BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 33, Number 2, April 1996

Variational methods in image segmentation, by Jean-Michel Morel and Sergio Solimini, Progress in Nonlinear Differential Equations and Their Applications, vol. 14, Birkhäuser, Boston, 1995, xvi + 245 pp., $59.00, ISBN 0-8176-3720-6

This is a remarkable multifaceted book in a field which is, from my perspective, a remarkable multifaceted area. Let me state up front so there is no confusion: it is the same field in which I have been working for a dozen years, and the book itself is concerned with problems and a conjecture worked on by Jayant Shah and me. So there is no question of whether this is a truly objective review! With that out of the way, let me sketch the background for this field and this book.

The scientific study of vision began with the work of the two great German psychophysicists Ernst Mach [Mc] and Heinrich von Helmholtz [vH] in the nineteenth century. Bits and pieces of interesting math came into it at various times (the spherical trig of the eyeball and its motions—Listing's law; the saliency of non-C^1 points of a perceived image—Mach bands; the Gestalt grouping laws; etc.). But by and large the field remained the province of psychologists and neurobiologists studying vision in animals and especially in man. Quite independently, however, engineers trying to build robots with sight started to grapple with image processing and image analysis in the '60s. Several people brought these two groups together. The work of Bela Julesz on stereoscopic binocular vision at Bell Labs and that of David Marr, starting out as a neural modeller and putting together a unified view of the field [Mr] at the MIT AI lab, are especially notable. The basic idea is that there is *one computational problem of vision*, which is solved in many different animal groups (e.g. mammals, birds and octopi, with independently evolved structures) and which we hope to solve by computer. This field has come to be called computer vision, but it is meant to include all computational aspects of vision in both robots and animals. To put it succinctly, the computational problem is to use the information in the raw visual input, which is an array of intensity values measured by the retina or TV camera, and infer the three-dimensional structure of the world in front of the camera and as much as possible about the identity or category of the objects present. Thus the field includes large parts of the related field of image processing and the fields of image analysis and object recognition. It has very strong parallels with the fields of speech recognition, computational linguistics and the AI (artificial intelligence) study of natural language. Before the field grew, it was often considered as a subfield of AI, but is has now largely gone its own way, and its practitioners can be found in engineering, computer science and psychology departments as well as biology, biomedical and statistics departments, etc.

In the meantime, it has also grown as an area of applied mathematics. There are two quite beautiful applications of mathematics to vision which have been extensively developed in the last twenty years. One of these is the application of the geometry of three-dimensional space (and its projections to two-dimensional image planes) to the inference of three-dimensional structure in the scene producing an image and to matching the objects in the image with object models of various types. To give a few examples, one problem is to identify parts of each of a sequence of images which can be interpreted as successive views of a single rigidly translating and rotating 3-D object. This goes under the name of *structure from motion*: cf.

the recent survey [Fau]. If multiple geometric object models are available, one also seeks to match specific views of specific objects to each image. This can proceed by using various projective geometry ideas such as cross ratio [M-Z] or, if a precise enough 3-D reconstruction can be done, by using curvature properties to identify the object [K]. The second application of mathematics to vision uses the ideas of statistics and recasts the fundamental problem of vision as making statistical inferences about the world from raw image data rather than exact conclusions. This set of ideas was pioneered by the work of Grenander (see his recent synthesis [Gr] and my overview [Mu]), Cooper [Ceta1] and Stuart and Donald Geman [G-G], who saw how to bring in ideas from statistical physics. This work has been done primarily in a Bayesian statistical framework, some of which has also involved use of entropy, coding theory and minimum description-length ideas.

Morel and Solimini's book belongs to the second of these developments. Specifically it is concerned with the problem of segmenting images. To fix notation, let $I(x, y)$ be a function on a plane domain D given by the measured incident light on the focal plane of an eye or camera: we call such an I an *image*. In every image, various three-dimensional objects will be visible, and their visible surfaces will be projected on subsets $R_i \subset D$ of the domain of I, while the restriction $I|_{R_i}$ is the subimage representing the ith object. In many cases, since the functions $I|_{R_i}$ are images of one object, these functions are smooth and slowly varying. This happens, for instance, if the surfaces have slowly varying albedo and normal vector and no strong shadows or specular reflections are present. In other cases, the surface is textured by variable albedo (e.g. patterns on clothes) or by micro-geometry (e.g. a lawn or pond), but one can still expect that $I|_{R_i}$ has slowly varying power spectrum or nearly stationary statistics of some kind. In still other cases, such as the presence of normal vector discontinuities (think of polyhedral objects) or sharp shadows, it may be best to subdivide the subsets R_i further into parts of the surfaces of objects in order to achieve some local homogeneity. In any case, we are led to seek a decomposition $D = \bigcup_i R_i$ of the domain of I into disjoint parts on each of which I (or in the textured case, some vector of local statistics of I) is slowly varying. Computing the correct decomposition of the domain of an image has long been recognized as the first major computational step in the analysis of images, the central problem of what people refer to as *low-level* vision.

The first part of Morel and Solimini's book is an extensive survey of the many methods that have been introduced in computer vision for the solution of the segmentation problem. As an aside, they entitle this section *modelization* following a mistaken but universal French belief that *to modelize* is an English verb. This section is a real tour de force, especially as it is accompanied by an exhaustive bibliography. Nothing comparable to *Math Reviews* exists in hybrid fields like computer vision, and it is worthwhile pointing out that the authors cite papers from all over the map, e.g. *J. Optical Society of America, Comp. Vision, Graphics, Image Proc., IEEE Transactions, J. Amer. Stat. Assoc., Comm. ACM, SIAM J. Numerical Anal.*, as well as the mainline math journals. This part is a first class introduction to all the basic algorithms for edge detection, region growing and image processing by nonlinear PDEs, and I recommend it strongly to anyone seeking to get a sense of the field. The central theme of this part is that all these diverse approaches to segmentation make better sense and are better understood when posed as variational problems. This means you should define a cost functional or energy functional $E(I, \{R_i\})$ which evaluates how well a proposed segmentation $\{R_i\}$ explains the

structure of a given image I. What you want is that the correct segmentation of most reasonable images I will be either the minimum of E, I being fixed, or at least close to a local minimum on which E is close to the global minimum of E. They argue (p. 5) in four ways that this is correct. First, without such a guiding variational principle, the heuristic algorithms in the literature get extremely complex; second, that any decent algorithm should state when one segmentation is better than another, hence should lead to an E; third, that most practical methods that have been used can be recast variationally; and finally, that variational principles arise naturally from *multiscale analysis*.

Multiscale analysis is indeed a second central theme in this book. I think it will help explain to the reader much of what follows in this book if we briefly review these ideas. In tactile sensing, you must touch an object to sense it, hence its size on the tactile sensor array is always the same; in auditory sensing, nature provides an absolute temporal scale by the bodily rhythms such as your gait, your heartbeat or the vibration of your vocal cords. But in vision, *there is no given scale*: if you move close to a scene, all objects get bigger, and if you move farther away, all objects get smaller. Thus the statistics of images should be scale-invariant: $I(x, y)$ and $I(\sigma x, \sigma y)$ should be equally likely. In particular, if you are looking for a face, it may take up all of D or it may be a tiny part of D. Now a central part of the job of the applied mathematician is to think about what kind of mathematical abstraction is best suited to a particular physical reality. If you think about the full implications of scale invariance, it becomes clear that images are not functions at all: they are better considered as distributions. In fact, if you had Superman's super vision, you would see the microbes on every surface; and as the normals to their surfaces vary as much as those of macroscopic objects, the laws of shading imply that I would vary as much from one side of the microbe to the other as it varies across the whole of D. This infinitely fine everywhere-present detail is incompatible with I being a function.

Actually this problem comes up again and again in many guises, another one of which is called *clutter*. If you were critical in reading the previous paragraph, you might have wondered whether the segmentation model introduced there was reasonable for things like your desktop or your bookshelf. Books on a shelf, for instance, vary so much in color, size, orientation (if some are leaning over or horizontal) that it is hard to assert confidently that an image of a book shelf has stationary statistics! This is clutter, and it makes clear that segmentation can only be done if you also break scale invariance. I would argue that the presence of clutter in visual scenes shows that modeling (modelizing?) images by distributions is not a mere mathematical abstraction but the natural setting for a property of images which is already clear on the limited range of scales (roughly four orders of magnitude) that our eyes deal with.

The basic variational problem studied in Morel and Solimini's book does indeed break scale invariance, but it does so in a minimal way (as opposed to more drastic solutions like (2.3) on p. 14), and it pays a price for this: it is not clear whether it is well posed. This is the main issue studied in parts 2 and 3 of the book and is known as the Mumford-Shah conjecture. The variational problem in question is the one Shah and I introduced in 1985–89 [M-S1, M-S2] which was based on looking for another way to analyze the ideas of Cooper [Ceta1], the Gemans [G-G], and Blake

and Zisserman [B-Z]. The definition is:

$$E(I, J, \{R_i\}) = \int_D (I - J)^2 + \sum_i \int_{R_i} \|\nabla J\|^2 + \text{length}(\bigcup_i \partial R_i).$$

(We follow here the book's refreshing tendency to drop constants which can be normalized away.) Here a new variable J has made its entrance. J has been called the *cartoon* of the full image I and is an idealization of the full image I in which clutter and noise have been stripped away. J lives in the Sobolev-Hilbert space of functions whose restriction to each R_i is in $H^1(R_i)$. Thus J will be, in general, discontinuous across the boundaries of the R_i, which we write, following the book, as $K = \bigcup_i \partial R_i$. The functional E is quadratic in J, so for each K, it has a unique minimum J_K, and we write $E(I, K) = E(I, J_K, K)$ for simplicity. Intuitively, minimizing E for a fixed I is finding *edges* K which minimize a sum (in general a weighted sum) of the length of K and the error and size of the best approximations of $I|_{R_i}$ by H^1-functions $J|_{R_i}$.

When an image is discretely sampled and E is approximated by a finite sum, then $E(I, J, \{R_i\})$ is still quadratic in J; and since there are only a finite number of possible regions R_i, E certainly has a minimum. Experiments show that minimizing E generally produces reasonable segmentations. More precisely, since no algorithm is known even in the discrete case for finding the exact minimum of E, these experiments show that various algorithms which approximate the minimum produce reasonable segmentations. Sections 4.4 and 5.4 describe two such approximate algorithms, and examples of the output of the second are given in 5.5.

However, discrete sampling introduces another length scale and is a radical sort of low-pass filtering. Does the original E in the continuous domain have good minima or minima of any kind? The middle part of the book, entitled "Elements of Geometric Measure Theory", develops the tools needed to study this question. The part is an excellent exposition of the theory of rectifiable sets in $\mathbf{R}^n$ starting with the basic theory of Hausdorff measures $\mathcal{H}^\alpha$. It then introduces the basic tool of analyzing a set $K \subset \mathbf{R}^n$ by its densities

$$\mathcal{H}^\alpha(K \cap B(x, r))/r^\alpha,$$

where $B(x, r)$ is a ball of radius R centered at a point x. This leads to the decomposition of a set K with finite α-Hausdorff measure into its regular and irregular parts. The rest of the section is devoted to the analysis of the regular part, proving that regularity is equivalent to being rectifiable. The authors suggest at various places that these densities are very natural tools in the computer vision context. Similar ideas have recently been investigated by S. Zucker and his student Dubuc [D-Z] to find numerical measures that distinguish isolated edges in images from "dense" sets of edges that are found in textured areas of images. My only quibble with Morel and Solimini's exposition is that the authors do not display and number their definitions, only their theorems, lemmas and corollaries. This sometimes makes it quite hard to locate the key definitions. This part of the book parallels the book of Falconer [Fal] which deals with original Besicovitch theory of 1-measurable subsets of the plane. But by incorporating the approach of Mattila and Marstrand and the reflection lemmas, the authors manage to give a concise and elementary treatment in all dimensions.

The last section of the book is entitled "Existence and Structural Properties of the Minimal Segmentations for the Mumford-Shah Model". This section is a

very coherent exposition of the results of the authors and their collaborators on the well posedness of E. This is approached by extending E to a suitable large set of K's within which a weak minima can be proven to exist. This was first done by De Giorgi and his many collaborators by introducing the concept of SBV functions (special bounded variation) J, defining K to be the set of jumps of J or the singular part of ∇J. This led Ambrosio to a proof that E admits a weak minimum for such a J (see [A]). Morel and Solimini take a different approach by developing several key a priori inequalities on the densities of minimizers K first. This leads them directly to the existence of a "less weak" minimum of E, namely, one for which K is a rectifiable *Ahlfors* set, meaning one for which uniform upper and lower bounds exist for their densities:

$$c^{-1}r \leq \mathcal{H}^1(K \cap B(x,r)) \leq cr$$

for a fixed c and for every disk $B(x,r)$ in D such that $x \in K$. The key tool in their work is to compare a segmentation with a given K with one obtained from this by *excision*. This means that $K \cap B(x,r)$ is removed from K, but a finite set of closed arcs T in the boundary $\partial B(x,r)$ of the disk are added to K:

$$K' = K - K \cap B(x,r) + T.$$

At the same time J is changed inside $B(x,r)$ but remains continuous on $\partial B(x,r) - T$.

Since the completion of the manuscript for this book, there has been dramatic progress in proving the regularity of K by Alexis Bonnet, Guy David and Ambrosio, Fusco and Pallara. Bonnet's work is still unpublished, but I want to describe a key idea in his method which reintroduces multiscale ideas: what he does is blow up K infinitely around any point $x \in K$ while suitably rescaling J and show that some subsequence approaches a limit $\overline{K}, \overline{J}$ which solves a much simpler variational problem without an image I. More precisely, he shows that there exists $\varepsilon_n \to 0$, locally constant functions J_n on disks $D(x,\varepsilon_n) - K \cap D(x,\varepsilon_n)$ and $\eta_n \to 0$ with $\eta_n/\varepsilon_n \to 0$ too such that

- $\frac{1}{\eta_n} K$ converges to a closed set $\overline{K} \in \mathbf{R}^2$ and
- $\frac{1}{\sqrt{\eta_n}}(J(\eta_n x) - J_n(\eta_n)) \to \overline{J}(x)$ where $\overline{J}$ is defined on $\mathbf{R}^2 - \overline{K}$.

Then $(\overline{K}, \overline{J})$ "locally minimizes" the improper functional

$$\overline{E}(\overline{K}, \overline{J}) = \int_{\mathbf{R}^2 - \overline{K}} \|\nabla \overline{J}\|^2 + \mathcal{H}^1(\overline{K})$$

in the sense that any change of $(\overline{K}, \overline{J})$ on a *compact* subset B of $\mathbf{R}^2$—which does not connect components of $\mathbb{R}^2 - \overline{K} - B$ that are disconnected in $\mathbb{R}^2 - \overline{K}$—cannot decrease the finite part of this functional given by the points in this compact set.

DeGiorgi conjectures in 1989 that the local minima of $\overline{E}$ are

1. $\overline{J}$ constant, $\overline{K} = \varnothing$;
2. $\overline{J}$ locally constant, $\overline{K}$ a line;
3. $\overline{J}$ locally constant, $\overline{K}$ three half lines meeting at a point with angles $2\pi/3$;
4. $\overline{K}$ a half line, $\overline{J} = c\sqrt{r}\cos(\theta/2)$, where r, θ are polar coordinates in which $\overline{K}$ is the positive x-axis.

Bonnet has proved this under the additional assumption that $\overline{K}$ is connected. As a consequence, he can prove that if K_0 is an *isolated* component of the original K,

then K_0 is a finite union of C^1-arcs, as conjectured. Whether all the components of K are isolated remains open however.

References

[A] L. Ambrosio, *Variational problems in SBV and image segmentation*, Acta Appl. Math. **17** (1989) MR **91d:**49003

[B-Z] A. Blake and Z. Zisserman, *Visual reconstruction*, MIT Press, 1987. MR **89k:**92086

[Ceta1] D. Cooper, H. Elliott, F. Cohen, L. Reiss, and P. Symosek, *Stochastic boundary estimation and object recognition*, Image Modeling (A. Rosenfeld, ed.), Academic Press, 1981.

[D-Z] B. Dubuc and S. Zucker, *Indexing visual representations through the complexity map*, Proc. 5th Int. Conf. Comp. Vision, IEEE Comp. Soc. Press, 1995.

[Fal] K. Falconer, *The geometry of fractal sets*, Cambridge Univ. Press, 1985. MR **88d:**28001

[Fau] O. Faugeras, *Three-dimensional computer vision*, MIT Press, Cambridge, MA, 1993.

[G-G] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Trans. Patt. Anal. Mach. Int. **6** (1984).

[Gr] U. Grenader, *General pattern theory*, Oxford Univ. Press, 1995.

[vH] H. von Helmholtz, *Physiological optics*, orig. German edition, 1909; Dover translation, 1962.

[K] J. Koenderink, *Solid shape*, MIT Press, 1990. MR **91f:**65040

[Mc] Ernst Mach, *The analysis of sensations and the relation of the physical to the psychical*, orig. German edition, 1895; Dover translation, 1959.

[Mr] D. Marr, *Vision*, Freeman and Co., 1982.

[Mu] D. Mumford, *Pattern theory*, Proc. 1st European Congr. Math., Birkhäuser, Boston, 1994. MR **1:**341 824

[M-S1] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, Proc. IEEE Conf. Comp. Vis. Pattern Recognition, 1985.

[M-S2] ———, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math. **42** (1989). MR **90g:**49033

[M-Z] J. Mundy and A. Zisserman, *Geometric invariance in computer vision*, MIT Press, 1992. MR **94d:**68114

David Mumford

Harvard University

E-mail address: mumford@math.harvard.edu