

# Computational vision and regularization theory

Tomaso Poggio, Vincent Torre\* & Christof Koch

Artificial Intelligence Laboratory and Center for Biological Information Processing, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Massachusetts 02139, USA

\* Istituto di Fisica, Università di Genova, Genova, Italy

*Descriptions of physical properties of visible surfaces, such as their distance and the presence of edges, must be recovered from the primary image data. Computational vision aims to understand how such descriptions can be obtained from inherently ambiguous and noisy data. A recent development in this field sees early vision as a set of ill-posed problems, which can be solved by the use of regularization methods. These lead to algorithms and parallel analog circuits that can solve 'ill-posed problems' and which are suggestive of neural equivalents in the brain.*

COMPUTATIONAL vision denotes a new field in artificial intelligence, centred on theoretical studies of visual information processing. Its two main goals are to develop image understanding systems, which automatically construct scene descriptions from image input data, and to understand human vision.

Early vision is the set of visual modules that aim to extract the physical properties of the surfaces around the viewer, that is, distance, surface orientation and material properties (reflectance, colour, texture). Much current research has analysed processes in early vision because the inputs and the goals of the computation can be well characterized at this stage (see refs 1–4 for reviews). Several problems have been solved and several specific algorithms have been successfully developed. Examples are stereomatching, the computation of the optical flow, structure from motion, shape from shading and surface reconstruction.

A new theoretical development has now emerged that unifies much of these results within a single framework. The approach has its roots in the recognition of a common structure of early vision problems. Problems in early vision are 'ill-posed', requiring specific algorithms and parallel hardware. Here we introduce a specific regularization approach, and discuss its implications for computer vision and parallel computer architectures, including parallel hardware that could be used by biological visual systems.

## Early vision processes

Early vision consists of a set of processes that recover physical properties of the visible three-dimensional surfaces from the two-dimensional intensity arrays. Their combined output roughly corresponds to Marr's 2.5D sketch<sup>1</sup>, and to Barrow and Tennenbaum's intrinsic images<sup>2</sup>. Recently, it has been customary to assume that these early vision processes are general and do not require domain-dependent knowledge, but only

generic constraints about the physical world and the imaging stage (see box). They represent conceptually independent modules that can be studied, to a first approximation, in isolation. Information from the different processes, however, has to be combined. Furthermore, different modules may interact early on. Finally, the processing cannot be purely 'bottom-up': specific knowledge may trickle down to the point of influencing some of the very first steps in visual information processing.

Computational theories of early vision modules typically deal with the dual issues of representation and process. They must specify the form of the input and the desired output (the representation) and provide the algorithms that transform one into the other (the process). Here we focus on the issue of processes and algorithms for which we describe the unifying theoretical framework of regularization theories. We do not consider the equally important problem of the primitive tokens that represent the input of each specific process.

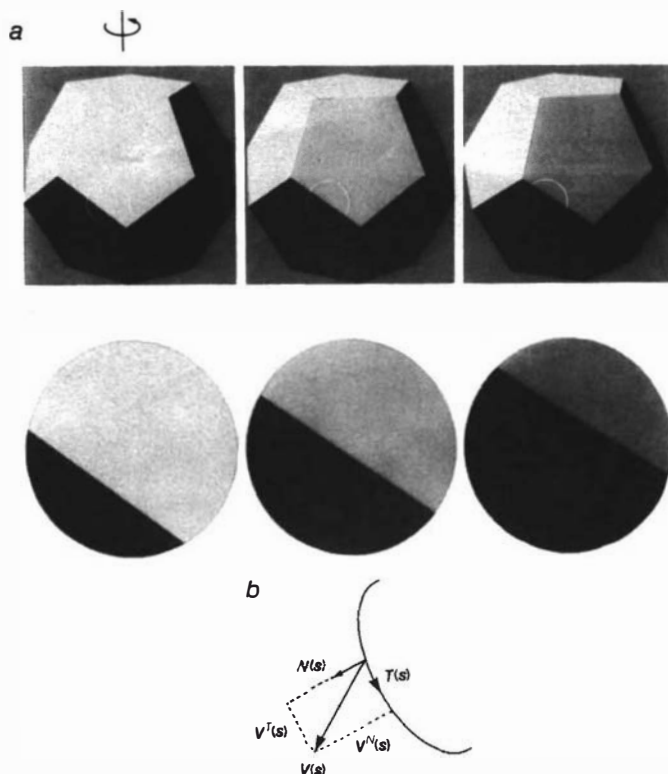
A good definition of early vision is that it is inverse optics. In classical optics or in computer graphics the basic problem is to determine the images of three-dimensional objects, whereas vision is confronted with the inverse problem of recovering surfaces from images. As so much information is lost during the imaging process that projects the three-dimensional world into the two-dimensional images, vision must often rely on natural constraints, that is, assumptions about the physical world, to derive unambiguous output. The identification and use of such constraints is a recurring theme in the analysis of specific vision problems.

Two important problems in early vision are the computation of motion and the detection of sharp changes in image intensity (for detecting physical edges). They illustrate well the difficulty of the problems of early vision. The computation of the two-dimensional field of velocities in the image is a critical step in several schemes for recovering the motion and the three-dimensional structure of objects. Consider the problem of determining the velocity vector  $V$  at each point along a smooth contour in the image. Following Marr and Ullman<sup>6</sup>, one can assume that the contour corresponds to locations of significant intensity change. Figure 1 shows how the local velocity vector is decomposed into a normal and a tangential component to the curve. Local motion measurements provide only the normal component of velocity. The tangential component remains 'invisible' to purely local measurements (unless they refer to some discontinuous features of the contour such as a corner). The problem of estimating the full velocity field is thus, in general, underdetermined by the measurements that are directly available from the image. The measurement of the optical flow is inherently ambiguous. It can be made unique only by adding information or assumptions.

The difficulties of the problem of edge detection are somewhat different. Edge detection denotes the process of identifying the

### Examples of early vision processes

- Edge detection
- Spatio-temporal interpolation and approximation
- Computation of optical flow
- Computation of lightness and albedo
- Shape from contours
- Shape from texture
- Shape from shading
- Binocular stereo matching
- Structure from motion
- Structure from stereo
- Surface reconstruction
- Computation of surface colour



**Fig. 1** Ambiguity of the velocity field. *a*, Local measurements cannot measure the full velocity field in the image plane, originated here by three-dimensional rotation of a solid object (three frames are shown). Any process operating within the aperture (shown as a white circle) can compute only the component of motion perpendicular to the contour. *b*, Decomposition of the velocity vector along the contour, parametrized by the arc length  $s$  into components normal ( $V^N(s)$ ) and tangential ( $V^T(s)$ ) to the curve. The computer drawing was kindly provided by Karl Sims.

physical boundaries of three-dimensional surfaces from intensity changes in their image. What is usually intended with edge detection is a first step towards this goal, that is, detecting and localizing sharp changes in image intensity. This is a problem of numerical differentiation of image data, which is plagued by the noise unavoidable during the imaging and the sampling processes. Differentiation amplifies noise and this process is thus inherently unstable. Figure 3 shows an example of an edge profile and its second derivative, where noise is significantly amplified. Most problems in early vision present similar difficulties. They are mostly underconstrained, as in the computation of the optical flow, or not robust against noise, as in edge detection.

## Ill-posed problems

The common characteristics of most early vision problems (in a sense, their deep structure) can be formalized: **most early vision problems are ill-posed problems in the precise sense defined by Hadamard<sup>7,8</sup>. This claim captures the importance of constraints and reflects the definition of vision as inverse optics.**

**Hadamard first introduced the definition of ill-posedness in the field of partial differential equations<sup>9</sup>.** Although ill-posed problems have been considered for many years as almost exclusively mathematical curiosities, it is now clear that many ill-posed problems, typically inverse problems, are of great practical interest (for instance, computer tomography). A problem is well-posed when its solution exists, is unique and depends continuously on the initial data. Ill-posed problems fail to satisfy one or more of these criteria. Note that the third condition does not imply that the solution is robust against noise in practice. For this, the problem must not only be well-posed but also be well conditioned to ensure numerical stability<sup>10</sup>.

It is easy to show formally that several problems in early vision are ill-posed in the sense of Hadamard<sup>8</sup>: stereo matching, structure from motion, computation of the optical flow, edge detection, shape from shading, the computation of lightness and surface reconstruction. Computation of the optical flow is ill-posed because the 'inverse' problem of recovering the full velocity field from its normal component along a contour fails to satisfy the uniqueness condition. Edge detection, intended as numerical differentiation, is ill-posed because the solution does not depend continuously on the data.

**The main idea for 'solving' ill-posed problems, that is for restoring 'well-posedness', is to restrict the class of admissible solutions by introducing suitable *a priori* knowledge. *A priori* knowledge can be exploited, for example, under the form of either variational principles that impose constraints on the possible solutions or as statistical properties of the solution space.** We will use the general term regularization for any method used to make an ill-posed problem well-posed. **Variational regularization will indicate the regularization methods that reformulate an ill-posed problem in terms of a variational principle.** We will next outline specific variational methods that we will denote as the standard regularization methods, attributable mainly to **Tikhonov<sup>11,12</sup>** (see also refs 13, 14). We will also outline future extensions of the standard theory from the perspective of early vision.

**The regularization of the ill-posed problem of finding  $z$  from the 'data'  $y$**

$$Az = y \quad (1)$$

**requires the choice of norms  $\|\cdot\|$  and of a stabilizing functional  $\|Pz\|$ . In standard regularization theory,  $A$  is a linear operator, the norms are quadratic and  $P$  is linear. Two methods that can be applied are<sup>8,13</sup>: (1) among  $z$  that satisfy  $\|Az - y\| < \epsilon$  find  $z$  that minimizes  $\epsilon$  depends on the estimated measurement errors and is zero if the data are noiseless)**

$$\|Pz\|^2 \quad (2)$$

**(2) find  $z$  that minimizes**

$$\|Az - y\|^2 + \lambda \|Pz\|^2 \quad (3)$$

**where  $\lambda$  is a so-called regularization parameter.**

**The first method computes the function  $z$  that is sufficiently close to the data and is most 'regular', that is minimizes the 'criterion'  $\|Pz\|$ . In the second method,  $\lambda$  controls the compromise between the degree of regularization of the solution and its closeness to the data. Standard regularization theory provides techniques for determining the best  $\lambda$ <sup>12,15</sup>. Thus, standard regularization methods impose the constraints on the problem by a variational principle, such as the cost functional of equation (3). The cost that is minimized reflects physical constraints about what represents a good solution: it has to be both close to the data and regular by making the quantity  $\|Pz\|^2$  small.  $P$  embodies the physical constraints of the problem. It can be shown**

for quadratic variational principles that under mild conditions the solution space is convex and a unique solution exists. It must be pointed out that standard regularization methods have to be applied after a careful analysis of the ill-posed nature

of the problem. The choice of the norm  $\|\cdot\|$ , of the stabilizing functional  $\|Pz\|$  and of the functional spaces involved is dictated both by mathematical properties and by physical plausibility. They determine whether the precise conditions for a correct regularization hold for any specific case.

Variational principles are used widely in physics, economics and engineering. In physics, for instance, most of the basic laws have a compact formulation in terms of variational principles that require minimization of a suitable functional, such as the energy or the Lagrangian.

## Examples

Variational principles of the form of equation (3) have been used in the past in early vision<sup>16-25</sup>. Other problems have now been approached in terms of standard regularization methods

**Table 1** Regularization in early vision

Problem	Regularization principle
Edge detection	$\int [(Sf - i)^2 + \lambda (f_{xx})^2] dx$
Optical flow (area based)	$\int [i_x u + i_y v + i_t]^2 + \lambda (u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy$
Optical flow (contour based)	$\int [(V \cdot N - V^N)^2 + \lambda ((\partial/\partial_s)V)^2] ds$
Surface reconstruction	$\int [S \cdot f - d]^2 + \lambda (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2) dx dy$
Spatiotemporal approximation	$\int [(S \cdot f - i)^2 + \lambda (\nabla f \cdot V + ft)^2] dx dy dt$
Colour	$\ I^v - Az\ ^2 + \lambda \ Pz\ ^2$
Shape from shading	$\int [(E - R(f, g))^2 + \lambda (f_x^2 + f_y^2 + g_x^2 + g_y^2)] dx dy$
Stereo	$\int \{[\nabla^2 G * (L(x, y) - R(x + d(x, y), y))]^2 + \lambda (\nabla d)^2\} dx dy$

Some of the early vision problems that have been solved in terms of variational principles. The first five are standard quadratic regularization principles. In edge detection<sup>26,27</sup> the data on image intensity ( $i = i(x)$ ) (for simplicity in one dimension) are given on a discrete lattice: the operator  $S$  is the sampling operator on the continuous distribution  $f$  to be recovered. A similar functional may be used to approximate time-varying imagery. The spatio-temporal intensity to be recovered from the data  $i(x, y, t)$  is  $f(x, y, t)$ ; the stabilizer imposes the constraint of constant velocity  $V$  in the image plane (ref. 61). In area-based optical flow<sup>18</sup>,  $i$  is the image intensity,  $u$  and  $v$  are the two components of the velocity field. In surface reconstruction<sup>21,22</sup> the surface  $f(x, y)$  is computed from sparse depth data  $d(x, y)$ . In the case of colour<sup>32</sup> the brightness is measured on each of three appropriate colour coordinates  $I^v$  ( $v = 1, 2, 3$ ). The solution vector  $z$  contains the illumination and the albedo components separately; it is mapped by  $A$  into the ideal data. Minimization of an appropriate stabilizer enforces the constraint of spatially smooth illumination and either constant or sharply varying albedo. For shape from shading<sup>19</sup> and stereo (T.P. and A. Yuille, unpublished), we show two non-quadratic regularization functionals.  $R$  is the reflectance map,  $f$  and  $g$  are related to the components of the surface gradient,  $E$  is the brightness distribution<sup>19</sup>. The regularization of the disparity field  $d$  involves convolution with the laplacian of a gaussian of the left ( $L$ ) and the right ( $R$ ) images and a Tikhonov stabilizer corresponding to the disparity gradient.

(see Table 1). Most stabilizing functionals used so far in early vision are of the Tikhonov type, being linear combinations of the first  $p$  derivatives of the desired solution  $z$  (ref. 12). The solutions arising from these stabilizers correspond to either interpolating or approximating splines. We return now to our examples of motion and edge detection, and show how standard regularization techniques can be applied.

Intuitively, the set of measurements of the normal component of velocity over an extended contour should provide considerable constraint on the global motion of the contour. Some additional assumptions about the nature of the real world are needed, however, in order to combine local measurements at different locations. For instance, the assumption of rigid motion on the image plane is sufficient to determine  $V$  uniquely<sup>23,24</sup>. In this case, local measurements of the normal component at different locations can be used directly to find the optical flow, which is the same everywhere. The assumption, however, is overly restrictive, because it does not cover the case of motion of a rigid object in three-dimensional space (see Fig. 1). Hildreth suggested<sup>23,24</sup>, following Horn and Schunck<sup>18</sup>, a more general smoothness constraint on the velocity field. The underlying physical consideration is that the real world consists of solid objects with smooth surfaces, whose projected velocity field is usually smooth. The specific form of the stabilizer (a Tikhonov stabilizer) was dictated by mathematical considerations, especially uniqueness of the solution. The two regularizing methods correspond to the two algorithms proposed and implemented by Hildreth<sup>23</sup>. The first one, which assumes that the

measurements of the normal velocity components  $V^N(s)$  are exact, minimizes

$$\|PV\|^2 = \int \left( \frac{\partial V}{\partial s} \right)^2 ds \quad (4)$$

subject to the measurements of the normal component of velocity (where  $s$  is arc length). The integral is evaluated along the contour. For non-exact data the second method provides the solution by minimizing

$$\|V \cdot N - V^N\|^2 + \lambda \int \left( \frac{\partial V}{\partial s} \right)^2 ds \quad (5)$$

where  $N$  is the normal unit vector to the contour and  $\lambda^{-1}$  expresses the reliability of the data. Figure 2a shows an example of a successful computation of the optical flow by the first algorithm.

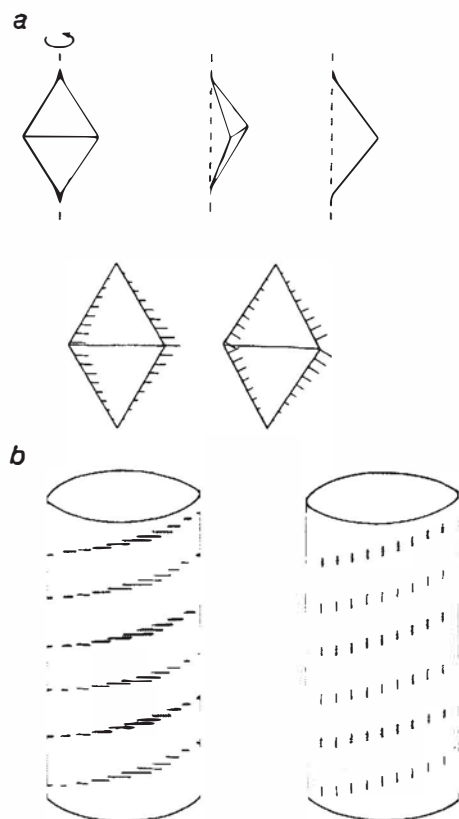
Recently, regularization techniques have been applied to edge detection<sup>26,27</sup>. The problem of numerical differentiation can be regularized by the second method with a Tikhonov stabilizer that reflects a constraint of smoothness on the image (see Table 1). The physical justification is that the image is an analytical function with bounded derivatives, because of the band-limiting properties of the optics that cuts off high spatial frequencies. This regularized solution is equivalent, under mild conditions, to convolving the intensity data with the derivative of a filter similar to the gaussian<sup>26</sup> (see Fig. 3), proposed earlier<sup>28,29</sup>.

Other early vision problems can be solved by standard regularization techniques. Surface reconstruction, for example, can be performed from a sparse set of depth values by imposing smoothness of the surface<sup>20-22</sup>. Optical flow can be computed at each point in the image, rather than along a contour, using a constraint of smooth variation, in the form of a Tikhonov stabilizer<sup>17</sup>. Variational principles that are not exactly quadratic but have the form of equation (3) can be used for other problems in early vision. The main results of Tikhonov can in fact be extended to the case in which the operators  $A$  and  $P$  are nonlinear, provided they satisfy certain conditions<sup>30</sup>. The variation of an object's brightness gives clues to its shape: the surface orientation can be computed from an intensity image in terms of the variational principle shown in Table 1, which penalizes orientations violating the smoothness constraint and the irradiance constraint<sup>18</sup>. Stereo matching is the problem of inferring the correct binocular disparity (and therefore depth) from a pair of binocular images, by finding which feature in one image corresponds to which feature in the other image. This is an ill-posed problem which, under some restrictive conditions corresponding to the absence of occlusions, can be regularized by a variational principle that contains a term measuring the discrepancy between the feature maps extracted from the two images and a stabilizer that penalizes large disparity gradients (see Table 1) and effectively imposes a disparity gradient limit. The algorithm can reduce to an area-based correlation algorithm of the Nishihara type<sup>31</sup> if the disparity gradient is small. A standard regularization principle has been proposed for solving the problem of separating a material reflectance from a spatially varying illumination in colour images<sup>32</sup>. The algorithm addresses the problem known in visual psychophysics as colour constancy<sup>33</sup>.

## Physical plausibility and illusions

Physical plausibility of the solution, rather than its uniqueness, is the most important concern in regularization analysis. A physical analysis of the problem, and of its significant constraints, plays the main role<sup>8</sup>. The *a priori* assumptions required to solve ill-posed problems may be violated in specific instances where the regularized solution does not correspond to the physical solution. The algorithm suffers an optical illusion. A good example is provided by the computation of motion. The smoothness assumption of equation (5) gives correct results under some general conditions (for example, when objects have images consisting of connected straight lines<sup>34</sup>). For some classes of motion and contours, the smoothness principle will not yield





**Fig. 2** Computing the smoothest velocity field along contours. *a*, Three-dimensional stimulus first used by Wallach<sup>62</sup> to demonstrate the ability of the human visual system to derive three-dimensional structure from the projected two-dimensional motion of an object (kinetic depth effect). The top part shows three views of a figure as it is rotated around the vertical axis. The initial measurements of the normal velocity components  $V_i^N$  are shown on the lower right. The velocity field computed using equation (4) is shown on the lower left. The final solution corresponds to the physical correct velocity distribution. Recent electrophysiological evidence implicates the middle temporal area of the monkey as a site where a similar motion integration may occur<sup>63</sup>. *b*, Circular helix on an imaginary three-dimensional cylinder, rotating about its vertical axis (barber pole). The projection of the curve onto the image plane, together with the resulting two-dimensional velocity vectors are drawn on the left. Although the true velocity field  $V$  is strictly horizontal (left), the smoothest velocity field (right) is vertical. This example illustrates a case where both the algorithm and the human visual system suffer the same optical illusion. Adapted from ref. 23.

the correct velocity field. In several of these cases, however, the human visual system also seems to derive a similar, incorrect velocity field, thereby possibly revealing *a priori* assumptions the brain is making about the world. A striking instance is the barber-pole illusion<sup>23</sup> (illustrated in Fig. 2*b*).

### Analog networks

One of the mysteries of biological vision is its speed. Parallel processing has often been advocated as the answer to this problem. The model of computation provided by digital processes is, however, unsatisfactory, especially given the increasing evidence that neurones are complex devices, very different from simple digital switches. It is, therefore, interesting to consider whether the regularization approach to early vision may lead to a different type of parallel computation. We have recently suggested that linear, analog networks (either electrical or chemical) are, in fact, a natural way of solving the variational principles dictated by standard regularization theory<sup>7</sup> (see also refs 22, 35).

The fundamental reason for such a mapping between variational principles and electrical or chemical networks is Hamilton's least action principle. The class of variational principles

that can be computed by analog networks is given by Kirchhoff's current and voltage laws, which represent conservation and continuity restrictions satisfied by each network component (appropriate variables are usually voltage and current for electrical networks and affinity and turnover rate for chemical systems<sup>36</sup>; see also ref. 37). There is in general no unique network but possibly many networks implementing the same variational principle. For example, graded networks of the type proposed by Hopfield in the context of associative memory<sup>38</sup> can solve standard regularization principles<sup>39</sup>.

From Kirchhoff's law, it can be proved<sup>7</sup> that for every quadratic variational problem with a unique solution (which is usually the case<sup>8</sup>), there exists a corresponding electrical network consisting of resistances and voltage or current sources having the same solution. In other words, the steady-state current (or voltage) distribution in the network corresponds to the solution, for example to the tangential velocity distribution  $V^T(s)$ , of the standard regularization problem (Fig. 4). Furthermore, when capacitances are added to the system, thereby introducing dynamics, the system is stable. The data are supplied by injecting currents or by introducing batteries, that is by constant current or voltage sources<sup>7</sup>.

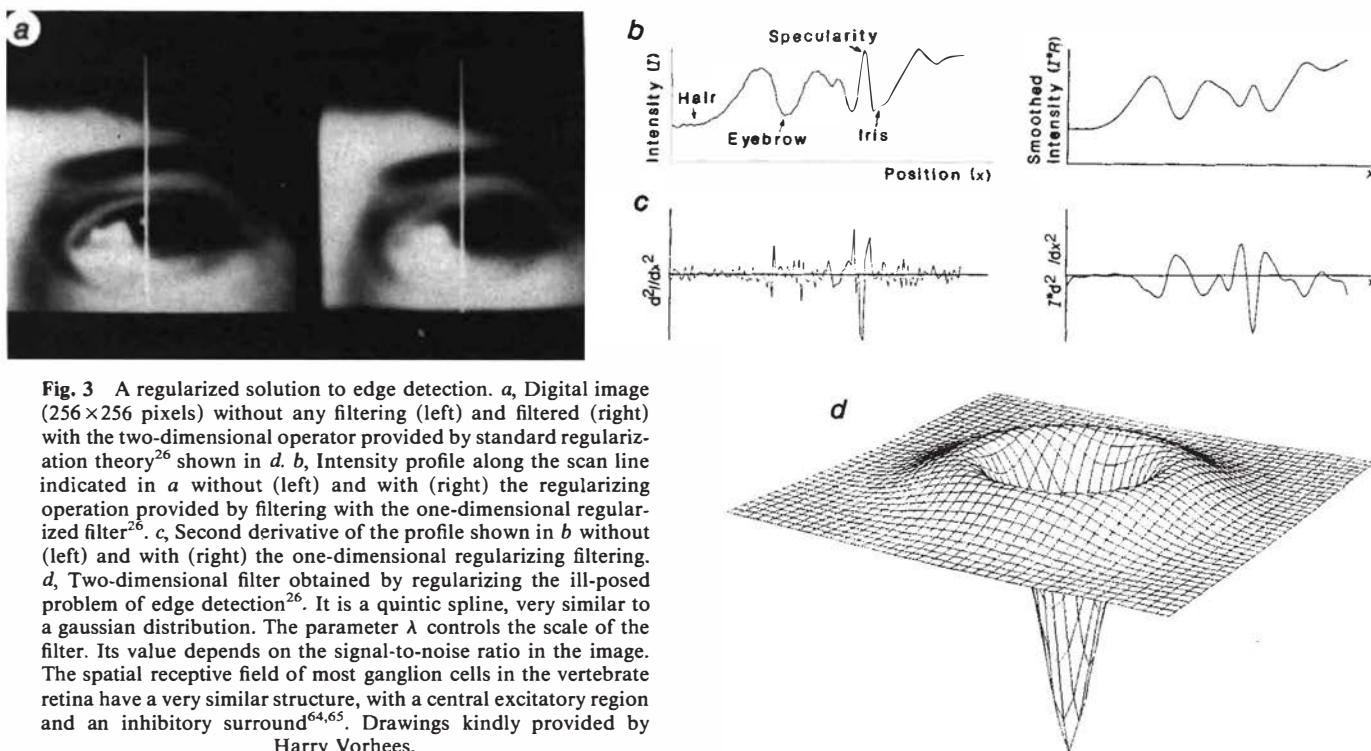
This analog parallel model of computation is especially interesting from the point of view of the present understanding of the biophysics of neurones, membranes and synapses. Increasing evidence shows that electrotonic potentials play a primary role in many neurones<sup>40</sup>. Mechanisms as diverse as dendrodendritic synapses<sup>41,42</sup>, gap junctions<sup>43</sup>, neurotransmitters acting over different times and distances<sup>44</sup>, voltage-dependent channels that can be modulated by neuropeptides<sup>45</sup> and interactions between synaptic conductance changes<sup>46</sup> provide neurones with various different circuit elements. Patches of neural membrane are equivalent to resistances, capacitances and phenomenological inductances<sup>47</sup>. Synapses on dendritic spines mimic voltage sources, whereas synapses on thick dendrites or the soma act as current sources<sup>48,49</sup>. Thus, single neurones or small networks of neurones could implement analog solutions of regularization principles. Hypothetical neuronal implementations of the analog circuits of Fig. 4 have been devised, involving only one or two separate dendrites<sup>7</sup>.

### Beyond standard regularization theory

The new theoretical framework for early vision clearly shows the attractions and the limitations that are intrinsic to the standard Tikhonov form of regularization theory. The main problem is the degree of smoothness required for the unknown function that has to be recovered. For instance, in surface interpolation, the degree of smoothness corresponding to the so-called thin-plate splines smooths depth discontinuities too much, and often leads to unrealistic results<sup>20</sup> (discontinuities may, however, be detected and then used in a second regularization step<sup>66</sup>).

Standard regularization theory deals with linear problems and is based on quadratic stabilizers. It leads therefore to the minimization of quadratic functionals and to linear Euler-Lagrange equations. Non-quadratic functionals may be needed to enforce the correct physical constraints (Table 1 shows the non-quadratic case of shape-from-shading). Even in this case, methods of standard regularization theory can be used<sup>30</sup>, but the solution space is no longer convex and many local minima can be found in the process of minimization.

A non-quadratic stabilizer has been proposed for the problem of preserving discontinuities in the reconstruction of surfaces from depth data<sup>50</sup>. The stabilizer, in its basic form attributable to Geman and Geman<sup>51</sup> (a similar principle but without a rigorous justification was proposed by Blake<sup>52</sup>; see also the variational continuity control of Terzopoulos<sup>67</sup>), embeds prior knowledge about the geometry of the discontinuities (the line process) and, in particular, that they are continuous and often straight contours. In standard regularization principles, the search space has only one local minimum to which suitable algorithms always converge. For non-quadratic functionals, the search space may be similar to a mountain range with many



**Fig. 3** A regularized solution to edge detection. *a*, Digital image ( $256 \times 256$  pixels) without any filtering (left) and filtered (right) with the two-dimensional operator provided by standard regularization theory<sup>26</sup> shown in *d*. *b*, Intensity profile along the scan line indicated in *a* without (left) and with (right) the regularizing operation provided by filtering with the one-dimensional regularized filter<sup>26</sup>. *c*, Second derivative of the profile shown in *b* without (left) and with (right) the one-dimensional regularizing filtering. *d*, Two-dimensional filter obtained by regularizing the ill-posed problem of edge detection<sup>26</sup>. It is a quintic spline, very similar to a Gaussian distribution. The parameter  $\lambda$  controls the scale of the filter. Its value depends on the signal-to-noise ratio in the image. The spatial receptive field of most ganglion cells in the vertebrate retina have a very similar structure, with a central excitatory region and an inhibitory surround<sup>64,65</sup>. Drawings kindly provided by Harry Vorhees.

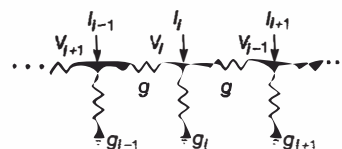
local minima. Stochastic algorithms for solving minimization problems of this type have been proposed recently, to escape from local minima at which simple hill-climbing algorithms would be trapped<sup>53–55</sup>. The basic idea is somewhat similar to adding a forcing noise term to the search algorithm. If the non-quadratic variational principle can be represented in a nonlinear analog network (as in ref. 39), an appropriate source of gaussian noise could drive the analog network. The dynamics of the system would then be described by a nonlinear stochastic differential equation, representing a diffusion process.

The challenge now for the regularization theory of vision is to extend it beyond standard regularization methods. The universe of computations that can be performed in terms of quadratic functionals is rather restricted. To see this, it is sufficient to realize that minimization of quadratic cost functionals leads to a linear regularization operator, that is, to a linear mapping of the input data into the solution space. In the special case when the data are on a regular grid and obey suitable conditions, the linear operator may become a convolution, that is, a simple filtering operation on the data. Similar to linear models in physics, standard regularization theory is an extremely useful approximation in many cases, but cannot deal with the full complexity of vision.

### Stochastic route to regularization

A different rigorous approach to regularization is based on Bayes estimation and Markov random fields models. In this approach the *a priori* knowledge is represented in terms of appropriate probability distributions, whereas in standard regularization *a priori* knowledge leads to restrictions on the solution space. Consider as an example the case of surface reconstruction. A *priori* knowledge can be formulated in terms of a Markov random field (MRF) model of the surface. In a MRF the value at one discrete location depends only on the values within a given neighbourhood. In this approach the best surface maximizes some likelihood criterion such as the maximum *a posteriori* estimate or the *a posteriori* mean of the MRF. It has been pointed out<sup>50</sup> that the maximum *a posteriori* estimate of a MRF is equivalent to a variational principle of the general form of equation (3); the first term measures the discrepancy between the data and the solution, the second term is now an arbitrary potential function of the solution (defined on a discrete lattice). The overall variational principle, in general not quadratic,

reduces to a quadratic functional of the standard regularization type when the noise is additive and gaussian and first-order differences of the field are zero-mean, independent, gaussian random variables. In this case the maximum *a posteriori* estimate (MAP) coincides with all estimates and, in particular, with the *a posteriori* mean. But Marroquin<sup>56</sup> has shown recently that this is not true in general: in most cases the MAP estimate is not optimal with respect to natural error measures and better estimates such as the *a posteriori* mean can be found. In these cases the problem is not equivalent to finding the global minimum of an energy functional: simulated annealing is not needed, and a Metropolis-type algorithm<sup>55</sup> can be used instead.



**Fig. 4** Analog networks. A resistive network computing the smoothest velocity field<sup>23</sup>. The network corresponds to the situation where the measurements of the normal velocity component  $V_i^N$  are assumed to be exact. Discretizing the associated variational equation (4) along the contour yields the Euler-Lagrange equations  $(2 + \kappa_i^2) V_i^T - V_{i+1}^T - V_{i-1}^T = d_i$ , where  $\kappa_i$  is the curvature of the contour at location  $i$ ,  $d_i$  is a function of the data  $V_i^N$  and the contour and  $V_i^T$  is the unknown tangential component of the velocity  $V$  at location  $i$  along the contour. The equation describing the  $i$ th node in the electrical circuit is  $(2g + g_i) V_i - g V_{i+1} - g_{i-1} V_{i-1} = I_i$ , where  $V_i$  is the voltage corresponding to the unknown  $V_i^T$ , and  $I_i$  is the injected current at node  $i$  depending on the measurement  $V_i^N$ . A slightly more complicated circuit can be designed for the case when the measurements of  $V_i^N$  are not exact<sup>7</sup> (equation (5)). Uniqueness of the regularized solution always ensures stability of the corresponding network, even if capacities are introduced. Equivalent analog networks can be implemented by diffusion-reaction systems, where the interaction between neighbouring locations are mimicked using diffusion or chemical reactions with first-order kinetics. Hypothetical neuronal implementations may be envisaged. The conductance  $g$  may correspond to a small segment of a dendrite, the variable conductance  $g_i$  to a synaptic input with a reversal potential close or equal to the resting potential of the dendrite (that is, silent or shunting inhibition) and the current source to a conventional chemical synapse injecting current  $I_i$  into the dendrite. The output is sampled at location  $i$  by a chemical synapse. Adapted from ref. 7.



In the case of Hildreth's motion computation<sup>23</sup> the smoothness assumption corresponds to the hypothesis that the changes in velocity between neighbouring points along the contour are zero-mean, independent, gaussian random variables. This connection between the stochastic approach and standard regularization methods gives an interesting perspective on the nature of the constraints and the choice of the stabilizer. The variational principles used to solve the inverse problems of vision correspond to the Markov structure that generates plausible solutions.

A related area of future investigation concerns the problem of learning a regularizing operator. In the case of standard regularization, the corresponding linear operator mapping the data into the solution may be learned by an associative learning scheme<sup>57</sup> of the type proposed in connection with biological memory<sup>58</sup>.

## Towards symbolic descriptions

So far, we have restricted our discussion to the early stages of vision that create image-like representations of the physical three-dimensional surfaces around the viewer. The step beyond these representations, also called intrinsic images<sup>5</sup>, or 2-1/2D sketches<sup>1</sup>, is a large one. Intrinsic images are still image-like numerical representations, not yet described in terms of objects. They are already sufficient for some of the high-level tasks of a vision system such as manipulation and navigation. They cannot be used directly for the tasks of recognition and description that require the generation and use of more symbolic representations. It seems at first difficult to see how the computation of symbolic representations may fit at all in the perspective of regularizing ill-posed problems.

The basic idea of all regularization methods is to restrict the space of possible solutions. If this space is constrained to have finite dimensions, there is a good chance that an inverse problem will be well-posed. Thus, a representation based on a finite set of discrete symbols regularizes a possibly ill-posed problem. From this point of view, the problem of perception (regularizing an otherwise underconstrained problem using generic constraints of the physical world) becomes practically equivalent to the classical artificial intelligence problem of solving and inference, that is, of finding ways of solving intractable problems (such as chess) by limiting the search for solutions.

## Conclusions

We suggest a classification of vision algorithms that maps naturally into parallel digital computer architectures now under development. Standard regularization, when sufficient, leads to two classes of parallel algorithms. Algorithms for finding minima of a convex functional such as steepest descent or the more efficient multigrid algorithms developed for vision<sup>59</sup> can always be used. They can be replaced by convolution algorithms if the data are given on a regular grid and  $A$  in equation (1) is space invariant. In the latter case, the regularized solution is obtained by convolving the data through a precomputed filter.

All these algorithms may be implemented by parallel architectures of many processors with only local connections. Problems that cannot be approached in terms of regularization and that require symbolic representations and operations on them, may need parallel architectures with a global communication facility, such as the Connection Machine currently under development<sup>60</sup>.

The concept of ill-posed problems and the associated old and new regularization theories seem to provide a satisfactory theoretical framework for much of early vision. This new perspective also provides a link between the computational (ill-posed) nature of early vision problems, the structure of the algorithms for solving them and the parallel hardware that can be used for efficient visual information processing. It also shows the intrinsic limitations of the variational principles used so far in early vision, indicating at the same time how to extend regularization analysis beyond the standard theory.

We thank E. Hildreth, A. Hurlbert, J. Marroquin, G. Mitchison, D. Terzopoulos, H. Voorhees and A. Yuille for discussions and suggestions. Mario Bertero first pointed out to us that numerical differentiation is an ill-posed problem. E. Hildreth, L. Ardrey and especially H. Voorhees, K. Sims and M. Drumheller helped with some of the figures. Support for the Artificial Intelligence Laboratory's research in artificial intelligence is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505. The Center for Biological Information Processing is supported in part by the Sloan Foundation and in part by Whitaker College. C.K. is supported by a grant from the Office of Naval Research, Engineering Psychology Division.

- Marr, D. *Vision* (Freeman, San Francisco, 1982).
- Brady, J. M. *Computing Surv.* **14**, 3-71 (1982).
- Ballard, D. H., Hinton, G. E. & Sejnowski, T. J. *Nature* **306**, 21-26 (1983).
- Brown, C. M. *Science* **224**, 1299-1305 (1984).
- Barrow, H. G. & Tenenbaum, J. M. *Artif. Intell.* **17**, 75-117 (1981).
- Marr, D. & Ullman, S. *Proc. R. Soc. B211*, 151-180 (1981).
- Poggio, T. & Koch, C. *Proc. R. Soc. B* (in the press).
- Poggio, T. & Torre, V. *Artif. Intell. Lab. Memo No. 773* (MIT, Cambridge, 1984).
- Hadamard, J. *Lectures on the Cauchy Problem in Linear Partial Differential Equations* (Yale University Press, 1923).
- Bertero, M., Del Mol, C. & Pike, E. R. *J. inverse Prob.* (in the press).
- Tikhonov, A. N. *Sov. Math. Dokl.* **4**, 1035-1038 (1963).
- Tikhonov, A. N. & Arsenin, V. Y. *Solutions of Ill-posed Problems* (Winston, Washington, DC, 1977).
- Bertero, M. in *Problema non ben posti ed inversi* (Istituto di Analisi Globale, Firenze, 1982).
- Nashed, M. Z. (ed.) *Generalized Inverses and Applications* (Academic, New York, 1976).
- Wahba, G. *Tech. Rep. No. 595* (University of Wisconsin, 1980).
- Horn, B. K. P. *Computer Graphics Image Processing* **3**, 111-299 (1974).
- Horn, B. K. P. *Robot Vision* (MIT Press & McGraw-Hill, Cambridge & New York, 1985).
- Horn, B. K. P. & Schunck, B. G. *Artif. Intell.* **17**, 185-203 (1981).
- Ikeuchi, K. & Horn, B. K. P. *Artif. Intell.* **17**, 141-184 (1981).
- Grimson, W. E. L. *From Images to Surfaces: A Computational Study of the Human Early Visual System* (MIT, Cambridge, 1981).
- Grimson, W. E. L. *Phil. Trans. R. Soc. B298*, 395-427 (1982).
- Terzopoulos, D. *Computer Graphics Image Processing* **24**, 52-96 (1983).
- Hildreth, E. C. *The Measurement of Visual Motion* (MIT Press, Cambridge, 1984).
- Hildreth, E. C. *Proc. R. Soc. B221*, 189-220 (1984).
- Horn, B. K. P. & Brooks, M. J. *Artif. Intell. Lab. Memo No. 813* (MIT, Cambridge, 1985).
- Poggio, T., Voorhees, H. & Yuille, A. *Artif. Intell. Lab. Memo No. 833* (MIT, Cambridge, 1985).
- Torre, V. & Poggio, T. *IEEE Trans. Pattern Analysis Machine Intelligence* (in the press).
- Marr, D. & Poggio, T. *Proc. R. Soc. B204*, 301-328 (1979).
- Marr, D. & Hildreth, E. C. *Proc. R. Soc. B207*, 187-217 (1980).
- Morozov, V. A. *Methods for Solving Incorrectly Posed Problems* (Springer, New York, 1984).
- Nishihara, H. K. *Artif. Intell. Lab. Memo No. 780* (MIT, Cambridge, 1984).
- Hurlbert, A. *Artif. Intell. Lab. Memo No. 814* (MIT, Cambridge, 1985).
- Land, E. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5163-5169 (1984).
- Yuille, A. *Artif. Intell. Lab. Memo No. 724* (MIT, Cambridge, 1983); *Advances in Artificial Intelligence* (ed. O'Shea, T. M. M.) (Elsevier, Amsterdam, in the press).
- Ullman, S. *Computer Graphics Image Processing* **9**, 115-125 (1979).
- Eigen, M. in *The Neurosciences: 3rd Study Program* (eds Schmitt, F. O. & Worden, F. G.) xix-xxviii (MIT Press, Cambridge, 1974).
- Oster, G. F., Perelson, A. & Katchalsky, A. *Nature* **234**, 393-399 (1971).
- Hopfield, J. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3088-3092 (1984).
- Koch, C., Marroquin, J. & Yuille, A. *Artif. Intell. Lab. Memo No. 751* (MIT, Cambridge, 1985).
- Schmitt, F. O., Dev, P. & Smith, B. H. *Science* **193**, 114-120 (1976).
- Gaubard, K. & Calvin, W. H. in *The Neurosciences: 4th Study Program* (eds Schmitt, F. O. & Worden, F. G.) 317-332 (MIT Press, Cambridge, 1979).
- Shepherd, G. M. & Brayton, R. K. *Brain Res.* **175**, 377-382 (1979).
- Bennett, M. V. L. in *Handbook of Physiology*, 221-250 (American Physiological Society, Bethesda, 1977).
- Marder, E. *Trends Neurosci.* **7**, 48-53 (1984).
- Schmitt, F. D. *Neuroscience* **13**, 991-1002 (1984).
- Koch, C., Poggio, T. & Torre, V. *Phil. Trans. R. Soc. B298*, 227-268 (1982).
- Cole, K. S. *Membranes, Ions and Impulses* (University of California Press, Berkeley, 1968).
- Jack, J. J., Noble, D. & Tsien, R. W. *Electric Current Flow in Excitable Cells* (Clarendon, Oxford, 1975).
- Koch, C. & Poggio, T. *Proc. R. Soc. B218*, 455-477 (1983).
- Marroquin, J. *Artif. Intell. Lab. Memo No. 792* (MIT, Cambridge, 1984).
- Geman, S. & Geman, D. *IEEE Trans. Pattern Analysis Machine Intelligence* **6**, 721-741 (1984).
- Blake, A. *Pattern Recognition Lett.* **1**, 393-399 (1983).
- Hinton, G. E. & Sejnowski, T. J. *Proc. IEEE 1983 Conf. Computer Vision and Pattern Recognition* (Washington, DC, 1983).
- Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. *Science* **220**, 671-680 (1983).
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. *J. chem. Phys.* **21**(6), 1087-1092 (1953).
- Marroquin, J. *Artif. Intell. Lab. Memo No. 839* (MIT, Cambridge, 1985).
- Poggio, T. & Hurlbert, A. *Artif. Intell. Lab. Working Pap. No. 264* (MIT, Cambridge, 1984).
- Kohonen, T. *Self-Organization and Associative Memory* (Springer, Berlin, 1984).
- Terzopoulos, D. *IEEE Trans. Pattern Analysis Machine Intelligence* (in the press).
- Hillis, W. D. *The Connection Machine* (MIT Press, Cambridge, 1985).
- Fahle, M. & Poggio, T. *Proc. R. Soc. B213*, 451-477 (1981).
- Wallach, H. & O'Connell, D. N. *J. exp. Psychol.* **45**, 205-217 (1953).
- Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. in *Pattern Recognition Mechanisms* (eds Chagas, C., Gattar, R. & Gross, C. G.) 95-107 (Vatican, Rome, 1984); *Expl Brain Res.* (in the press).
- Barlow, H. B. *J. Physiol., Lond.* **119**, 69-88 (1953).
- Kuffler, S. W. *J. Neurophysiol.* **16**, 37-68 (1953).
- Terzopoulos, D. thesis, Massachusetts Inst. Technol. (1948).
- Terzopoulos, D. *Artif. Intell. Lab. Memo No. 800* (MIT, Cambridge, 1985).

# 計算論的視覚と正則化理論

## 要旨

目に見える表面の物理的特性 (距離やエッジの有無など) の記述は、一次画像データから復元する必要がある。計算論的視覚は、本質的に曖昧でノイズの多いデータから、どのようにしてそのような記述を得ることができるかを理解することを目的としている。この分野の最近の進展は、初期視覚を、正則化法を用いて解決できる一連の不良設定問題として捉えている。これにより、「不良設定問題」を解くことができるアルゴリズムや並列アナログ回路が開発され、脳内の神経に相当する部分が示唆されている。

計算論的視覚は、視覚情報処理の理論的研究を中心とした人工知能の新しい分野である。画像の入力データから自動的に情景描写を行う画像理解システムの開発と、人間の視覚を理解することを目的としている。

初期視覚とは、見る人の周囲にある表面の物理的特性、すなわち、距離、表面の向き、材料特性 (反射率、色、テクスチャ) を抽出することを目的とした一連の視覚モジュールである。現在の研究の多くは、初期視覚の処理過程を分析している。なぜなら、この段階では、入力と計算の目標が十分に特徴づけられるからである (レビューは文献 1~4 を参照)。いくつかの問題が解決され、いくつかの具体的なアルゴリズムの開発に成功している。例えば、両眼視対応、光学フローの計算、動きからの構造、陰影からの形状、表面再構成などである。

現在、これらの結果の多くを 1 つの枠組みに統合する新しい理論的展開が生まれている。このアプローチは、初期視覚問題に共通する構造を認識したことに端を発している。初期視覚問題は、特定のアルゴリズムと並列ハードウェアを必要とする「不良設定問題」である。ここでは、特定の正則化アプローチを紹介し、コンピュータビジョンと並列コンピュータアーキテクチャ (生物学的視覚系で使用される可能性のある並列ハードウェアを含む) への影響を議論する。

## 1. 初期視覚過程

初期視覚は、2 次元の強度配列から、目に見える 3 次元の表面の物理的特性を回復する一連の処理過程で構成されている。これらの処理過程を組み合わせた出力は、Marr の 2-1/2D スケッチ(1) や Barrow と Tennenbaum の固有画像(5) にほぼ対応している。最近では、これらの初期視覚過程は一般的なものであり、領域依存の知識を必要とせず、物理的な単語と画像化段階に関する一般的な制約のみを必要とすると仮定するのが通例となっている (ボックスを参照)。これらの処理過程は、概念的に独立したモジュールを表しており、第一近似的には分離して研究することができる。しかし、それぞれの処理過程から得られる情報を組み合わせる必要がある。さらに、異なるモジュールが早い段階で相互作用することもある。さらに、処理は純粋に「ボトムアップ」ではない。特定の知識が浸透して、視覚情報処理の最初の段階のいくつかに影響を与えることもある。

ボックス 初期視覚過程の例

- エッジ検出 Edge detection
- 時空間の補間と近似 Spatio-temporal interpolation and approximation
- 光学フローの計算 Computation of optical flow
- 明るさとアルベドの計算 Computation of lightness and albedo
- 輪郭からの形状 Shape from contours
- テクスチャからの形状 Shape from texture
- 陰影による形状 Shape from shading
- 両眼視対応 Binocular stereo matching
- 運動からの構造 Structure from motion
- 両眼視からの構造 Structure from stereo
- 表面再構成 Surface reconstruction
- 表面色の計算 Computation of surface colour

初期視覚モジュールの計算理論は、典型的には、表現と処理過程の二重の問題を扱う。つまり、入力と出力の形式を指定し (表現)、一方を他方に変換するアルゴリズムを提供しなければならない (処理過程)。ここでは、処理過程とアルゴリズムの問題に焦点を当て、正則化理論という統一された理論的枠組みを説明する。ここでは、処理過程とアルゴリズムの問題に焦点を当て、正則化理論という統一的な理論的枠組みを説明する。同様に重要な問題である、各特定の処理過程の入力を表す原生トークンについては考慮しない。

初期視覚の定義は、「逆光学」であるということである。古典的な光学やコンピュータグラフィックスでは、3 次元物体の画像を決定することが基本的な問題であるのに対し、視覚では画像から表面を復元するという逆の問題に直面する。3 次元の世界を 2 次元の画像に投影する画像処理の過程で非常に多くの情報が失われるため、視覚では、曖昧さのない出力を得るために、自然の制約、すなわち物理的世界に関する仮定に頼らざるを得ないことが多い。このような制約条件を特定して利用することは、特定の視覚問題を分析する際の繰り返しのテーマとなる。

初期視覚における 2 つの重要な問題は、動きの計算と、画像強度の急激な変化の検出 (物理的なエッジを検出するため) である。これらの問題は、初期視覚の問題の難しさをよく表している。画像内の 2 次元視野における物体速度の計算は、物体の動きと 3 次元構造を回復するためのいくつかのスキームの重要な段階である。画像内の滑らかな輪郭に沿った各点の速度ベクトル  $V$  を決定する問題を考えてみよう。Marr and Ullman (6) に倣い、輪郭は強度が大きく変化する場所に対応していると仮定する。図1 は、局所的な速度ベクトルが、曲線の法線方向と接線方向の成分に分解される様子を示している。局所的な動きの測定では、速度の法線成分のみが得られる。接線方向の成分は、(角などの輪郭の不連続な部分を参照しない限り) 純粋な局所計測では「見えない」ままである。したがって、完全な速度場を推定する問題は、一般的に、画像から直接得られる測定値では決定できない。光学フローの測定は本質的に曖昧である。情報や仮定を追加することによってのみ、一意にすることができる。

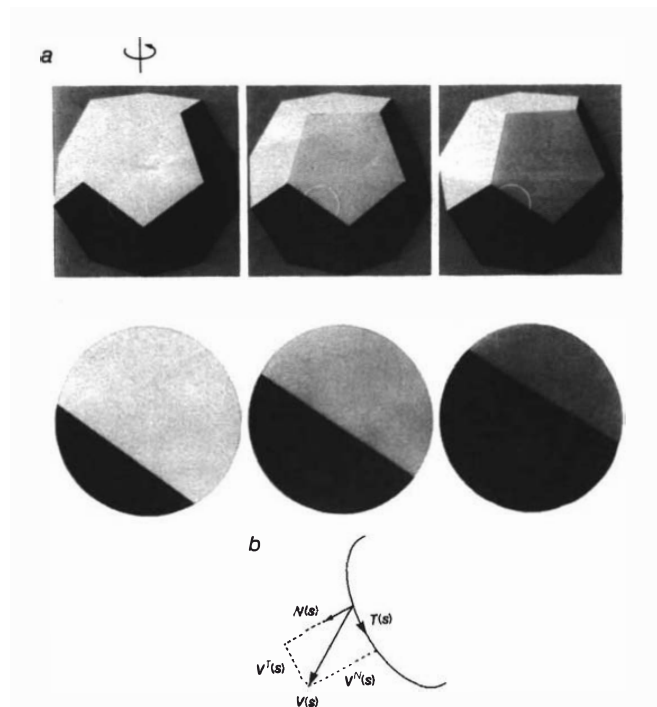


Fig. 1 速度場の曖昧性

**a** 局所的な計測では、剛体の3次元回転に起因する画像平面内の完全な速度場を計測することはできない(3つのフレームを示す)。開口部(白丸)内で動作する処理過程は、輪郭に垂直な動きの成分のみを計算することができる。**b** 弧の長さ  $s$  で表される輪郭に沿った速度ベクトルを、曲線に垂直な成分 ( $V^N(s)$ ) と接線方向の成分 ( $V^T(s)$ ) に分解している。

The computer drawing was kindly provided by Karl Sims.

エッジ検出の問題は、その難しさがやや異なる。エッジ検出とは、画像の強度変化から3次元表面の物理的境界を特定する処理過程を意味する。エッジ検出で意図しているのは、この目的のための第一歩、つまり画像の強度の急激な変化を検出して、それを局所化することである。これは、画像データの数値微分の問題であり、撮像時やサンプリング時に避けられないノイズに悩まされる。微分することでノイズが増幅されてしまうため、本質的に不安定な処理となってしまう。図3は、エッジプロファイルとその2次微分の例で、ノイズが大きく増幅されている。初期視覚の問題の多くは、同様の問題を抱えている。光学フローの計算のように制約が少なかったり、エッジ検出のようにノイズに対してロバストでなかったりする。

## 2. 不良設定問題

ほとんどの初期視覚問題に共通する特徴(ある意味、その深い構造)は、形式化することができる。すなわち、ほとんどの初期視覚問題は、Hadamardによって定義された正確な意味での不良設定問題である(7,8)。この主張は制約の重要性を表しており、逆光学としての視覚の定義を反映している。

Hadamardは、偏微分方程式の分野で、不良設定問題の定義を初めて導入した(9)。これまで、非線形問題は、ほとんど数学的な好奇心でしか考えられていなかった。だが、現在では、非線形問題の多くは、典型的な逆問題であり、実用的な関心事であることが明らかになっている(例えば、コンピュータモグラフィーなど)。問題は、その解が存在し、一意であり、初期データに連続的に依存する場合には、正解となる。逆問題とは、これらの条件のうち1つ以上を満たさない問題のことである。なお、3番目の条件は、実際に解がノイズに対して頑健であることを意味するものではない。そのためには、問題がよく分解されているだけでなく、数値的な安定性を確保するための条件が整っている必要がある(10)。

両眼視対応、動きからの構造、光学フローの計算、エッジ検出、陰影からの形状、明るさの計算、表面の再構成など、初期視覚におけるいくつかの問題がHadamard(8)の意味での不良設定であることを形式的に示すことは容易い。光学フローの計算は、輪郭に沿った法線成分から完全な速度場を復元するという「逆」問題が、一意性条件を満たさないため、不良設定である。数値微分を目的としたエッジ検出は、解がデータに連続的に依存しないため、不良設定となる。

不良設定問題を「解決」する、つまり「良好な設定」性を回復するための主なアイデアは、適切な事前知識を導入することで、許容される解のクラスを制限することである。事前知識とは、例えば、可能な解に制約を与える変分原理や、解空間の統計的な性質のような形で利用することができる。ここでは、一般的な正則化という用語を、不良設定問題を正則化するために用いられるあらゆる方法に使用する。変分正則化とは、変分原理の観点から不良問題を再定式化する正則化手法を示す。ここでは、主にTikhonov(11,12)に由来する、標準的な正則化法と呼ぶべき変分正則化法を紹介する(文献13,14参照)。また、標準的な理論を将来的に拡張するために、初期視覚の観点からも説明する。

データ  $y$  から  $z$  を求めるという不良設定問題を正則化すると

$$\mathbf{A}z = y, \quad (1)$$

は、ノルム  $\|\cdot\|$  と安定化汎関数  $\|Pz\|$  の選択が必要である。標準正則化理論では、 $\mathbf{A}$  は線形演算子であり、ノルムは2次であり、 $\mathbf{P}$  は線形である。適用可能な2つの方法がある(8,13)。(1)  $\|y - \mathbf{A}z\| < \epsilon$  を満たす  $z$  の中から、 $\epsilon$  を最小化する  $z$  を見つける。 $\epsilon$  は推定測定誤差に依存し、データがノイズ無しであればゼロとなる。

$$\|Pz\|^2 \quad (2)$$

(2) は次式を最小化する  $z$  を探索する:

$$\|\mathbf{A}z - y\| + \lambda \|Pz\|^2 \quad (3)$$

ここで  $\lambda$  はいわゆる正則化パラメータである。



第1の方法は、データを十分に近似し、かつ最も「規則的」な、すなわち「基準」 $\|\mathbf{Pz}\|^2$ を最小化する関数 $z$ を計算することである。2つ目の方法は、 $\lambda$ は解の正則化の度合いとデータへの近さの間の妥協点を制御する方法である。標準正則化理論は、最適な $\lambda(12,15)$ を決定する技術を提供する。このように、標準正則化法は、式(3)のコスト関数のような変分原理によって問題に制約を与える。最小化されるコストは、何が良い解を表すかという物理的な制約を反映し、データに近く、かつ、量 $\|\mathbf{Pz}\|^2$ を小さくする必要がある。 $\mathbf{P}$ は問題の物理的制約を具体化したものである。二次変分原理では、穏やかな条件の下では、解空間は凸であり、唯一の解が存在することが示される。標準正則化法は、問題の非負の性質を注意深く分析した後に適用しなければならないことを指摘しなければならない。ノルム $\|\cdot\|$ の選択、安定化関数 $\|\mathbf{Pz}\|$ の選択、関係する関数空間の選択は、数学的特性と物理的妥当性の両方によって決定される。これらは、正しい正則化のための正確な条件が、特定のケースで成立するかどうかを決定する。

変分原理は、物理学、経済学、工学の分野で広く使われている。例えば物理学では、基本法則のほとんどが変分原理によってコンパクトに定式化されており、エネルギーやラグランジアンなどの適切な関数を最小化することが求められる。

### 3. 例

式(3)の形式の変分原理は、過去に初期視覚で使用されていた(16-25)。その他の問題については、標準正則化手法を用いてアプローチされている(表1参照)。これまでに初期視覚で使用されてきた安定化関数のほとんどはTikhonov型であり、目的の解 $z$ の最初の $p$ 個の導関数の線形結合である(文献12)。これらの安定化関数から得られる解は、補間スプラインまたは近似スプラインに対応する。ここで、動きとエッジの検出の例に戻り、標準正則化技術がどのように適用できるかを示す。

表 1. 初期視覚における正則化

- エッジ検出  $\int [(Sf - i)^2 + \lambda(f_{xx})^2] dx$
- 光学フロー (領域ベース)  $\int [(i_x\mu + i_y\nu + i_t)^2 + \lambda(\mu_x^2 + \mu_y^2 + \nu_x^2 + \nu_y^2)^2] dx dy$
- 光学フロー (輪郭ベース)  $\int [(\mathbf{V} \cdot \mathbf{N} - \mathbf{V}^N)^2 + \lambda((\partial/\partial s)\mathbf{V})^2] ds$
- 表面再構成  $\int [(S \cdot f - i)^2 + \lambda(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)^2] dx dy$
- 時空間近似  $\int [(S \cdot f - i)^2 + \lambda(\nabla f \cdot \mathbf{V} + ft)^2] dx dy dt$
- 色  $\|\mathbf{I}' - \mathbf{Az}\|^2 + \lambda\|\mathbf{Pz}\|^2$
- 陰影からの形状  $\int [(E - R(f, g))^2 - \lambda(f_x^2 + f_y^2 + g_x^2 + g_y^2)] dx dy$
- 両眼視  $\int \{[\nabla^2 G * (L(x, y) - R(x + d(x, y), y))]^2 + \lambda(\nabla d)^2\} dx dy$

初期の視覚問題の中には変分原理の観点から解決される。最初の5つは、標準的な二次正則化原理である。エッジ検出(26,27)では、画像強度のデータ( $i = i(x)$ ) (演算子 $S$ は、復元されるべき連続分布 $f$ に対するサンプリング演算子である。同様の関数を用いて、時間変動する画像を近似することができる。データ $i(x, y, t)$ から復元される時空間的な強度は $f(x, y, t)$ であり、安定化項は画像平面内での一定の速度 $V$ という制約を課している(61)。領域に基づくオプティカルフロー(18)では、 $i$ は画像強度、 $u$ と $v$ とは速度場の2つの成分である。表面再構成(21,22)では、表面 $f(x, y)$ は、疎な深さデータ $d(x, y)$ から計算される。色(32)の場合、明るさは3つの適切な色座標 $I_v(v = 1, 2, 3)$ のそれぞれで測定される。解ベクトル $z$ は、照明成分とアルベド成分を別々に含み、 $A$ によって理想データに写像される。適切な安定化項の最小化は、空間的に滑らかな照明と、一定または急激に変化するアルベドの制約を強いる。陰影シェーディング(19)と立体視からの形状(T.P.and A. Yuille, 未刊行)については、2つの非二次正則化関数を示す。 $R$ は反射率地図、 $f$ と $g$ とは表面勾配の成分に関係し、 $E$ は輝度分布である(19)。視差の視野 $d$ の正則化には、左画像(L)と右画像(R)のガウス分布の2階微分ラプラシアンと、視差勾配に対応するTikhonov安定化関数との畳み込みが必要である。

直感的には、拡張された輪郭上の速度の法線成分の測定値のセットは、輪郭の大域的な動きにかなりの制約を与えるはずである。しかし、異なる場所での局所的な測定値を組み合わせるためには、実世界の性質に関するいくつかの追加的な仮定が必要である。例えば、 $V$ を一意に決定するためには、画像平面上での剛体運動の仮定で十分である(23,4)。この場合、異なる場所での法線成分の局所的な測定値を直接使用して、どこでも同じであるオプティカルフローを求めることができる。しかし、この仮定は、3次元空間内の剛体の動きの場合をカバーしていないので、過度に制限されている(図1)。HildrethはHomとSchunck(18)に続いて、速度場に対するより一般的な滑らかさの制約を提案した(23,24)。実世界は滑らかな表面を持つ固体で構成されており、その投影された速度場は通常滑らかである、という物理的な考察が背景にある。安定化剤の具体的な形(Tikhonov安定化項)は、数学的な考察、特に解の一意性によって決定された。2つの正則化法は、Hildreth(23)によって提案され、実装された2つのアルゴリズムに対応する。最初のものは、法線方向の速度成分 $V_N(s)$ の測定値が正確であることを前提とし、次のように最小化する:

$$\|\mathbf{Pz}\|^2 = \int \left(\frac{\partial \mathbf{V}}{\partial s}\right)^2 ds$$

(4)

は、速度の法線成分の測定値を対象とする( $s$ は弧長)。積分は輪郭に沿って評価される。正確ではないデータの場合、2番目の方法は、次のように最小化することで解を得る。

$$\|\mathbf{V} \cdot \mathbf{N} - \mathbf{V}^N\|^2 + \lambda \int \left(\frac{\partial \mathbf{V}}{\partial s}\right)^2 ds$$

(5)

ここで、 $N$ は輪郭に対する法線方向の単位ベクトル、 $\lambda^i s - 1$ はデータの信頼性を表す。図2aは、第1のアルゴリズムでオプティカルフローの計算に成功した例である。

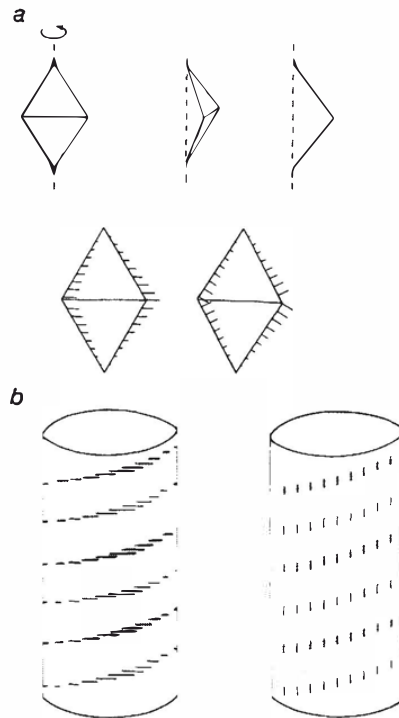


図2 輪郭に沿って最も滑らかな速度場を計算 a. Wallach (62) が最初に使用した3次元刺激。物体の2次元の動きを投影して3次元の構造を導き出す人間の視覚システム的能力(運動奥行き効果)を示す。上の部分は、ある図形を垂直軸の周りに回転させたときの3つの見えを示している。右下には、法線方向の速度成分  $V_f$  の初期測定値を示している。左下には式(4)を用いて計算された速度場が示されている。最終的な解は、物理的に正しい速度分布に対応している。最近の電気生理学的証拠によると、サルの中側頭領域は、同様の運動統合が起こる部位であると考えられている(63)。

b. 垂直軸を中心に回転する架空の3次元円柱上の円形螺旋(床屋の円柱)。曲線の画像平面への投影と、その結果としての2次元速度ベクトルが左に描かれている。真の速度場  $V$  は厳密には水平(左)であるが、最も滑らかな速度場(右)は垂直である。この例は、アルゴリズムと人間の視覚系の両方が同じ目の錯覚に陥る場合を示している。文献23より引用。

近年、正則化技術がエッジ検出に応用されている(26,27)。数値微分の問題は、画像に滑らかさの制約を反映させた Tikhonov 安定化項を用いた第2の方法で正則化することができる(表1)。物理的に正当化されるのは、高い空間周波数をカットする光学系の帯域制限特性のため、画像が有界導関数であることである。この正則化された解は、穏やかな条件の下では、以前に提案されたガウシアン(26, 図3)に似たフィルタの導関数で強度データを畳み込むことと同等である(28,29)。

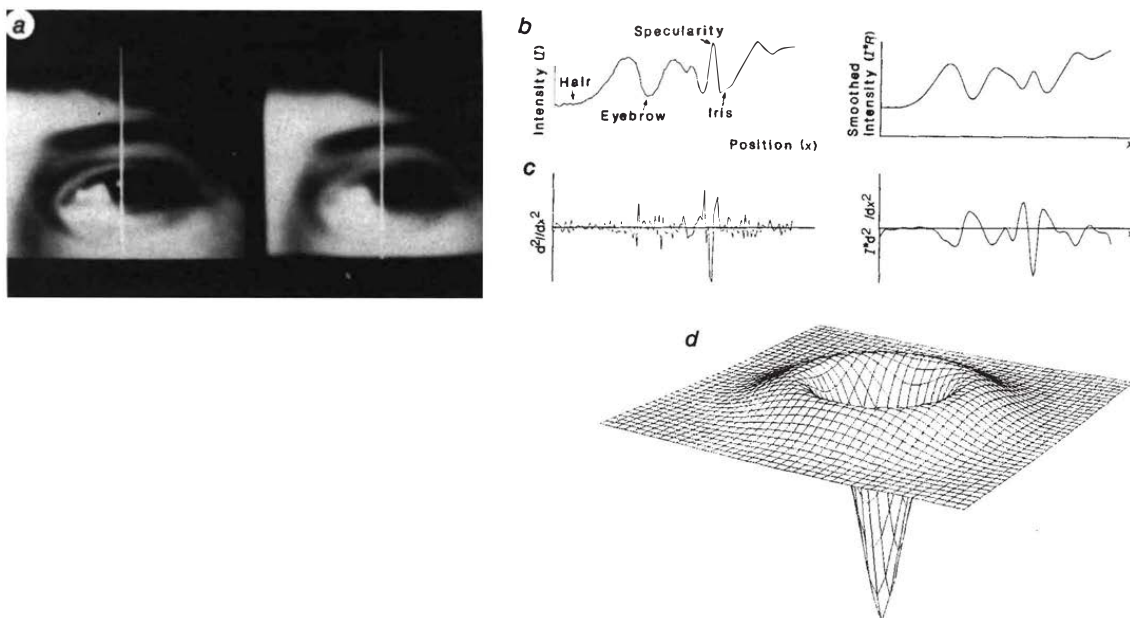


図3. エッジ検出のための正則化解。a. デジタル画像(256x256画素)に何もフィルタリングしない状態(左)と、dに示した標準的な正則化理論(26)による2次元演算子でフィルタリングした状態(右)。b. 一次元正則化フィルタ(26)によるフィルタリングを行っていない場合(左)と行った場合(右)の、aで示した走査線に沿った強度プロファイル。c. 一次元正則化フィルタリングを行っていない場合(左)と行った場合(右)のbで示されたプロファイルの二階微分。d. エッジ検出の非正規配置問題を正規化して得られた2次元フィルタ(26)。これは、ガウス分布によく似た五角形のスプラインである。パラメータAは、フィルタのスケールを制御する。その値は、画像のS/N比に依存する。脊椎動物の網膜にあるほとんどの神経節細胞の空間受容野は、中央に興奮性領域があり、周囲に抑制性領域があるという非常に似た構造を持っている(64,65)。図はHarry Vorhees氏よりご提供いただいた。

その他の初期視覚問題は、標準正則化技術によって解決することができる。例えば、表面の再構成は、表面の滑らかさを仮定することで、疎な深度値の集合から実行することができる(20-22)。オプティカルフローは、Tikhonov 安定化項の形で、滑らかな変化の制約を用いることで、輪郭に沿ってではなく、画像の各点で計算す

ることができる(17)。正確には二次式ではないが、式(3)の形をした変分原理は、初期視覚の他の問題にも利用できる。実際、Tikhonovの主要な結果は、ある条件を満たせば、演算子 $A$ と $P$ が非線形である場合にも拡張できる(30)。物体の明るさの変化は、その形状を知る手がかりとなる。表面の向きは、表1に示す変分原理を用いて、強度画像から計算することができる。この変分原理は、平滑性制約と照度制約に違反する向きに罰則を与える(18)。両眼視対応とは、一対の両眼画像から、一方の画像のどの特徴が他方の画像のどの特徴に対応しているかを見つけて、正しい両眼視差(したがって奥行き)を推定する問題である。この問題は非可決であり、隠蔽がない場合の制限的な条件の下では、2つの画像から抽出された特徴地図間の不一致を測定する項を含む変分原理と、大きな視差勾配に罰則を課して実質的に視差勾配制限を課す安定化項によって正則化することができる(表1)。このアルゴリズムは、視差勾配が小さければ、西原型の面積ベースの相関アルゴリズム(31)に還元できる。色画像において、空間的に変化する照明から物質の反射を分離する問題を解決するために、標準正則化原理が提案されている(32)。このアルゴリズムは、視覚心理物理学で「色の恒常性」として知られている問題を解決するものである(33)。

## 4. 物理的妥当性と錯覚

正則化解析では、解の一意性よりも、解の物理的妥当性が最も重要な関心事である。問題とその重要な制約条件の物理的分析が主な役割を果たす(8)。正則化された解が物理的な解に対応していない特定のケースでは、非負値の問題を解くために必要な先験的な仮定が破られる可能性がある。アルゴリズムが目錯覚を起こしてしまうのである。その良い例が、運動視の計算である。式(5)の滑らかさの仮定は、いくつかの一般的な条件下では正しい結果を与える(例えば、物体の画像が連結した直線で構成されている場合など、34)。動きや輪郭のクラスによっては、平滑性の原理では正しい速度場が得られないことがある。しかし、これらのいくつかのケースでは、人間の視覚システムも同様に正しくない速度場を導き出しているようで、脳が世界について作っている先験的な仮定を明らかにしている可能性がある。顕著な例としては、床屋の柱錯視(23, 図2b)がある。

## 5. アナログネットワーク

生物の視覚の謎の一つは、その処理速度である。この問題を解決する方法として、並列処理がしばしば提唱されてきた。しかし、ニューロンが単純なデジタルスイッチとは全く異なる複雑なデバイスであるという証拠が増えていることを考えると、デジタル処理によって提供される計算モデルは満足いくものではない。したがって、初期視覚に対する正則化アプローチが、別種の並列計算につながるかどうかを考えるのは興味深いことである。我々は最近、線形のアナログネットワーク(電気的または化学的)が、標準正則化理論(7)で規定されている変分原理を解く自然な方法であることを提案した(文献22, 35 参照)。

変分原理と電気・化学ネットワークとの間にこのような対応関係が成立する基本的な理由は、ハミルトンの最小作用原理にある。アナログネットワークで計算可能な変分原理のクラスは、キルヒホッフの電流法と電圧法によって与えられ、ネットワークの各構成要素が満たす保存性と連続性の制約を表している(適切な変数は、電気ネットワークでは通常、電圧と電流、化学システムでは親和性と回転速度(36, 37)。一般に、ユニークなネットワークは存在しないが、同じ変分原理を持つネットワークは多数存在する可能性がある。例えば、連想記憶の文脈でHopfieldが提案したタイプの段階的ネットワーク(38)は、標準的な正則化原理を解くことができる(39)。

キルヒホッフの法則から、ユニークな解を持つ二次変分問題(通常はそうなる、8)には、同じ解を持つ抵抗と電圧または電流源からなる対応する電気ネットワークが存在することが証明できる(7)。つまり、ネットワークの定常電流(または電圧)分布は、標準正則化問題の解、たとえば接線速度分布 $v^T(s)$ に対応している(図4)。さらに、システムにキャパシタンスを追加してダイナミクスを導入しても、システムは安定している。データは、電流を注入したり、電池を導入したりして、つまり定電流源や定電圧源によって供給される(7)。

このアナログ並列計算モデルは、ニューロン、膜、シナプスの生物物理学を理解する上で、特に興味深い。多くのニューロンでは、電位が主要な役割を果たしていることを示す証拠が増えている(40)。樹状突起のシナプス(41, 42)、ギャップ結合(43)、異なる時間と距離で作用する神経伝達物質(44)、神経ペプチドで調節可能な電圧依存性チャネル(45)、シナプスのコンダクタンス変化の相互作用(46)などの多様なメカニズムが、神経細胞にさまざまな異なる回路要素を提供している。神経膜のパッチは、抵抗、容量、現象論的インダクタンスに相当する(47)。樹状突起上のシナプスは電圧源に似ており、太い樹状突起や体幹上のシナプスは電流源として機能する(48, 49)。このように、単一のニューロンや小さなニューロンネットワークは、正則化原理のアナログ解を実装することができる。図4のアナログ回路の仮想的なニューロン実装が考案されているが、これには1本または2本の樹状突起しか関与していない(7)。

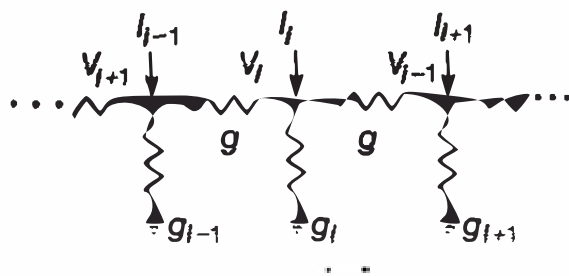


図4 アナログネットワーク。最も滑らかな速度場を計算する抵抗性ネットワーク(23)。このネットワークは、法線方向の速度成分 $V^N$ の測定値が正確であると仮定した場合に対応する。関連する変分方程式(4)を輪郭に沿って離散化すると、オイラー・ラグランジュ方程式 $\mathcal{L}(2+\kappa_i^2)V^T - v_{i+1} - v_{i-1} = d$ 、ここで、 $\kappa_i$ は位置における輪郭の曲率、 $d$ はデータ $V^T$ と輪郭の関数であり、 $v_{i+1}$ は輪郭に沿った位置における速度 $V$ の未知の接線成分である。電気回路の $i$ 番目のノードを記述する方程式は、 $(2g+g_i)V_i - V_{i+1} - V_{i-1} = d$ であり、ここで $\kappa_i$ は未知の $V^T$ に対応する電圧であり、 $d$ は測定値 $V_i^T$ に依存するノード $i$ での注入電流である。 $V_i^T$ の測定値が正確でない場合には、もう少し複雑な回路を設計することができる(7, 式(5))。正則化された解の一意性は、容量が導入された場合でも、対応するネットワークの安定性を常に保証する。同等のアナログネットワークは、拡散反応系で実装することができる。ここでは、近隣の場所間の相互作用を、一次の速度論による拡散または化学反応を用いて模倣する。仮想的なニューロンの実装を想定することもできる。コンダクタンス $g$ は、樹状突起の小さなセグメントに対応し、可変コンダクタンス $g$ は、樹状突起の静止電位に近いか等しい反転電位を持つシナプス入力(すなわち、サイレントまたはシャント抑制)に対応し、電流源は、樹状突起に電流 $I$ を注入する従来の化学シナプスに対応している。出力は化学シナプスによって位置 $i$ でサンプリングされる。文献7より引用

## 6. 標準正則化理論を超えて

初期視覚の新しい理論的枠組みは、標準正則化理論のTikhonov形式に内在する魅力と限界を明確に示している。主な問題は、復元しなければならない未知の関数に必要な滑らかさの度合いです。例えば、表面補間において、いわゆる薄板スプラインに対応する滑らかさの度合いは、深さ方向の不連続性を滑らかにしすぎ、しばしば非現実的な結果をもたらす(20 ただし、不連続性は検出され、その後、第2の正則化ステップで使用されることがある、66)。



標準正則化理論は、線形問題を扱い、二次安定化項に基づいている。そのため、二次関数の最小化と線形オイラー=ラグランジュ方程式が導き出される。正しい物理的制約を与えるためには、非二次関数が必要となる場合がある(表1は陰影からの形状復元の非二次の場合を示している)。この場合でも、標準正則化理論の手法を用いることができるが(30)、解空間はもはや凸形ではなく、最小化の過程で多くの局所的な最小値が見つかる可能性がある。

奥行きから表面を再構成する際に不連続性を保存する問題に対して、非二次の安定化項が提案されている(50)。この安定化項は、基本的にはGemanとGeman(51)に起因するもので(同様の原理がBlake(52)によって提案されているが、厳密な正当化はされていない)、Terzopoulos(67)の変分連続性制御も参照)、不連続性の幾何学的性質(ライン仮定)、特に連続的でしばしば直線的な輪郭であることについての事前知識を埋め込む。標準正則化原理では、探索空間には1つの局所的な最小値しかなく、適切なアルゴリズムは常にそこに収束する。非二次関数の場合、探索空間は多くの局所最小値を持つ山脈に似ているかもしれない。最近、この種の最小化問題を解くための確率的アルゴリズムが提案されているが、これは、単純な山登りアルゴリズムが引つかかるような局所最小値から逃れるためである(53-55)。その基本的な考え方は、探索アルゴリズムに強制的なノイズ項を追加することに多少似ている。非二次変分原理を非線形アナログネットワークで表現することができれば(文献39のように)、適切なガウシアンノイズ源がアナログネットワークを駆動することができる。システムのダイナミクスは、拡散過程を表す非線形確率微分方程式で記述される。

視覚の正則化理論にとって現在の課題は、標準正則化法を超えて拡張することである。二次関数の観点から実行できる計算の世界はかなり限定されている。これを理解するには、二次損失関数の最小化が線形正則化演算子、つまり入力データの解空間への線形写像につながることを理解すれば十分である。データが規則的なグリッド上にあり、適切な条件に従っている特殊な場合は、線形演算子は畳み込み、つまりデータに対する単純なフィルタリング操作になる。物理学における線形モデルと同様に、標準正則化理論は、多くの場合、非常に有用な近似だが、視覚の複雑さを完全に扱うことはできない。

## 7. 正則化への確率的経路

ベイズ推定とマルコフ確率場モデルに基づく正則化の別の厳密なアプローチがある。このアプローチでは、事前知識は適切な確率分布で表現されるが、標準正則化では事前知識は解空間の制限につながる。例として、表面の再構成の場合を考えてみよう。知識は、表面のマルコフ確率場(MRF)モデルの観点から定式化することができる。MRFでは、ある離散的な場所での値は、与えられた近隣の値にのみ依存する。この方法では、最適な表面は、最大事後推定値やMRFの事後平均値など、ある種の尤度基準を最大化する。MRFの最大事後推定は、式(3)の一般的な形式の変分原理と同等であることが指摘されている(50)。第1項はデータと解の間の不一致を測定し、第2項は解の任意のポテンシャル関数(離散格子上で定義される)となる。一般的には二次関数ではない変分原理は、ノイズが加法性でガウス性であり、場の一次差分がゼロ平均で独立したガウス性の確率変数である場合、標準的な正則化タイプの二次関数に減少する。この場合、最大事後推定値(MAP)はすべての推定値と一致し、特に事後平均と一致する。しかし、Marroquin(56)は最近、一般的にはそうではないことを示した。ほとんどの場合、MAP推定値は自然な誤差測定に関して最適ではなく、事後平均のようなより良い推定値を見つけることができる。このような場合、この問題はエネルギー関数の大域最小値を求めることと同等ではない。シミュレーテッドアニーリングは必要なく、代わりにメトロポリス・アルゴリズム(55)を使用することができる。

Hildrethの運動視の計算(23)の場合、滑らかさの仮定は、輪郭に沿った近隣点間の速度変化がゼロ平均の独立したガウス分布の確率変数であるという仮説に対応する。確率論的アプローチと標準正則化法の間この接続は、制約条件の性質と安定化項の選択について興味深い視点を与えてくれる。視覚の逆問題を解くのに使われる変分原理は、もっともらしい解を生成するマルコフ構造に対応している。

今後の研究課題として、正則化演算子の学習問題が挙げられる。標準正則化の場合、データを解に写像する対応する線形演算子は、生物学的記憶(58)に関連して提案されたタイプの連想学習スキーム(57)によって学習することができる。

## 8. 記号的描写を目指して

ここまでは、初期視覚段階である、見る人の周りにある物理的な3次元の表面を画像のように表現することに限定して説明してきた。固有画像(5)や2-1/2Dスケッチ(1)と呼ばれるこれらの表現を超えるステップは、大きなものである。固有画像は、まだ画像的な数値表現であり、物体としては記述されていない。固有画像は、操作やナビゲーションなど、視覚系の高レベルな課題の一部にはすでに十分である。しかし、より記号的な表現の生成と使用を必要とする認識や記述の課題には、直接使用することはできない。非正則化問題の正則化という観点から、記号的表現の計算がどのように適合するのか、最初は理解できないように思われる。

すべての正則化手法の基本的な考え方は、可能な解の空間を制限することである。この空間が有限の次元を持つように制限されていれば、逆問題が正則化される可能性が高くなる。このように、離散的な記号の有限集合に基づく表現は、不良設定問題を正則化する。この観点から、知覚の問題(物理世界の一般的な制約を用いて、制約の少ない問題を正則化する)は、古典的な人工知能の解法と推論の問題と実質的に同等になる。すなわち、解法の探索を制限することで、難解な問題(チェスなど)を解く方法を見つけることができる。

## 9. おわりに

我々は、現在開発中の並列デジタルコンピュータアーキテクチャに自然に対応する視覚アルゴリズムの分類を提案する。標準正則化が十分であれば、2つのクラスの並列アルゴリズムになる。最急降下法のような凸関数の最小値を求めるアルゴリズムや、ビジョン用に開発されたより効率的なマルチグリッドアルゴリズム(59)は常に使用することができる。データが規則的なグリッド上に与えられ、式(1)のAが空間不変である場合、これらのアルゴリズムは畳み込みアルゴリズムで置き換えることができる。後者の場合、正則化解は、あらかじめ計算されたフィルタを通してデータを畳み込むことで得られる。

これらのアルゴリズムはすべて、局所的接続しか持たない多数のプロセッサからなる並列アーキテクチャで実装できる可能性がある。正則化の観点からアプローチできない問題や、記号的な表現とその演算を必要とする問題では、現在開発中のConnection Machine(60)のような大域的な通信機能を備えた並列アーキテクチャが必要になるかもしれない。

初期視覚の問題の多くは、不良設定問題という概念と、それに関連する新旧の正則化理論によって、満足のいく理論的枠組みが提供されてるようである。また、この新しい視点は、初期視覚問題の計算上の性質(非負の問題)と、それを解くアルゴリズムの構造、そして効率的な視覚情報処理に使用できる並列ハードウェアとの間に関連性をもたらす。また、初期視覚でこれまで使われてきた変分原理の本質的な限界を示し、同時に正則化分析を標準化理論を超えて拡張する方法を示している。