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 02193, USA

* Istituto di Fisica, Universita 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 2-1/2D sketch[1], and to Barrow and Tennenbaum's intrinsic images[5]. 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 word 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[6], 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

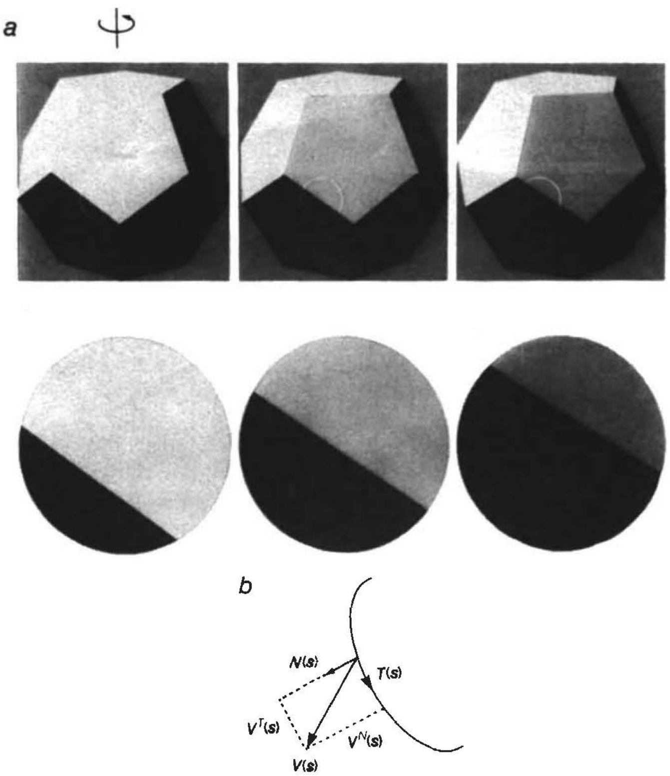

Fig. 1 Ambiguity of the velocity field. *a*, Local measurements cannot measure the full velocity field in the image plane, originated here by three-dimensional rotation of a solid object (three frames are shown). Any process operating within the aperture (shown as a white circle) can compute only the 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[7,8]. 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[9]. 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[10].

It is easy to show formally that several problems in early vision are ill-posed in the sense of Hadamard[8]: 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[11,12] (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

$$Az = y \tag{1}$$

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[8,13]: (1) among z that satisfy $\|Az-y\| \leq \varepsilon$ find z that minimizes (ε depends on the estimated measurement errors and is zero if the data are noiseless)

$$\|Pz\|^2 \tag{2}$$

(2) find z that minimizes

$$\|Az-y\|^2 + \lambda\|Pz\|^2 \tag{3}$$

where λ 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' $\|Pz\|^2$. In the second method, λ 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 λ[12,15]. 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[16-25]. 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	$\int[(S\cdot f-d)^2+\lambda(f_{xx}^2+2f_{xy}^2+f_{yy}^2)^2]\,dx\,dy$
Spatiotemporal approximation	$\int[(S\cdot f-i)^2+\lambda(\nabla f\cdot V+ft)^2]\,dx\,dy\,dt$
Colour	$\|I^\nu-Az\|^2+\lambda\|Pz\|^2$
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[26,27] 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[18], i is the image intensity, u and v are the two components of the velocity field. In surface reconstruction[21,22] the surface $f(x,y)$ is computed from sparse depth data $d(x,y)$. In the case of colour[32] the brightness is measured on each of three appropriate colour coordinates $I^\nu(\nu=1,2,3)$. 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[19] 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[19]. 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[23,24]. 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[23,24], following Horn and Schunck[18], 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[23]. The first one, which assumes that the measurements of the normal velocity components $V^N(s)$ are exact, minimizes

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

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

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

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

Recently, regularization techniques have been applied to edge detection[26,27]. 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[26] (see Fig. 3), proposed earlier[28,29].

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[20-22]. 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[17]. 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[30]. 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[18]. 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[31] 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[32]. The algorithm addresses the problem known in visual psychophysics as colour constancy[33].

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[8]. 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[34]). For some classes of motion and contours, the smoothness principle will not yield