Artificial Intelligence 59 (1993) 81–87
Elsevier

ARTINT 1000

"Determining optical flow": a retrospective

Berthold K.P. Horn

MIT, Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139, USA

B.G. Schunck

Artificial Intelligence Laboratory, University of Michigan, 144 ATL, 1101 Beal Avenue, Ann Arbor, MI 48109-2110, USA

1. Overview

Our work on estimating optical flow [8] was important not because it solved (in some limited way) the problem of estimating optical flow, but because it represented the start of the variational approach to machine vision. The variational approach to machine vision problems was applied to shape-from-shading soon afterwards (Ikeuchi and Horn [10]), and has since found its way into many areas of machine vision. The variational approach lends itself particularly well to methods that are not feature-based. It provides a way of taking into account contributions from all parts of the image rather than just special isolated points, and it makes it possible to incorporate prior knowledge about what may be expected in particular imaging situations. It can also suggest methods for solving vision problems using nonlinear analog networks (Horn [6]), and it may lead to new ways of integrating multiple cues in images (Thompson [15]).

2. Origins

The idea behind the "optical flow" paper took form in the summer of 1978 when I was invited to the Motion Vision Laboratory established by

Correspondence to: B.K.P. Horn, MIT, Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139, USA. E-mail: bkph@ai.mit.edu.

0004-3702/93/$06.00

Hans-Helmut Nagel at the University of Hamburg. I had no real exposure to motion vision problems before this. My interest was heightened by a paper presented by Clocksin that summer at a conference in Hamburg [3] that irritated me for some reason.

I was struck right away by the local ambiguity in recovering "optical flow", which I formulated as a linear constraint between the two components of the optical flow, with the derivatives of brightness as coefficients. This "brightness change constraint" can be conveniently exhibited as a line in velocity space. The local ambiguity inherent in this constraint was later termed the "aperture problem" (Marr [11]). An aside: this illustrates (a) how important it is to give an evocative name to an idea, and (b) how the simplest ideas are the ones that stick in peoples minds—not the ones that *you* think are important.

In attempting to deal with the local ambiguity it occurred to me that, in many cases, flow at nearby places in the image will be similar. But this is not a hard constraint that could be easily exploited. And the "constraint-propagation" methods popular in the field in the early seventies did not appeal to me. I started instead to think of the problem in least-squares terms, where there would be penalties for violating the basic brightness change constraint equation, and also penalties for having a flow that varied "too quickly" from place to place—the latter needed simply because the first "constraint" was not enough to provide a unique solution.

It took a bit of time, however, before I realized that this was really a problem in the calculus of variation (not too surprising, given that I knew nothing about the calculus of variation). While trying to read the work of Courant and Hilbert [4], I discretized the problem so that I could make some progress using traditional least-squares methods. After solving the discrete version, I would grovel over the local weighted sums that appeared, in order to try and guess what partial derivative operators were being approximated by these "computational molecules"!

Even-order partial differential equations of similar form to those arising in this variational problem were not entirely foreign to me, since we had used Poisson's equation (thin membrane) and the biharmonic equation (thin plate) earlier for interpolation of digital terrain models (DTMs) from contours (topographic maps) in our work for DARPA on image understanding—hill shading in particular [5].

Upon my return from Hamburg to MIT, I had Anni Bruss summarize the calculus of variations for me, and handed the problem over to Brian Schunck, who later did his Ph.D. thesis on this topic. We checked the basic ideas mostly on synthetic data, using various shapes with sinusoidal kinds of "textures". The images were created using a ray-tracing approach—as usual, this "forward optics" part was much simpler than the "inverse optics" part.

3. Variational approach to machine vision problems

The interesting part of this work was, in my opinion, the realization that tasks in machine vision can be posed as variational problems. A number of other tasks were later tackled in this way, starting with the shape-from-shading problem, which Katsushi Ikeuchi and I worked on using a penalty function containing the error in image brightness and again a "lack-of-smoothness" term [10].

There are several things to say about the variational approach to machine vision. First of all, it typically is not feature-based. Marvin Minsky told me, when Patrick Winston, Gene Freuder and I completed the "copy demo" in December 1970, that the time had come to leave the blocks world behind. To me this meant forgetting about edges. Most of my work has focused on image cues other than edges, since "edge detection" is only really a well-defined problem in a world of polyhedral shapes. (Note that I said "well-defined", not "well-posed".)

Another aspect of the variational approach is that it permits one to introduce information about how imaging works, as well as prior information about what is likely to happen in the scene. One part of the error being minimized is invariably some measure of how much the image actually observed differs from an image predicted from the computed solution. This seems like an eminently reasonable term to have in the penalty function!

Another aspect of the variational approach is that it leads to methods that use information from all over the image, *not* just isolated points. This provides for more robust results, since there is the opportunity for many small errors to more or less cancel out—an error at one pixel is not usually catastrophic. True, information from some areas, where brightness changes rapidly, may be more important than that from others, but why draw a line and say that some of this information should be thrown out altogether. Instead, just weight it in such a way that contributions from "important" areas have a stronger influence. The least-squares approach allows one to formalize this notion, and actually get the optimal weighting automatically.

4. False leads

There were several attempts to analyze our method and to improve upon it. Not all were productive. One notion was to note that "edges" provide strong constraint in one direction ("grey-level corners" provide strong constraint in two directions). So the propagation of velocities should perhaps be anisotropic—favored in certain directions. But the original formulation already took care of that, what looks like an isotropic "smoothing" interpolation is actually distinctly directional, and in just the right way—and how could it be otherwise? After all, the method is solving the least-squares problem by finding the optical flow that gives the best fit.

Some efforts to replicate our results, particularly on real data, failed because of lack of attention to basic numerical analysis. For example, it is important that the brightness derivatives be estimated so as to refer to the same point in time and space, and that the estimator do some smoothing, yet not lose locality. The estimator we used was described in detail in the paper, but some people simply used forward differences. Also, there are serious aliasing effects that may occur when the images moves more than a fraction of the wavelength of the dominant components of the scene "texture". Estimating derivatives from aliased data is not a productive activity. Not unexpectedly, there is a certain range of velocities over which these types of method work well. For large disparities between successive images one has to resort to a feature matching method.

5. Unfortunate side-effects

A few years after the birth of the variational approach to machine vision, Tomaso Poggio noticed that some of the variational methods used could be viewed as regularization of ill-posed problems. This has led to at least two mistaken ideas:

(1) that the variational approach *is* regularization, and
(2) that all vision problems are ill-posed.

To some extent the second error was fueled by the approach that Ikeuchi and I took in the paper "Numerical shape from shading and occluding boundaries" [10] where we abandoned the integrability constraint—because we couldn't find a convergent iterative scheme based on it—and instead used a departure-from-smoothness penalty term—quite analogous to that appearing in the optical flow method. This is basically the approach one would take if shape-from-shading was an ill-posed problem. But it is not, as has been forcefully pointed out by John Oliensis [12], Bror Saxberg [13], and Mike Brooks (and as should actually also be apparent from the original solution involving characteristic strips).

6. What was left out

Perhaps the main omission in the approach taken to optical flow at that time was the neglect of boundaries between different regions moving differently. The variational approach chosen was based on the idea that flow at neighboring places in the image is similar—without exception. Not surprisingly, large errors can occur at occluding boundaries in the image of rigid body motion—as illustrated in the original paper. Independently moving objects will lead to errors where one obscures another.

The reason the segmentation problem was not addressed in the original paper is that we had no reasonable ideas about how to solve it, other than simply omitting contributions to the departure-from-smoothness penalty term in places where there appeared to be rapid changes in the estimated optical flow. But this was neither a principled approach nor particularly effective in practice (although several papers have been written since then that basically pursue this heuristic).

Recently a number of approaches to the flow segmentation have arisen that look more promising. Starting with the idea of a line process and simulated annealing—which is not computationally reasonable—Andrew Blake, Christof Koch, John Harris, Tomaso Poggio and others (see [1]) have developed approximations thereof that are computationally tractable.

Another thing we did not do enough of in the paper is an analysis of what the sources of errors might be and what circumstances contributed to successful recovery of optical flow. We obviously knew something about this, since we chose experimental image sequences that worked! But we did not say enough about when the algorithm would not produce useful output. And in the tradition of machine vision programs based on single cues, there are plenty of situations that confuse this algorithm.

7. Optical flow and the motion field

One thing that I regret now in looking back at the paper is that we did not draw a clear distinction between what I would now call the "motion field" and the "optical flow". The optical flow is a velocity field in the image that transforms one image into the next image in a sequence. As such, it is not uniquely determined. One needs to add additional constraint to obtain a particular "optical flow". The motion field, on the other hand, is a purely geometric concept, without any ambiguity—it is the projection into the image of three-dimensional motion vectors. One endeavours to recover an optical flow that is close to the motion field—which is what one would really like in order to estimate shapes and motions. Much confusion has resulted from a lack of distinction between these two quite distinct concepts.

8. Where did the work go from there

There are really two "directions" to discuss. The first is in work on motion vision *per se*. Here my own work with Anni Bruss [2], Shariar

Negahdaripour [7] and Ned Weldon [9] focused on introducing the rigid body constraint. This is such a powerful constraint, reducing the "number of unknowns" from two at every picture cell (components of flow) to roughly one at every picture cell (depth alone), that it would be silly not to exploit it when it applies. With rigid body motion, the problems are no longer variational, but instead "ordinary" (albeit complex and nonlinear) least-squares problems. And in several special cases (such as pure rotation, pure translation, motion with respect to a planar surface), closed form solutions exist. Again, the problem of segmentation was not addressed in our work, because it throws a real spanner in the works. Yet in many practical situations, depth discontinuities and independently moving objects require that the scene be segmented before the methods we developed can be applied to the regions therein.

The second "direction" is broader, that of variational methods for machine vision in general. Variational methods have now been brought to bear on a variety of problems, including optical flow, shape-from-shading, interpolation from sparse data and yes, even edge detection and segmentation.

9. What can be built more easily as a result of the paper?

The optical flow method described in the paper allows the flow velocity to vary from place to place in the image—albeit slowly. An even simpler situation is one where the flow is the same everywhere—as might be a reasonable assumption if we consider a small enough patch of an image. This problem has a simple closed form solution involving two linear equations in two unknowns, which is variously attributed to Ned Weldon or Hans-Helmut Nagel, although it was also a long-standing homework problem in my course on machine vision here at MIT.

John Tanner and Carver Mead at Caltech built an analog VLSI circuit that solves this simplified optical flow problem [14]. They did it using a feedback scheme quite analogous to gradient descent, rather than by working with the closed form solution directly. This is particularly nice, since a simple extension (replacing a global bus with a resistive network) would solve the variational problem, where flow *is* allowed to vary from place to place. A circuit for doing this has not been built, to my knowledge, but is well within the state of the art.

More interesting perhaps than computation of optical flow itself is the recovery of rigid body motion and a depth map, if possible. Work on "direct" methods (that is, based on derivatives of brightness at all picture cells) for motion vision has led to some schemes that lend themselves to implementation in analog VLSI hardware. The pure rotation case is so simple (and has so little application in practice) that nobody has bothered to built a chip for it. Pure translation is much harder, and Ignacio Sean McQuirk is now building an analog chip here at MIT to find the focus of expansion.

Such efforts may lead to more complex chips that find both translation and rotation of a camera in a fixed environment. These should be useful in robotics and guidance of vehicles.

References

[1] A. Blake and A. Zisserman, *Visual Reconstruction* (MIT Press, Cambridge, MA, 1987).

[2] A.R. Bruss and B.K.P. Horn, Passive navigation, AI Memo 662, MIT, Cambridge, MA (1981).

[3] W.F. Clocksin, Determining the orientation of surfaces from optical flow, in: *Proceedings AISB Conference*, Hamburg, Germany (1978).

[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1937/1953).

[5] B.K.P. Horn, Hill shading and the reflectance map, *Proc. IEEE* **69** (1) (1981).

[6] B.K.P. Horn, Parallel analog networks for machine vision, AI Memo 1071, MIT, Cambridge, MA (1988).

[7] B.K.P. Horn and S. Negahdaripour, Direct passive navigation: analytical solution for planes, *IEEE Trans. Pattern Anal. Mach. Intell.* **9** (1) (1987) 168–176.

[8] B.K.P. Horn and B.G. Schunck, Determining optical flow, *Artif. Intell.* **17** (1981) 185–203.

[9] B.K.P. Horn and E.J. Weldon Jr, Direct methods for recovering motion, *Int. J. Comput. Vis.* **2** (1) (1988) 51–76.

[10] K. Ikeuchi and B.K.P. Horn, Numerical shape from shading and occluding boundaries, *Artif. Intell.* **17** (1981) 141–184.

[11] D. Marr, *Vision* (Freeman, San Francisco, CA, 1982).

[12] J. Oliensis, Existence and uniqueness in shape form shading, Ph.D. Thesis, University of Massachusetts, Amherst, MA (1989).

[13] B.V.H. Saxberg, A modern differential geometry approach to shape from shading, Ph.D. Thesis, Electrical Engineering and Computer Science Department, MIT, Cambridge, MA (1989).

[14] J.E. Tanner and C.A. Mead, An integrated optical motion sensor, *VLSI Signal Processing* II (Proceedings of the ASSP Conference on VLSI Signal Processing), Los Angeles, CA (1986) 59–76.

[15] C.M. Thompson, Robust photo-topography by fusing shape-from-shading and stereo, Ph.D. Thesis, Mechanical Engineering Department, MIT, Cambridge, MA (1993).