the scene, and this is what makes concert hall acoustics a difficult matter. Thus we see again that the difference between vision and audition is not much a difference of signals or waves, but rather a difference of purpose: vision looks at objects in the scene, not light waves, while audition listens to sound waves, not objects.

The next question is: what matters more, the fact that objects are opaque to light, or that they have sharp edges in geometrical optics (as we explained in Subsection 2.2). Consider indeed X-ray images. Following Serra's argument, methods deriving from sound analysis would be more appropriate for dealing with them than those from image analysis, for example mathematical morphology. This is not the case; standard image analysis methods, including mathematical morphology, have been applied to X-ray images. Following our analysis, as X-rays have very short wavelengths, they produce sharp edges, and so obviously edge detection algorithms can be applied to the images that they make; the fact that objects are translucent implies that such edges can be superposed, and hence that segmentation must be done in a different way than with opaque objects. Next, image analysis deals also with images of two-dimensional objects, for example flat sections of materials, where the problem of occlusion is absent; in fact, this type of images is very sucessfully dealt with by mathematical morphology. Note finally that in the case of acoustics, objects are not completely 'transparent' to sound waves; a wall separating completely a source from a listener will generally act as a filter. The fact that an object does not completely 'hide' in the acoustic sense another one behind it is due to the fact that the sound wave scattered by that object is propagated in all directions and produces no sharp-edged shadow hiding another object behind (see again Subsection 2.2): a sound 'turns around' obstacles, even 'opaque' ones.

From the above considerations Serra deduces that a set approach is necessary:

> "Stating that A is in front of B is equivalent to asserting that we see the visible contour of B minus, in the set sense, that of A. Stating that A hides B is equivalent to saying that the contour of B is included in the contour of A, etc. A

morphological description, i.e., a description of their shapes, must primarily use *portions* of space."

Other similar considerations are found in Section 5.2 of [49]. Note first that the above statements are not exact: A is in front of B if and only if we see the visible contour of B minus the *projection* (not the contour) of B; A hides B if and only if the contour of B is *masked by* the projection of A (not contained in it), and in this case the contour of A will be *surrounded by* the contour of B. We will discuss further such confusions between masking and inclusion in Subsection 5.2.2.

Second, we could add further related statements like: the shadow of a union of objects is the union of their shadows. All this is sensible in the case of objects and Boolean images, but mathematical morphology does not restrict its scope to them, it studies other things, in particular grey-level images. How can the above considerations be translated in a meaningful way for grey-level images? What then is the theoretical justification for grey-level morphologigal operations? The traditional approach (see [46], Chapter XII) is to consider a grey-level image on a two-dimensional plane as a three-dimensional set (with the grey-levels forming the third dimension), but then there is in general no direct correspondence between the shape of a three-dimensional physical object and the 'shape' of the three-dimensional set corresponding to its two-dimensional grey-level image. This problem will be discussed in more detail in Subsection 5.2.1.

3. Linear Filtering, Fourier Analysis, and their Limitations

Given a complex-valued function f in n real variables, the Fourier transform of f is the complex-valued function $\mathcal{F}(f)$ in n real variables defined by

$$\mathcal{F}(f)(\omega_1 \ldots, \omega_n) = \int_{\mathbb{R}^n} f(t_1, \ldots, t_n)\ exp\Big[-2\pi i(t_1\omega_1 + \cdots + t_n\omega_n)\Big]\, dt_1 \cdots dt_n. \qquad (1)$$

In some textbooks the factor 2π does not appear in the imaginary exponential; in quantum mechanics it becomes $2\pi/h$, where h is Planck's constant, and then the above formula links the 'wave functions' of position and momentum [22]. However with these modifications the theory remains the same, with only factors 2π or $1/h$ to be added in some formulas.

Given two complex-valued functions f and g in n real variables, one defines the Hermitian product

$$\langle f, g\rangle = \int_{\mathbb{R}^n} f(t_1, \ldots, t_n)\, g^*(t_1, \ldots, t_n)\, dt_1 \cdots dt_n,$$

where g^* is the complex conjugate of g, leading to the L^2 norm

$$\|f\| = \sqrt{\langle f, f\rangle}.$$

The convolution operation $*$ is defined as follows:

$$(f * g)(s_1, \ldots, s_n) = \int_{\mathbb{R}^n} f(t_1, \ldots, t_n)\, g(s_1 - t_1, \ldots, s_n - t_n)\, dt_1 \cdots dt_n.$$

Now the main properties of the Fourier transform is that for functions f, g satisfying certain mathematical properties (which we will not consider here), we have

$$\begin{aligned} \mathcal{F}(\mathcal{F}(f))(t_1, \ldots, t_n) &= f(-t_1, \ldots, -t_n) \qquad \text{almost everywhere,} \\ \|f\| &= \|\mathcal{F}(f)\|\,, \\ \text{and} \qquad \mathcal{F}(f * g) &= \mathcal{F}(f)\mathcal{F}(g). \end{aligned} \tag{2}$$

3.1. The uncertainty relations and their consequences

In 1946 Gabor [11] showed the mathematical relations underlying a general principle of information communication: to transmit a certain amount of information in a unit time, a certain minimum of frequency bandwidth is necessary. In other words, a meaningful signal cannot be arbitrarily compressed in both the time and frequency (Fourier) domains. This principle is mathematically identical with the 'uncertainty principle' in quantum mechanics, which sets a theoretical limit on the precision of joint measurements of the position and momentum of a particle (because the 'wave functions' of position and momentum form a Fourier pair).

The precise mathematical formulation of these relations was given in [11] for one-dimensional signals (in other words, for $n = 1$ in (1)). For the sake of clarity we will indeed first consider functions in one variable, and next the generalization to functions in $n > 1$ variables, in particular for $n = 2$ (pictorial signals).

Given a function having norm $\|f\| = 1$, the measure of its spread around a point a can be given by

$$\|f(t) \cdot (t-a)\| = \Big(\int_{\mathbb{R}} |f(t)|^2 \ (t-a)^2 \, dt\Big)^{\frac{1}{2}}.$$

For example in quantum mechanics $|f(t)|^2$ represents the probability distribution for the measurement associated to the 'wave function' f, and if a is its mean, then $\|f(t) \cdot (t-a)\|$ is its standard deviation. If $\|f\| \neq 1$, we can 'normalize' f (divide its values by the norm $\|f\|$), or equivalently the measure of the spread of f is $\|f(t) \cdot (t-a)\| \,/\, \|f(t)\|$. One can then show that under general mathematical conditions on f (which we will again not consider), the spread of f around a and that of its Fourier transform $\mathcal{F}(f)$ around b have their product bounded by below:

$$\frac{\|f(t) \cdot (t-a)\|}{\|f(t)\|} \, \frac{\|\mathcal{F}(f)(\omega) \cdot (\omega - b)\|}{\|\mathcal{F}(f)(\omega)\|} \geq \frac{1}{4\pi}, \tag{3}$$

with the equality if and only if f takes the form

$$f(t) = C\, exp[-A(t-a)^2]\, exp[2\pi i b(t-a)], \tag{4}$$

where C is a complex constant and A a positive real one (in other words f is, up to a multiplicative constant, a Gaussian envelope centered about a modulated by an imaginary exponential of frequency b). Functions of the form (4) are called *Gabor functions*.

Consider now n-dimensional signals. Here the generalization of (3) is given by:

$$\frac{\|f(t_1, \ldots, t_n) \cdot (t_i - a_i)\|}{\|f(t_1, \ldots, t_n)\|} \, \frac{\|\mathcal{F}(f)(\omega_1, \ldots, \omega_n) \cdot (\omega_j - b_j)\|}{\|\mathcal{F}(f)(\omega_1, \ldots, \omega_n)\|} \geq \frac{1}{4\pi} \qquad \text{for } i, j = 1, \ldots, n. \tag{5}$$

The equality holds for every $i, j = 1, \dots, n$ if and only if

$$\prod_{i=1}^{n} \frac{\|f(t_1, \dots, t_n) \cdot (t_i - a_i)\|}{\|f(t_1, \dots, t_n)\|} \prod_{j=1}^{n} \frac{\|\mathcal{F}(f)(\omega_1, \dots, \omega_n) \cdot (\omega_j - b_j)\|}{\|\mathcal{F}(f)(\omega_1, \dots, \omega_n)\|} = \left(\frac{1}{4\pi}\right)^{2n},$$

and this happens if and only if f is an n-dimensional generalization of Gabor functions of the form:

$$\begin{aligned} f(t_1, \dots, t_n) = C\, exp\big[-\big(A_1(t_1 - a_1)^2 + \cdots + A_n(t_n - a_n)^2\big)\big] \\ \cdot\, exp\big[2\pi i\big(b_1(t_1 - a_1) + \cdots + b_n(t_n - a_n)\big)\big], \end{aligned} \tag{6}$$

where C is a complex constant and $A_1, \dots, A_n$ are positive real ones. By analogy with quantum mechanics, inequalities (3) and (5) are called the *uncertainty relations*.

In order to understand their meaning in practice, suppose that we convolve a signal g by a linear filtering function f. As $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$, the above implies that if the spread of f is small, in other words the action of the convolution is localized in the spatiotemporal domain, then the spread of $\mathcal{F}(f)$ is large, in other words the convolution operates on a large frequency bandwidth. Thus localization in the spatiotemporal and Fourier domains are contradictory requirements. The two extreme cases illustrating this opposition are:

— f being the Dirac distribution about $(t_1, \dots, t_n)$, which is completely localized in the spatiotemporal domain and gives $(f * g)(s_1, \dots, s_n) = f(s_1 - t_1, \dots, s_n - t_n)$, but with $\mathcal{F}(f)$ having infinite spread.

— $f(t_1, \dots, t_n) = exp[2\pi i(\omega_1 t_1 + \cdots + \omega_n t_n)]$, a function completely unlocalized in the spatiotemporal domain, but with a complete localization in the Fourier domain, since $(f * g)(s_1, \dots, s_n) = exp[2\pi i(\omega_1 s_1 + \cdots + \omega_n s_n)]\mathcal{F}(g)(\omega_1, \dots, \omega_n)$.

Between these two extremes, Gabor functions realize the best compromise between localization in the spatiotemporal domain and in the frequency bandwidth.

Gabor [11] noted that frequencies defined globally in the Fourier transform contradict our physical intuition of local frequencies, in particular changing ones (for example in music). Gabor functions were thus proposed as local frequency units, the Gaussian envelope providing the temporal extent and the imaginary exponential the frequency.

In image analysis, interesting features (such as edges) form patterns localized both in space and frequency. Thus the Fourier transform is inappropriate for analysing the meaning of images. It simply turns the problem upside down, from an information perfectly localized in space and completely unlocalized in frequency to one vice-versa. Although Gabor's work dates from 1946, until 1980 most researchers in the field of human or machine vision remained unaware of it. The Fourier transform as a model of visual coding and analysis was fashionable in the early seventies, sometimes under the appealing denomination of 'holographic model' of vision and memory. In parallel and opposition to it was developed a local methodology for feature analysis which generalized template matching and produced the 'difference of boxes' technique for edge detection [43]. The defect of such an approach is that it ignores the effects of such operations on the frequency spectrum and typically gives 'dirty' results (see [30]).

Since then the work of Gabor has been 'rediscovered' and its consequence become understood. Some approaches refer explicitly to the 'uncertainty principle' [56,57]. The optimality of Gabor functions w.r.t. the uncertainty relations has been exploited in the now current practice of preprocessing images (before gradient measurement) by convolving them with Gaussians (Gabor functions with frequency zero) instead of boxes or pure low-pass filtering functions. Daugman [7] has studied two-dimensional Gabor functions in more details, and presented evidence that they are implemented in cells of the mammalian visual cortex.

Moreover, uncertainty relations imply that any linear processing stage in visual analysis needs the concurrent use of several filters having different spatial resolutions. Indeed, the set of linear shift-invariant operators is closed under composition, in other words it is useless to repeat linear convolutions in succession, since the result will always remain a linear convolution. We must thus do something else afterwards, in particular if we want to pass from filtering to recognition, which involves classification into discrete categories. For example we can do a measurement, or apply a non-linear shift invariant operator, such as an edge detector. Now if we have used only one linear filter, a convolution by a function f, then by the uncertainty relations f cannot have a good localization in the both spatial and Fourier domains. A bad localization in spatial domain means that features (such as edges) detected afterwards will have an imprecise position. A bad localization in Fourier domain means that the spatial extent of these features will be imprecise, for example we will not be able to tell of an edge segment if it is purely local or if it separates two large regions.

It is possible to escape from this dilemma by convolving the image with several functions. For example we take f_1 and f_2 having a fine localization in the spatial and Fourier domains respectively (say, f_1 and f_2 are Gabor functions, f_1 with a narrow Gaussian envelope, but f_2 with a wide one); by correlating the results of a non-linear operation (such as edge detection) on the two convolutions of the image by f_1 and f_2, we can determine with precision both the position and spatial extent of features. This methodology is called *multiresolution image processing*, it has been applied in [30], where several (3 or 4) Gaussians corresponding to different resolutions were used for convolving the original image into 'frequency channels', and edges were detected in each 'channel'; the spatial coincidence of edge elements in all 'channels' was the criterion for selecting the meaningful edges of the image.

In human vision, four 'frequency channels' have been observed [55], each corresponding to a distinct resolution in spatial and Fourier domains. Furthermore, anatomical and physiological studies by Livingstone and Hubel [24,25] have revealed two complementary pathways in the monkey visual system, called 'magno' and 'parvo'; 'magno' neurons generally have a coarser spatial resolution but a finer temporal one than corresponding 'parvo' neurons, so that both combined realize multiresolution in the spatio-temporal domain. This is a simple means for escaping the limitations of the uncertainty relations.

Subtler methods involve convolution by Gaussians at all possible scales and the study of the effects of scale variations on detected edges [1,58,60]. It seems also that in humans attention can modify the size and shape of the receptive field of neurons of the visual cortex (see [18], Section 5.4); this corresponds somewhat to adaptive modifications of scale in edge detection, called 'edge focusing' [1].

3.2. The importance of phase in images

Apart from the above criticism of using the (global) Fourier transform, it is also important to stress some errors concerning what is the most important information contained in it. Indeed, by some kind of analogy with the method of spectrometry used to analyse the contents of light waves (for example in radioastronomy), many people think that the information in a signal is mainly contained in the amplitude spectrum of the Fourier transform, not in its phase spectrum. While it is clear that such an assumption is reasonable for light (the frequency of a photon is related to its energy, but its phase is meaningless or unmeasurable), this is not true for acoustic or pictorial signals. As shown in [37], for practical purposes, in the Fourier spectrum the phase is more important than the amplitude!

A typical experiment consisted in taking two photographs and building from them two hybrid images having each the amplitude spectrum from one of them and the phase spectrum from the other one. Then each hybrid contained the details of the original picture having the same phase spectrum, and not those of the other one whith which it shared the amplitude spectrum. The same experiment was performed with spoken sentences. To a sentence one associates its spectrogram, a two-dimensional signal giving the amplitude at each time t of each frequency ω. One knows [45] that our audition detects frequencies and is sensitive mainly to the local amplitudes of frequencies, and very weakly to their local phases, and so a spectrogram is a good encoding of a sound. By interverting phases in the Fourier spectrum of two sentences, each of the new hybrid acoustic signals resembled more (both acoustically and in its spectrogram) to the original sentence with which it shared its phase spectrum than to the other one having the same amplitude spectrum.

Similar experiments (with images only) were also made in [39]. In other words it was shown in practice that meaningful features, including local frequency amplitudes, depend more on global phases than on global frequency amplitudes.

In the specific case of visual signals, the importance of phase can be understood by the fact that blurring a picture by a convolution with an even-symmetric function (say, a Gaussian) preserves the phases, but not the amplitudes, in the Fourier spectrum.

There have been studies of the meaning of phase for features in an image. In [34] sharp edges and Mach bands were analysed in trapezoid periodic vertical gratings, that is images with the grey-level $g(x, y)$ of a point (x, y) given by

$$g(x, y) = \sum_{k=0}^{\infty} a_{2k+1} \sin\bigl(2\pi(2k + 1)x/T\bigr),$$

where $a_1 > 0$ and T is the period of the grating. When all perceptible harmonics were in phase concordance, in other words the portion in the series consisting of all terms having $a_{2k+1} < 0$ represented an imperceptible signal, a sharp edge was seen between dark and bright bands at positions where terms of the series with $a_{2k+1} > 0$ aligned all either in the $0°$ or in the $180°$ phase (corresponding to positive and negative edges respectively). On the other hand when there were perceptible harmonics in phase opposition, a Mach band appeared at junctions of ramps and plateaux, around positions where all terms of the series had a $\pm 90^\circ$ phase. Sharp edges and Mach bands excluded each other.

Further experiments were made with periodic vertical gratings [35] to show that perceptible features (edges, peaks, Mach bands, etc.) correspond to points of 'maximum phase congruency' in the Fourier decomposition of the signal (in other words points (x, y) where the Fourier terms $a_k \cos(\omega_k x + \varphi_k)$, with $a_k \geq 0$, are such that all phases $\omega_k x + \varphi_k$ are maximally concentrated about some value).

As remarked in [34], the coincidence of a sharp edge with points were all sines of the series are aligned either in 0° or in 180° phase is consistent with the proposal in [30] to take as edges the curves consisting of points where gradient maxima are detected in all 'frequency channels'. However when the match between gradient maxima in distinct 'frequency channels' is not exact, or equivalently the phases of the Fourier spectrum are not perfectly aligned, the method of [30] fails, but the feature can still be detected, thanks to a measure of 'phase congruency' given by the sum of the squares of the signal and its Hilbert transform (the signal having the same Fourier amplitudes, but with a 90° phase shift): the feature corresponds to a 'maximum phase congruency', and occurs at a point where that sum of squares has a local maximum. This leads to a practical method for edge and feature detection using pairs of 'matched filters' (convolution of the image with two functions forming a Hilbert transform pair, the first one even-symmetric, and the second one odd-symmetric). See [35] for more details.

As these two studies [34,35] dealt with periodic gratings, global phases were considered. However since features (edges, Mach bands, etc.) can be seen in localized portions of an image, one can suppose that the above analysis can be extended to local phases.

Therefore, while for acoustic signals global phases but local amplitudes count, it seems that for visual signals, phases count both globally and locally.

3.3. What can linear filtering do in images?

It is generally acknowledged that early human visual processing of images involves linear filtering, and even that there are distinct (and relatively independent) 'frequency channels', where various processings are made (see [29], Chapters 2 and 3), such as feature extraction. Note that by the uncertainty relations these 'frequency channels' do not correspond to sharp band-pass filters, because they must retain some spatial localization of features (and this is why we put the expression under inverted commas). The proper term would rather be 'spatial/frequency channels'.

It would be naïve to consider that low resolution 'spatial/frequency channels' correspond to global features while high resolution 'spatial/frequency channels' represent local details. This was for example the case in [5], where these 'channels' were also identified with band-pass filters. As pointed out by Marr (see Figure 2.2 of [29]), herringbone textures have perceptible global spatial structures of stripes which cannot be detected from low frequencies in the Fourier spectrum. Let us examine this example in more detail. Consider the texture shown in Figure 1, consisting of an succession of vertical stripes of width W having two alternating textures made of high-frequency periodic stripes in the two diagonal directions respectively. We assume that the average grey-level in both textures is a. Thus the grey-level is given by $g(x, y) = a + F(\omega(x + y))$ in even stripes and $g(x, y) = a + G(\omega(x - y))$ in odd stripes, where F and G are periodic functions (of period 2π) with zero mean (in other words whose Fourier series do not contain a constant term). On the lines separating the stripes we take the average between the two, that is $g(x, y) = a + \frac{1}{2}\big[F(\omega(x + y)) + G(\omega(x - y))\big]$. Consider the periodic function SQ corresponding to a square wave of amplitude ± 1 and period 2π, that is

$$SQ(t) = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin((2n+1)t).$$

Thus $SQ(t) = 0$ if t/π is an integer, and $-1^{\lfloor t/\pi \rfloor}$ otherwise. Then the grey-level of points in the image is given by

$$g(x, y) = a + \frac{1}{2}\big[F(\omega(x + y)) + G(\omega(x - y))\big] + \frac{1}{2}SQ(\frac{\pi}{W}x)\cdot\big[F(\omega(x + y)) - G(\omega(x - y))\big].$$

As F and G have zero mean (that is, no constant term in their Fourier series), the Fourier series of the grey-level function $g(x, y)$ will have as non-constant terms sine and cosine functions with the following arguments only:

$$m\omega(x + y) \pm n\frac{\pi}{W}x \qquad \text{and} \qquad m\omega(x - y) \pm n\frac{\pi}{W}x \qquad \text{where} \qquad \begin{array}{c} m = 1, 2, 3, 4, 5, \ldots \\ n = 0, 1, 3, 5, 7, \ldots \end{array}$$

Now the texture within stripes is taken with high frequency compared with that of the stripes, that is $\omega \gg \pi/W$. In other words the Fourier series will contain only high frequencies! This implies that no linear filtering (classical low pass, by Gabor functions, or anything else) will make low frequencies (of the order of $1/W$) appear. Hence we have shown that perceptible global features can exist without corresponding low frequencies. Note however that these global features (stripes of width W) appear explicitly on the amplitude and orientation spectrogram of local high frequencies in the image.

A simpler example is given by low-frequency amplitude modulation of sounds in music ('tremolo'). If a pure 440Hz sound is modulated in amplitude by a 5Hz sine, the resulting sound will be the superposition of two pure sounds at 435 and 445Hz. The 5Hz modulation does not appear by linear filtering of the sound, but is clearly present in the amplitude spectrogram of local frequencies around 440Hz.

Hence it is clear that a distinction must be made between global features, including periodic ones, and low frequencies and 'spatial/frequency'channels. A simple linear filtering is not enough to make apparent the global organization of a signal.

On the other hand one can assume that conspicuous local features imply the existence of local high frequencies (in other words the convolution with Gabor functions with a narrow Gaussian enveloppe and a high frequency modulation must give a strong result for some frequency bands); extrapolating from the above example, these frequencies can perhaps be much higher than what we would expect from the size of the features. Marr again has shown that in some cases the dependence of the perception of a texture on its Fourier spectrum increases when the resolution becomes finer. For example the chessboard (see Figure 2.2 of [29]), being the product of a horizontal square wave by a vertical one, has a Fourier spectrum whose power is concentrated near the diagonal directions, and totally absent in the horizontal and vertical ones. When the squares are large enough, we see the horizontal and vertical organization in the chessboard, as well as the diagonal one. As the squares get smaller, the diagonal organization becomes slightly more prominent.

Experiments have also been made on the relations between the human discrimination of periodic textures and their Fourier spectrum [32]. Given textures made by the superposition of at most 3 sine gratings whose orientations differ mutually by 60°, differences in the widths of gratings are well perceived, but this is not the case for differences of orientations. As pointed out by the authors, these results are difficult to reconcile with a linear filtering approach to perception, but they are consistent with the approach proposed by Marr [27,29] of extracting features from an image and building from them a 'primal sketch' (some sort of symbolic pencil drawing).

What can linear filtering do in image analysis? Certainly no miracles. Probably much less than for the analysis of sound signals. It is not unwise to consider such a technique, as it is often the case in recent works, as a 'preprocessing', which prepares the image for more interesting operations.

4. Concurrent Processing of Complementary Informations

Perception does not limit itself to the transmission of sensory data to the brain. These data are analysed and interpreted in order to understand the physical reality which gives rise to them. The information contained in a sensory signal must be assimilated by us. Thus, as said Serra [46], "to perceive an image, is to transform it". The Nobel laureate F. Crick explained very well the problem involved in the opposite conception of perception as a passive reception of sense data [6]:

> "Is there any idea we should avoid? I think there is at least one: the fallacy of the homunculus. Recently I was trying to explain to an intelligent woman the problem of understanding how it is we perceive anything at all, and I was not having any success. She could not see why there was a problem. Finally in despair I asked her how she herself thought she saw the world. She replied that she probably had somewhere in her head something like a little television set. "So who", I asked, "is looking at it?" She now saw the problem immediately."

Serra is also aware of the problem [48]: "The brain does not add a third 'eye', which would then *look* at the visual zones and *be observed* itself." Thus perception involves a transformation of the sense data, and he elaborates on the question by considering the algebraic and information processing aspects of such a transformation, in particular the algebraic property of *reversibility*.

4.1. Irreversibility and morphological filtering

Serra distinguishes between reversible operations used in the transmission and preprocessing of information, and irreversible ones involved in its understanding [48]:

> "In computer vision, what are we seeking—to transmit information or rather to assimilate it? Reversibility is acceptable when we improve the images that provide the input to a system, as with the case of spectacles for shortsighted people."

However, the recognition of objects, that is the assimilation of information, implies an irreversible process. As there is no 'homunculus' in our brain, "the chain stops there, and with it the notion of reversibility."

How is this irreversible process envisaged? The tradition of mathematical morphology is heavily influenced by the 'filtering' approach, where the word must be interpreted in its litteral sense of removing anything that is unwanted but preserving everything else. In [48] we read: "Recognition of an object simply means that all the rest has been eliminated from the scene. This is a definitive irreversible operation." In the preface of [46] the use of basic tools called 'structuring elements' was already presented in this perspective: "Chosen by the morphologist, they interact with the object under study, modifying its shape and reducing it to a sort of caricature which is more expressive than the actual initial phenomenon." In particular such a transformation involves a loss of information: we eliminate the information which is irrelevant from our point of view (for example the noise in an image).

In [48] Serra gives several practical examples of visual tasks (such as recognition of features or objects in an image) solved by morphological methods. They generally verify the following scheme:

- (*i*) Recognition is achieved by subjecting the image to a sequence of transformations $\psi_1, \ldots, \psi_n$, which constitute the distinct stages of the processing.
- (*ii*) For $i = 2, \ldots, n$, the input of ψ_i is the output of ψ_{i-1}.
- (*iii*) Each ψ_i ($i = 1, \ldots, n$) is a map $\mathcal{L} \to \mathcal{L}$ for some object space $\mathcal{L}$ (for example the set of grey-level images); the range of ψ_i need not be the whole of $\mathcal{L}$, it can be a a subspace $\mathcal{M}$ of $\mathcal{L}$ (for example the set of binary images), and then ψ_i is a map $\mathcal{L} \to \mathcal{M}$ (for example a thresholding).
- (*iv*) Each ψ_i ($i = 1, \ldots, n$) is idempotent, that is $\psi_i \cdot \psi_i = \psi_i$.

We illustrate this scheme on Figure 2 (*a*). Note that (*iv*) is meaningful only thanks to (*iii*). Indeed, if ψ_i is a map $\mathcal{L} \to \mathcal{K}$, where $\mathcal{K}$ is completely different from $\mathcal{L}$, the map ψ_i is not necessarily defined on $\mathcal{K}$, and then $\psi_i \cdot \psi_i$ is undefined, in other words idempotence is meaningless. This is the case when we transform a binary or grey-level image into something more elaborate, such as a description of features or objects to which attributes are associated.

In the examples given by Serra, point (*i*) of the scheme is always verified, and whenever (*iii*) holds, point (*iv*) is also true, except in some relatively trivial cases. When (*iii*) does not hold, it is in general for 'natural' reasons, when the goal of the processing is not the same type of object as the input (for example in segmentation, where we derive from an image a partition of space). We will consider violations of point (*ii*) later.

The signification of idempotence is strongly stressed. First, it acts as a buffer to the propensity for the transformation ψ_i to lose information. Second, it characterizes that transformation as a stage in the processing. Indeed, if an operation is idempotent, then there is no point repeating it, and so we must do something else, go to another stage.

Conversely, a stage must produce a clear result, and not stop halfway (as a 'filter' which would eliminate only a part of the noise, and would need to be iterated), and this corresponds to an idempotent operation.

In Section 5 of [48], five heuristic rules are given for the proper choice and sequencing of operators ψ_i, in order to control the loss of information. It is interesting to note that Serra's fifth rule is violated in an example where precisely point (ii) above is not verified. The problem was to segment a grey-level image representing distinct neighbouring cells (see [48], Figure 8); the solution involved the binarization of the image (ψ_1), then the extraction of the components of that binary image (ψ_2), and finally these binary components were used as masks to separate the components of the original grey-level image (ψ_3). In other words we have 3 stages, and the input of ψ_3 consists of both the input of ψ_1 and the output of ψ_2. Such a method is called by Serra 'backtracking'; he states that it is quite common in mathematical morphology.

Serra's general scheme is acceptable for a special purpose image analysis system. However the several allusions made in [48] to human perception, artificial intelligence, computer vision, and their relations with the methodology proposed there, require the consideration of a wider scope for the terms 'vision' and 'recognition'. If we compare Serra's scheme with what we generally mean by 'perception', several criticisms based on 'everyday psychology' and 'computational common sense' can be raised.

First, do we always eliminate what is unwanted when perceiving something? While reading, we recognize the characters, but also the fonts used. In a noisy image, we see the underlying picture, but also the noise; we can even distinguish several distinct noises. When listening to someone, we recognize the speech, but also the speaker's voice, its intonation, etc.. Thus elimination of unwanted information is not so much a matter of sensory and perceptual 'filtering', but rather one of conscious attention. This is to expect from an advanced general purpose perceptual system, in contrast with a primitive one, such as the frog's eye [23] which 'sees' food only if it is of the appropriate size and moving. Single task algorithms in Serra's examples cannot be compared to human perception; in fact, most image analysis systems built today are rather on the level of the frog's eye and brain.

Second, do we lose much information? If we examine the human eye, we note that the retina has more than 100 million photoreceptors, but the optic nerve contains only about one million fibers. This looks like an important reduction in information content. What happens in fact is that there is a 'deliberate' increasing loss of optic resolution as one moves from the fovea to the periphery; but we do not see a blurred image with a sharp center because we are constantly moving our eyes throughout the scene. Thus the complexity of a visual stimulus is partially transformed from a spatial magnitude to a temporal one. One of the roles of conscious attention is to drive our eyes towards features of interest (see [18], Chapter 7). We have at our disposal the possibility to realize the desired trade-off between time and information. With enough time and attention, we can retain and use a great part of the information contained in a static scene (this is the case, for example, with painters).

When looking at something, our perception is not restricted to the final product of visual analysis. We see the points, the local features, the objects, and the spatial relations all together. On account of such an elementary fact E. Kent [18] proposed a processing paradigm where the result of each processing stage is added to a 'perceptual data bus', which gives an ever richer representation of the input data; each processing stage has access to the whole of that 'data bus' as it has already been obtained through the previous stages. We illustrate this scheme in Figure 2 (b). If we formalize this, we can replace points (ii, iii, iv) above by:

(ii') For $i = 2, \dots, n$, the input of ψ_i has access to the outputs of all ψ_j, for $j = 1, \dots, i-1$, and eventually the original signal (the input of ψ_1). The perceptual output of the whole process is given by the successive addition to the original signal of the outputs of $\psi_1, \dots, \psi_n$.

(iii') The initial data belong to an object space $\mathcal{L}_0$, the output of each ψ_i ($i = 1, \dots, n$) belongs to an object space $\mathcal{L}_i$, and ψ_i is a map $\mathcal{L}_0 \times \cdots \times \mathcal{L}_{i-1} \to \mathcal{L}_i$; the final output belongs to the object space $\mathcal{L}_0 \times \cdots \times \mathcal{L}_n$. The object spaces $\mathcal{L}_0, \dots, \mathcal{L}_n$ need not be similar.

Here the method of 'backtracking', which was presented by Serra as a violation of his fifth rule, appears instead as the rule. Thus we have not to worry at all if one processor ψ_i loses necessary information. An advantage of such a scheme is that in case of failure of one of the processing stages (because of damage or of a too difficult task), the remaining ones can still manage to get the best out of the results of previous stages. On the other hand, in Serra's scheme, failure in any processing stage leads to the absence of output. Kent's scheme seems confirmed in the case of human vision. As pointed out in [18], damage to the primary visual cortex leads to blindness (in the corresponding portion of visual space), while damage to the secondary visual cortex leads to several deficits in scene recognition, but not to blindness.

If we recall our above discussion of point (iii) in Serra's scheme, here idempotence is generally meaningless, since we are discussing maps from one object space to another one. This can happen in practice. For example in [27] symbolic representations of features such as edges, expressed as lists of parameters, are derived from a grey-level image in order to form a 'primal sketch'; moreover perceptual groups (what one calls 'gestalts') are defined there in a recursive way as geometrical arrangements of elements which are themselves either simple primitives (line, blob, etc.) or such (smaller) perceptual groups; clearly the operator which produces such a symbolic description from a grey-level image cannot be applied to its output, since the latter is not a grey-level image.

Hence restriction of processing to a single object space (say, that of grey-level images) means in practice that we remain in low-level image analysis, since we exclude any other forms of data, such as symbolic attributes, syntactic structures, etc.. We said at the end of the previous section that linear filtering of images can be considered as a 'preprocessing'. We can now suggest that operations transforming images into other images of the same kind, and having various interesting algebraic properties (such as idempotence, or the other ones mentioned in [48]), can be considered as a 'secondary processing', at the beginning of visual perception. A general purpose vision system must be able to produce symbolic representations, and needs thus a wider set of object spaces and operations.

A problem concerning both Serra's and Kent's schemes (and many paradigms, such as

Marr's [29]), is the possibility of feedback between processing stages. It is known [53] that most connections between distinct visual areas of the brain are reciprocal. We mentioned above that in human vision, attention can modify the size and shape of the receptive field of neurons in the visual cortex; this is possible by the reciprocal connections between the primary visual cortex and cortical areas devoted to higher level functions (see again [18], Section 5.4). In image analysis algorithms based on morphological methods, many operators depend on certain parameters whose actual values often vary with the type of image one deals with: for example the size and shape of structuring elements, the threshold for a thresholding, etc.. The fixation of these parameters is possible by a feedback from processing modules higher in the hierarchy. In practice the feedback is often provided by the human experimenter who selects best values according to their result. Once these best values are determined, they become fixed and the algorithm is then restricted to only a certain type of images. However for an autonomous general purpose vision system, one must not rely on human feedback, but rather implement feedback inside it.

In the design of an automatic programming system using morphological operations, Schmitt [44] defines the extraction of a type of feature as an algebraic operation having (possibly several) images and features as arguments. In particular point (*ii*) is Serra's scheme is not verified (and point (*iv*) is not even considered). Between the processing stages are inserted evaluation procedures which, in case of failure, lead to a 'backtracking' and the modification of the previous stage. This is a particular form of feedback between the different stages of the analysis. However the data resulting from this system seem to remain in the same object space (grey-level or binary images), which is clearly a limitation of such a system.

4.2. Complementary informations

The above criticism deals rather with matters of efficiency and good engineering. We will now consider more fundamental problems related to the multiform and sometimes contradictory nature of the extraction of visual information. A prime example of what we intend to discuss has been encountered above with the uncertainty relations which limit the joint resolution of linear operators in the spatiotemporal and Fourier domains. We saw that a way out of this dilemma was multiresolution image processing, that is to apply in parallel several linear operators having different resolutions, and to follow each by a non-linear operator. Linear operators with a fine spatial or temporal resolution allow the location of features in space or time, while those with a coarse resolution allow the estimation of their spatial or temporal extent.

This complementarity between the description of features or objects and their location, or to put it more briefly [18], between the 'what' and 'where', is apparent in the anatomy and physiology of the human (or primate) visual system. Experimental damage to the brain of monkeys suggest that the primary visual cortex is the source of two distinct pathways [33]. One is directed towards the inferior temporal areas, and enables the visual identification of objects. The other goes to inferior parietal areas, and allows the visual location of objects. In the parietal lobes, visual spatial information combines with spatial information from other senses (in particular touch) in order to give a unified perception of space (see [18], Chapter 6).

We mentioned above 'magno' and 'parvo' neurons studied by Livingstone and Hubel [24,25] in the monkey visual system, with their opposite spatial and temporal resolutions. 'Magno' neurons have generally a coarse spatial resolution, a fine temporal resolution, a high contrast sensitivity, and are nearly colour-blind. On the other hand 'parvo' neurons have generally the opposite features: a fine spatial resolution, a coarse temporal resolution, a low contrast sensitivity, and they are colour-sensitive. In fact, these two types of neurons form relatively distinct pathways in the brain. The 'magno' pathway is responsible for figure-ground discrimination, motion detection, and the perception of three-dimensional spatial relations, in particular depth. 'Parvo' neurons generate two pathways. First the 'parvo-interblob' pathway, which handles the static recognition of form, in particular a high-resolution analysis of details. Second the 'blob' pathway, although consisting mainly of 'parvo' cells, has some inputs from 'magno' cells, and it deals with colour perception. See [24,25] for more details.

This subdivision is rather schematic: cells in a pathway do not all have the above-mentioned features, and there is some cross-talk between the pathways. It is better to understand these pathways as concurrent functional streams, which are distinguished not so much by their anatomic connections, but by their goals and computational strategies [8]. A neuron which has some features can be 'naturally' associated to a particular task, but it can also contribute to another one. For example a cell sensitive to the velocity of points is 'obviously' involved in motion detection (this is 'short-range' or 'low-level' motion); however the disparity between the velocities of different groups of points allow us to segment the scene into separate moving objects. Conversely, a cell sensitive to grey-level gradients 'clearly' intervenes in edge detection and scene segmentation; but if distinct edges change their location through time, they lead to a perception of motion (this is 'long-range' or 'high-level' motion). Other examples are given in [8]. Thus the three functional streams ('magno', 'parvo-interblob', and 'blob') correspond to three distinct tasks in vision, having each different computational strategies.

Although the evidence presented is purely biological, this distinction between the identification and location of objects may well correspond to a computational necessity, and one should envisage its possibility in artificial systems. In any case we will see in the next section that the methods of mathematical morphology are more suited to the recognition of two-dimensional form than to the analysis of three-dimensional spatial relations.

5. The Possibilities and Limitations of Mathematical Morphology

We now come to a more detailed consideration of what mathematical morphology can or can't do. This requires first a discussion of the physical formation and meaning of grey-levels in pictures, as well as a consideration of the mathematical problems involved in the extraction of information from image intensities.

5.1. *Image grey-levels and their meaning*

Given a point on a picture corresponding to a location on the surface of an object, the grey-level of that point depends on many factors, mainly the material properties of the surface at that location (colour, specularity, etc.), its orientation, the spatial distribution of light sources, and the viewpoint (not to mention the pattern of transmission of light by air). A detailed analysis of the relations between these four factors, and of the formation of image grey-level intensities, can be found in [16] (a shortened exposition is given in [29], Section 3.8).

In the psychophysics litterature, the pattern of reflection of light by a surface (depending on its material structure) is called its *reflectance*, while the combination of spatial distribution of light sources, surface orientation, and viewpoint, is summarized under the word *illumination*. In Horn's terminology [16], the latter word has a more restricted meaning, namely the light energy flux per unit area perpendicular to it, independently of surface orientation and viewpoint; in order to avoid confusion, we will adopt Horn's definition, and will say 'illumination/orientation' when the broader meaning of the term 'illumination' is intended. Since Helmholtz, it is well-known that the luminous intensity of a surface element depends on both its reflectance and illumination/orientation, and that the human visual system can up to a certain point separate these two factors. It has long been a chicken-and-egg problem of determining which one of these two factors is compensated first in order to evaluate the other one. In general, none of them can be simply eliminated, and there is some ambiguity in images: the same pattern of intensities can arise from changes of reflectance or from changes of orientation in a surface (for example film actors use makeup to produce depth effects on their face). The human visual distinction between changes in illumination/orientation and changes in reflectance uses global context as well as several constraints corresponding to simple physical assumptions: we will discuss this in Subsection 5.3.

This has an important consequence for low-level image analysis. Many operators are supposed to extract information from local patterns of grey-levels in an image, without distinction between those arising from illumination/orientation features or from reflectance features (as we will explain in Section 5.3, this distinction is generally impossible without global context). This works in two cases:

(1) This information is significant, regardless of its origin in illumination/orientation or in reflectance. For example edges, lines.

(2) The illumination/orientation has no effect on image grey-levels. Two subcases are possible:

  (*a*) The surface is flat, subjected to a constant illumination, in a fronto-parallel plane and viewed at a long enough distance (so that the rules of orthographic projections can be applied). For example: tomographic images, flat sections of materials in petrography, mineralogy, cytology, etc..

  (*b*) The illumination is constant, objects are not really opaque, but their degree of transparency depends on their depth and material composition. For example: X-ray angiography.

Other subcases of (2) are possible.

It should be noted that almost all successful applications of mathematical morphology have been in images arising from situations in case (2). Indeed, this discipline uses as basic tools structuring functions whose interaction with an image should reveal some aspects of the material structure of the underlying object which are sought for. Thus local configurations of grey-levels are implicitly assumed to correspond to material properties of the object pictured. The question of the origin of image intensities, in particular problems of illumination and surface orientation, are not considered. For the analysis of three-dimensional scenes, they must however be taken into account.

5.2. *Some tools used in mathematical morphology*

Certain types of operations are widely used in this discipline, generally because they have certain particular mathematical properties. We will criticize these operations, or at least the justifications given for their importance.

5.2.1. Grey-level structuring functions and invariance under grey-level transformations. In both orthographic and perspective projections, a translation in a fronto-parallel plane induces a translation in the image plane. This is the physical basis for the invariance of image transformations under spatial translations.

What about invariance by certain grey-level transformations? The only possible physical principle underlying it is that a uniform modification of the intensity of light sources does not change the nature of the scene. We must explicit the notion of 'uniform modification'. As the reflectance of surfaces intervenes as a multiplicative factor in their perceived luminosity, this means that for a single light source, a multiplication of light intensity by a factor leads to a multiplication of all luminosities by that factor. Thus for several light sources, a uniform modification must be a multiplication of each light intensity by the same factor. Hence the invariance of an image transformation must be under a multiplication of all luminous intensities by the same factor. If the grey-level g of a point is related to its luminous intensity i by a function f, that is $g = f(i)$, then operators on grey-level images should be invariant under grey-level transformations corresponding to scalar multiplications of luminous intensities, that is maps $g \mapsto f\big(c \cdot f^{-1}(g)\big)$ for $c > 0$.

In mathematical morphology, dilations or erosions by (additive) grey-level structuring functions are invariant under grey-level translations, which implies the assumption that grey-levels are related to luminous intensities by a logarithmic function. For multiplicative structuring functions [15], the underlying assumption is that grey-levels are proportional to luminous intensities. However these assumptions are not explicitly stated. The underlying error is to consider the three-dimensional *profile* formed on a subset of an image by the points and their grey-levels, as a three-dimensional *shape*, generalizing two-dimensional shapes found in binary images. However, this does not correspond to shapes of objects in three-dimensional space, since the luminosity of points is not the same as their depth. As we explained in the previous subsection, changes in luminous intensity of points can originate from different factors (reflectance, orientation, illumination, etc.).

One generally feels that an image remains the same if its contrast is enhanced by a strictly increasing linear transformation of grey-levels: $g \mapsto a \cdot g + b$, where $a > 0$. Now morphological operations by a grey-level structuring function are invariant under such transformations only if that structuring function is *flat*. This explains why most practical applications of grey-level morphology use flat structuring elements. Non-flat structuring elements have been used by Sternberg [50] (for example spherical ones, in the 'rolling ball algorithm'); reasons for such a choice include: a smoother grey-level profile in the resulting image, or the assumption that the grey-level profile (improperly called 'shape') of the structuring function is related to that of the feature or object which must be enhanced in the image. Note also that some numerical transformations of binary images can be implemented as grey-level morphological operations, for example the distance transform through dilations and erosions by conical structuring elements [50], but then the grey-level profile of the structuring element has a clear objective meaning.

5.2.2. Increasing operators. Increasing operators have a prominent place in mathematical morphology. For example *morphological filters* (see Chapters 5 and 6 of [49]) are defined as idempotent increasing operators. Here the word 'filter' is used by analogy with band-pass filters of linear signal processing, and the increasingness of morphological operators is counterposed to the linearity of classical operations in Fourier analysis (see Section 5.2 of [49]). What justification is given for increasing operators?

In [19] Lantuéjoul and Serra explain that for an operator ψ, the relations $\psi(f \vee g) \geq \psi(f) \vee \psi(g)$ and $\psi(f \wedge g) \leq \psi(f) \wedge \psi(g)$, both equivalent to the fact that ψ is increasing, simply mean that "when f is a stage and g is an actor in the proscenium (i.e. who *masks* a part of f), then the transform $\psi(g)$ also masks the parts of the transform $\psi(f)$ which are behind it."

This is plain nonsense. First, occlusion (or masking) in three-dimensional space is not equivalent to inclusion in the image. We already touched this point at the end of Subsection 2.3. For example if A masks a part of B which is behind it, then the portion of the contour of B masked by A will not be seen, we will see instead the corresponding portion of the contour of A, but the operation of taking the contour is *not* increasing. The projection of Boolean objects into a Boolean image preserves the union (and so is increasing), but if the projection of A is contained in the projection of B, this does not tell whether A is in front of, or behind B. For grey-level images, nothing of the sort remains: a dark object can hide a light one, and so the grey-level image formed by the two objects is not the grey-level supremum of the two images formed separately by each of them.

In Section 5.2 of [49] this argument is summarized by stating that the notion of inclusion is used to express the fundamental law of occlusion in three-dimensional space. This is also placed in the framework of the arguments about different modes of combination for acoustical and visual signals which we criticized in Subsection 2.3.

In [26] increasing operators are justified by the principle of *psychophysical isomorphism* of Gestalt psychology, quoting Köhler's assertion that "experienced order in space is always structurally identical with a functional order in the distribution of underlying brain processes". Apart from the fact that 'order in space' (occlusion) is not the same as inclusion, the notion of 'psychophysical isomorphism' is rather mythical, and without empirical justification. We explained in Subsection 4.1 that perception involves a transformation of data allowing their interpretation, and not an 'isomorphic' reproduction (which would presumably be looked upon by a 'mind's eye' or 'homunculus'). Vision is a *process* (involving 'computations') of *representing* objects and spatial relations, and not a physical reproduction of them in the brain (see [29] for a criticism of the naïve conceptions of the Gestalt school).

There is thus no a priori justification for using increasing operators. It all depends on what properties are necessary. For example a 'sieve-type' filter (such as an opening or closing) is related to size criteria (see [46], Chapter X): if a feature is removed, any smaller one should also be removed; here the increasingness of the operator naturally comes in.

5.2.3. Thresholding and gradient. An operation often encountered in practical applications of mathematical morphology is *thresholding*: given a threshold t, mark all points of the image as white if their grey-level is $\geq t$, and as black otherwise. This is supposed to give a binary silhouette of the object represented.

An obvious question is to determine where the threshold should be set. In other words this operation is not invariant under grey-level translations. A simple remedy is *adaptive thresholding*, where the threshold is computed either globally from the histogram of the image, or locally, from the grey-level profile in the neighbourhood of each pixel. But there is a more fundamental problem: the grey-level of a point results from a combination of several factors in the scene (reflectance, illumination, orientation, etc.), and as we will see in Subsection 5.3, they intervene at a global scale. This means that grey-levels of individual pixels or even of small groups of pixels, taken into isolation, have generally no significance. The exception is the case (2) mentioned in the previous section, namely when grey-levels do not depend on illumination/orientation. We said that most applications of mathematical morphology belong to that case, so fortunately thresholding can be useful there. But for scene analysis, where the distinction between reflectance and illumination/orientation is crucial, this technique is bound to fail.

In mathematical morphology, authors are fond of defining morphological versions of operations in other methodologies. For example Sternberg's shading and shadowing algorithm [50] for representing the effect of oblique illumination on a three-dimensional terrain, which is based on Minkowski set operations, in contrast with the study of shading by Horn [16], which is based on the quantitative laws of optical reflection of materials.

One well-known operation for which morphological analogues have been defined several times is the gradient. Let us list some examples here. They are base on the two operations $\nabla^{\oplus}$ and $\nabla^{\ominus}$ which can be defined in the continuous case by

$$\nabla^{\oplus}(F) = \lim_{\lambda \to 0} \frac{(F \oplus \lambda B) - F}{\lambda}$$

$$\text{and} \qquad \nabla^{\ominus}(F) = \lim_{\lambda \to 0} \frac{F - (F \ominus \lambda B)}{\lambda},$$

where B is the compact unit disk centered at the origin, and in the digital case by

$$\nabla^{\oplus}(F) = (F \oplus N) - F$$

$$\text{and} \qquad \nabla^{\ominus}(F) = F - (F \ominus N),$$

where N is the 4- or 8-neighbourhood of the origin. Then at least three operators have been considered:

— The Beucher gradient (see [46], p. 441): $(\nabla^\oplus + \nabla^\ominus)/2$.
— The Lee gradient [21]: $\min(\nabla^\oplus, \nabla^\ominus)$.
— The 'nonlinear Laplace filter' [54]: $\nabla^\oplus - \nabla^\ominus$.

Which one is the best from a morphological point of view? We leave the choice to the reader.

A well-known problem concerning the usual (linear) gradient is that differentiation is an ill-posed problem (w.r.t. the usual L^p norm, $1 \leq p < \infty$) [52]. Given two metric spaces (U, ρ_U) and (F, ρ_F), suppose that we have the problem of finding a 'solution' $f \in F$ from 'initial data' $u \in U$; one says that this problem is *well-posed on the pair of metric spaces* (F, U) [51] if and only if:

(a) for every $u \in U$ there exists a solution $f \in F$;
(b) the solution is unique;
(c) the problem is *stable* on the pair (F, U), in other words a small enough variation in the data must lead to an arbitrarily small variation in the solution; more precisely for every $\varepsilon > 0$ there is some $\delta > 0$ such that for data $u_1, u_2 \in U$ leading to solutions $f_1, f_2 \in F$, $\rho_U(u_1, u_2) < \delta$ implies $\rho_F(f_1, f_2) < \varepsilon$.

If we consider the functions in the space $L^p(\mathbb{R})$ with the L^p norm $\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p \, dx\right)^{1/p}$, where $1 \leq p < \infty$, differentiation is ill-posed, because for two functions u_1, u_2 with $u_2(x) = u_1(x) + \eta \sin(\omega x)$ we have $u_2'(x) = u_1'(x) + \eta\omega \cos(\omega x)$, and by taking η very small and ω arbitrarily large, $\|u_1 - u_2\|_p$ can be as small as one wants while $\|u_1' - u_2'\|_p$ is arbitrarily large. One technique for dealing with ill-posed problems is *regularization* [51], which consists in finding an approximate solution which is close to the exact one and at the same time stable. In particular, differentiation is replaced by integration, which is stable. For example we can approximate differentiation by the convolution with the derivative of a Gaussian, which means in fact smoothing the original signal by a convolution with a Gaussion before differentiating it. A precise study of optimal regularization to the problem of differentiation in edge detection is given in [52]. Other methods for edge detection have been proposed, based on convolution with functions which are optimal w.r.t. some constraints.

It is easy to see that the morphological gradient is also ill-posed. In the case of one-dimensional signals, for $f(x) = \eta \sin(\omega x)$, we have $\nabla^\oplus(f)(x) = \nabla^\ominus(f)(x) = \eta\omega\,|\cos(\omega x)|$. However no 'morphological regularization' technique has been invented. Quite to the contrary, morphological gradients remain to the primitive level of the 'difference of boxes' methods used in early papers on edge detection.

5.2.4. Exact or approximate structuring elements? In mathematical morphology, the shape of sets is probed by their interaction with structuring elements. These are considered as *exact* templates. It is thus a yes-or-no procedure, where only exact 'matching' (i.e., inclusion, containment, equality, etc.) is considered. Now it is clear that perception can use elementary shape units, for example the *textons* proposed by Julesz [17] as basic primitives of texture, which have simple shapes: a bar, a cross, a T-juntion, an L, etc.. However they are not proposed in the form of exact templates: anything 'looking like' a T is a T, it is more a symbolic construction with bars than a group of points forming a given set. How does one recognize such an approximate or even 'symbolic' shape in mathematical morphology?

To consider each valid template approximating the shape of the structuring element is too costly. Other techniques involve replacing dilations and erosions by percentile operations (for example [42,59]), but this is not a complete solution to the problem.

5.3. Spatial relations and global image analysis

We will show here by a few concrete examples from the psychophysics literature that the analysis of spatial relations in a three-dimensional scene must be done at a global scale: local grey-level features taken outside context are unsufficient to determine these relations.

We mentioned at the beginning of Section 2 the first figure of [12]. Its upper part shows a scene illuminated from one side, with seveal planar surfaces having different orientations, and hence strongly contrasting luminous intensities; we do not see them as black and white patches, but as surfaces with different illuminations/orientations. The lower part shows a small portion of the image along the edge between two such surfaces, magnified in size and taken out of context; we see here a flat surface divided into a black patch and a white one. Thus the grey-level profile of an edge, taken into isolation, does not tell us if it is due to a change in reflectance or a change in illumination/orientation. Other examples showing the influence of global context in the discrimination between reflectance edges and illumination/orientation edges can be found in [12,13].

In [14] an astonishing demonstration was made. Two identical miniature rooms were made, one painted matte white, and the other matte black. Both were subjected to the same distribution of light sources, but in the black one the intensity of the light sources was stronger. Observers perceived the white one as white, and the black one as middle grey, even when its luminous intensity was higher than that of the white one. This is strange, because both scenes had the same orientations, a constant reflectance, and proportional illuminations. One would thus expect that by adjusting the intensity of light sources, both would give the same image. However there is a difference due to *secondary illuminations*: light can be reflected from a surface portion to illuminate a second one, and here the resulting intensity is proportional to the square of the reflectance. In other words, secondary illuminations effects are stronger in the white scene than in the black one. The result is that the global intensity profile is less contrasted in the image of the white room than in that of the black one. Of course this is a global effect, which cannot be detected from the analysis of an isolated local portion of the image.

How does one make sense from the intrication of effects of illumination, orientation and reflectance in an image? Gilchrist [13] proposes that the visual systems discriminates edges into two classes (due respectively to changes in illumination/orientation or to changes in reflectance), and then builds two images of the scene, one for orientations, and one for reflectances; in each of them all values are computed by integrating the changes along the edges. One must also take advantage of usual physical constraints to solve cases having ambiguous interpretations. For example in [40] it is shown that the human distinction between shadows and reflectance changes, or between concave and convex shaded surfaces, uses the implicit assumption that the illumination comes from one side only, and from above rather than from below. Similarly in [41], the apparent motion of groups of dots is interpreted under the assumption that points tend to have similar displacements.

As we said above, most applications of mathematical morphology deal with scenes where illumination/orientation does not matter (the case (2) above): flat sections of material under constant illumination, or X-ray images. For such restricted situations, the limitations of techniques based on local configurations of points and grey-levels (structuring functions, thresholding, etc.) are not evident. But for the analysis of three-dimensional scenes, this will be a serious problem. In fact other methodologies have been used for this purpose: see the numerous papers giving computational approaches to stereopsis, motion detection, extraction of structure for motion, three-dimensional shape from shading, perspective, etc.. Not much has been done on these subjects from the point of view of mathematical morphology. Some practical work is developed at the 'Centre de Morphologie Mathématique' on the detection of motion of cars on highways or ring roads. This situation is highly constrained by the constant perspective of road tracks, the uniform shape of cars, but even then many problems arise related to: the segmentation of some objects in low illumination possible only by optic flow, occlusions, shadows, specularities, etc. (S. Beucher, personal communication); each such problem is treated by 'ad hoc' techniques.

5.4. Concluding remarks

In Section 4 we pointed out the limitations inherent in operations which remain inside the same object space, and suggested that concurrent processing is necessary to circumvent the lack of universality of any operator, in particular w.r.t. the duality of the 'what' and 'where'. As seen in this section, mathematical morphology deals more with the 'what' than the 'where', and even with a purely two-dimensional 'what': its techniques are not suited to three-dimensional scenes, with their intricated global optical features. In fact this discipline does not base its analysis on quantitative optical properties of objects, and the only three-dimensional effect it mentions is occlusion and the ordering that it induces between objects. This allows the reduction of a scene to a set of flat figurines ordered by occlusion. We summarize these facts by the aphorism:

Mathematical morphology is flat.

We mentioned above the problem of structuring elements forming exact templates. Mathematical morphology deals more with minute details of shape than with its symbolic representations (for example, what is a T?), in particular groupings of symbolic tokens (see [27]). A question it has not taken up is figure-background discrimination. This fact, combined with its unsuitability to depth and motion analysis, allows us to say, in relation with the exposition in Subsection 4.2 of the functional streams in the anatomy of the human visual system:

Mathematical morphology is a 'parvo-interblob' system.

In other words, this discipline is one-legged, we need other legs (in particular a 'magno' system). It cannot be considered as a panacea, not even as the basic framework for exploring computer vision.

Acknowledgement

The author's ideas evolved through discussions with X. Aubert, H. Heijmans, J. Roerdink, J. Serra.

References

1. F. Bergholm: Edge focusing. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-9, no. 6, pp. 726–741, Nov. 1987.
2. M. Brady: Preface — the changing face of computer vision. *Artificial Intelligence*, Vol. 17, pp. 1–15, 1981.
3. P. Brou, T.R. Sciascia, L. Linden, J.Y. Lettvin: The colors of things. *Scientific American*, Vol. 255, no. 3, pp. 80–77, Sep. 1986.
4. G. Bruhat: *Cours de Physique Générale: Optique*, 5e. édition, Masson & Cie., Paris (1959).
5. F.W. Campbell: Sometimes a biologist has to make a noise like a mathematician. *Neurosciences Research Prog. Bull.*, Vol. 15, no. 3, pp. 417–424, 1977.
6. F.H.C. Crick: Thinking about the brain. *Scientific American*, Vol. 241, no. 3, pp. 181–189, Sep. 1989.
7. J.G. Daugman: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.*, Ser. A, Vol. 2, no. 7, pp. 1160–1169, Jul. 1985.
8. E.A. DeYoe, D.C. Van Essen: Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, Vol. 11, no. 5, pp. 219–226, 1988.
9. R.P. Feynman: *Q.E.D., the strange theory of light and matter*, Princeton University Press, Princeton (1983).
10. J.P. Frisby: *Seeing*, Oxford University Press, Oxford (1979).
11. D. Gabor: Theory of communication. *J. IEE*, Vol. 93, pp. 429–457, 1946.
12. A.L. Gilchrist: The perception of surface blacks and whites. *Scientific American*, Vol. 240, no. 3, pp. 88–97, Mar. 1979.
13. A.L. Gilchrist, S. Delman, A. Jacobsen: The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception & Psychophysics*, Vol. 33, no. 5, pp. 425–436, 1983.
14. A.L. Gilchrist, A. Jacobsen: Perception of lightness and illumination in a world of one reflectance. *Perception*, Vol. 13, no. 1, pp. 5–19, 1984.
15. S. Herman: Feature-size dependent selective edge enhancement of X-ray images, *SPIE Meeting on Medical Imaging* (1988).
16. B.K.P. Horn: Understanding image intensities. *Artificial Intelligence*, Vol. 8, pp. 201–231, 1977.
17. B. Julesz and J.R. Bergen: Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Sys. Tech. J.*, Vol. 62, no. 6, pp. 1619–1645, 1983.
18. E.W. Kent: *The Brains of Men and Machines*, Byte Books, McGraw-Hill, New York (1981).

19. C. Lantuéjoul, J. Serra: M-filters. *ICASSP'82, Internat. Conf. Acoustics, Speech & Signal Processing* (1982), pp. 2063–2066.
20. E.H. Land, J.J. McCann: Lightness and retinex theory. *J. Opt. Soc. Am.*, Vol. 61, pp. 1–11, 1971.
21. J.S.J. Lee, R.M. Haralick, L.G. Shapiro: Morphologic edge detection. *Proc. 8th. Internat. Conf. Pattern Recognition* (1986), pp. 369–373.
22. R. Leipnik: Entropy and the Uncertainty Principle. *Information & Control*, Vol. 2, pp. 64–79, 1959.
23. J.Y. Lettvin, H.R. Maturana, W.S. McCulloch, W.H. Pitts: What the frog's eye tells the frog's brain. *Proc. IRE*, Vol. 47, no. 11, pp. 1940–1951, Nov. 1959.
24. M.S. Livingstone: Art, illusion, and the visual system. *Scientific American*, Vol. 258, no. 1, pp. 68–75, Jan. 1988.
25. M.S. Livingstone, D.H. Hubel: Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, Vol. 240, pp. 740–749, 6 May 1988.
26. P. Maragos: A representation theory for morphological image and signal processing. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-11, no. 6, pp. 586–599, Jun. 1989.
27. D. Marr: Early processing of visual information. *Phil. Trans. Royal Soc. London*, Ser. B, Vol. 275, pp. 483–519, 1976.
27. D. Marr: Artificial intelligence — a personal view. *Artificial Intelligence*, Vol. 9, pp. 37–48, 1977.
29. D. Marr: *Vision*, W.H. Freeman & Co., San Francisco (1982).
30. D. Marr, E. Hildreth: Theory of edge detection. *Proc. Royal Soc. London*, Ser. B, Vol. 207, pp. 187–217, 1980.
31. G. Matheron: *Random Sets and Integral Geometry*, J. Wiley & Sons, New York (1975).
32. J.E.W. Mayhew, J.P. Frisby: Texture discrimination and Fourier analysis in human vision. *Nature*, Vol. 275, pp. 438–439, 5 Oct. 1978.
33. M. Mishkin, L.G. Ungerleider, K.A. Macko: Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, Vol. 6, pp. 414–417, Oct. 1983.
34. M.C. Morrone, J. Ross, D.C. Burr, R. Owens: Mach bands are phase dependent. *Nature*, Vol. 324, pp. 250–253, 20 Nov. 1986.
35. M.C. Morrone, D.C. Burr: Feature detection in human vision: a phase-dependent energy model. *Proc. Royal Soc. London*, Ser. B, Vol. 235, pp. 221–245, 1988.
36. P.M. Morse, K.U. Ingard: *Theoretical Acoustics*, McGraw-Hill, New York (1968).
37. A.V. Oppenheim, J.S. Lim: The importance of phase in signals. *Proc. IEEE*, Vol. 69, no. 5, pp. 529–541, May 1981.
38. T. Pavlidis: A critical survey of image analysis methods. *Proc. 8th. Internat. Conf. Pattern Recognition* (1986), pp. 502–511.
39. L.N. Piotrowski, F.W. Campbell: A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, Vol. 11, no. 3, pp. 227–346, 1982.

40. V.S. Ramachandran: Perceiving shape from shading. *Scientific American*, Vol. 259, no. 2, pp. 58–65, Aug. 1988.
41. V.S. Ramachandran, S.M. Anstis: The perception of apparent motion. *Scientific American*, Vol. 254, no. 6, pp. 80–87, Jun. 1986.
42. C. Ronse: Extraction of narrow peaks and ridges in images by combination of local low rank and max filters: implementation and applications to clinical angiography, *PRLB Working Document WD47*, (1988).
43. A. Rosenfeld, M. Thurston: Edge and curve detection for visual scene analysis. *IEEE Trans. Computers*, Vol. C-20, no. 5, pp. 562–569, May 1971.
44. M. Schmitt: Mathematical morphology and artificial intelligence: an automatic programming system. *Signal Processing*, Vol. 16, no. 4, pp. 389–401, Apr. 1989.
45. M.R. Schroeder: Models of hearing. *Proc. IEEE*, Vol. 63, no. 9, pp. 1332–1350, Sep. 1975.
46. J. Serra: *Image Analysis and Mathematical Morphology*, Academic Press, London (1982).
47. J. Serra: *Éléments de Théorie pour l'Optique Morphologique*, Ph. D. Thesis, Université P. & M. Curie, Paris (1986).
48. J. Serra: Morphological optics. *J. Microscopy*, Vol. 145, Pt. 1, pp. 1–22, 1987.
49. J. Serra, ed: *Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances*, Academic Press, London (1988).
50. S.R. Sternberg: Grayscale morphology, *Computer Vision, Graphics, and Image Processing*, Vol. 35, no. 3, pp. 333–355, 1986.
51. A.N. Tikhonov, V.Y. Arsenin: *Solutions of Ill-Posed Problems.* V.H. Winston & Sons, Washington (1977).
52. V. Torre, T.A. Poggio: On edge detection. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-8, no. 2, pp. 147–163, Mar. 1986.
53. D.C. Van Essen, J.H.R. Maunsell: Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, Vol. 6, pp. 370–375, Sep. 1983.
54. L.J. van Vliet, I.T. Young, G.L. Beckers: A nonlinear Laplace operator as edge detector in noisy images. *Computer Vision, Graphics, & Image Processing*, Vol. 45, no. 2, pp. 167–195, Feb. 1989.
55. H.R. Wilson, J.R. Bergen: A four mechanism model for threshold spatial vision. *Vision Research*, Vol. 19, pp. 19–32, 1979.
56. R. Wilson, G.H. Granlund: The uncertainty principle in image processing. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-6, no. 6, pp. 758–767, Nov. 1984.
57. R. Wilson, H. Knutsson: Uncertainty and inference in the visual system. *IEEE Trans. Systems, Man, & Cybernetics*, Vol. SMC-18, no. 2, pp. 305–312, Mar.-Apr. 1988.
58. A.P. Witkin: Scale-space filtering. *Proc. 8th. Internat. Joint Conf. Artificial Intelligence* (1983), pp. 1019–1021.
59. M.J. Wu: Fuzzy morphology and image analysis. *Proc. 9th. Internat. Conf. Pattern Recognition* (1988), pp.453–455.

60. A. Yuille, T.A. Poggio: Scaling theorems for zero crossings. *IEEE Trans. Pattern Analysis & Machine Intelligence*, Vol. PAMI-8, no. 1, pp. 15–25, Jan. 1986.
61. S.W. Zucker: Vision, early. *Encycl. Artificial Intelligence*, S.C. Shapiro & D. Eckroth eds., Wiley-Intersciences Publications, New York (1987), Vol. II, pp. 1131–1152.