Berthold K. P. Horn
The Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Non-Correlation Methods for Stereo Matching

Optimization methods and feature matching are being investigated.

Attempts at automation of the stereo compilation process have led to the development of a number of machines which are able to produce good results in many cases, particularly if aided by a human operator. On difficult terrain, they fail. These machines are based on modifications of the correlation idea for finding the disparity. The assumption has been that further refinements of this approach will solve the remaining problems or that they cannot be solved.

There are methods other than correlation available for identifying corresponding points. Some depend on optimization of the match found with varying distortions of the two images (but without using correlation)—others symbolically match features extracted from the image. Research on several alternate proposed methods is showing promise. Some approaches even lend themselves to implementation in special purpose hardware to attain rapid throughput.

Why has there been such single minded concentration on correlation and its modifications for matching image patches? One reason is speed. It was assumed that a machine had to be fast to be useful and that only simple computations could be performed at the required speed. Correlation lends itself to implementation in special-purpose hardware or more general-purpose parallel computers. This emphasis on speed was perhaps premature. A machine which rapidly produces output which requires extensive post-editing is not contributing to increased productivity.

If the two images were simply shifted versions of one another, then the integral of the square of their difference (with one image shifted) can be made equal to zero. This integral thus has a minimum when the images are shifted the correct amount relative to one another. Correspondingly, the correlation (the integral of their product) has a maximum. Does this maximum correspond to a unique displacement? If the two image patches are uniform in brightness, then unfortunately a whole range of displacements will do. Also, if the image patches contain a periodic pattern, there are multiple maxima. So there are problems, even when the images are just shifted versions of one another.

Even more serious is the fact that the two images are in fact *not* shifted versions of one another. For large base-height ratios in hilly terrain, or in built-up areas, some surfaces may be seen only in one image. Those that are seen in both will appear with different foreshortening in the two images; that is, the surfaces appear compressed along one dimension to differing degrees. Simple correlation will not work here. At the least, a kind of rubber-sheet distortion has to be introduced so that the patches being compared are in fact similar.

Even if this is taken into account, one is still faced with the fact that the same patch on the surface will appear with different brightnesses when viewed from two different directions. Only if the bidirectional reflectance distribution function (BRDF) happens to be independent of the direction of reflection will the image irradiance be the same in the two images (ignoring atmospheric effects). Specular reflections from water surfaces and hot-spots (black reflections) from vegetation canopies (and the atmosphere) are extreme examples of violations of this criterion. If a surface is viewed with a device which does not resolve surface patches of different orientation, then the brightness is a weighted average of the brightnesses of the individual patches, with weights proportional to the foreshortend areas. Thus, even if the material of a surface satisfies the above criterion, a region of it may appear with different brightnesses in different images if its surface undulations are not resolved. Shadows too small to be resolved contribute to this effect.

These effects and others (such as noise) guarantee that the basic model on which the idea of correlation is based is not sound. The result is that operator intervention is required during

Photogrammetric Engineering and Remote Sensing,
Vol. 49, No. 4, April 1983, pp. 535-536.
0099-1112/83/4904-0535$02.25/0

stereo compilation, that tedious post editing is needed and that there can be complete failure on really difficult areas like lakes, and parts of some deserts for example.

At this point it may be worth pointing out, too, that disparity is determined accurately at some points, where the grey level is "busy," and poorly determined in areas of "low contrast." The disparity in the latter regions must be "filled in" from surrounding, better known, values. This also suggests filtering to emphasize higher spatial frequencies.

A difficult issue is the choice for the size of the patch to be correlated. If it is too large, information from features at different disparities is smeared together, giving at best a kind of average. If it is too small, there may not be enough grey-level variation to lock into. This is analogous to the problem of determining the support for an edge detector. Working at several scales, using coarse resolution to guide the finer resolution seems to be the answer. Some existing machines use these ideas in some form.

It is necessary to take a look at methods other than correlation. This may be hard to believe because the idea that stereo means correlation is so well entrenched. As in some other fields, this one started with an under-estimate of the difficulties and premature announcement of success. It was assumed that simple methods that work sometimes could be refined to always work. This, and secrecy due to rivalry, stiffled further research. Only now are we admitting that the problem has *not* been solved.

What are the alternatives? There are many. Some retain the idea of grey-level matching but use methods other than correlation to achieve it. These include optimization methods such as dynamic programming and relaxation computations in disparity arrays. Others work on matching features. These can be edges, "corners," areas of high grey-level curvature, or more complicated elements of the image. Edges may for exaple be located by looking for the zero-crossings in the Laplacian of the image filtered by convolution with a Gaussian. The important point is that matching now is a symbol manipulation operation. Such a method has properties quite different from those of correlation of grey-levels. Symbolic matching is affected very little by differences in grey-level in the two images or by foreshortening, for example. It also makes explicit the interpolation needed to fill in the surface model between those places where disparity can be accurately determined.

Stereo compilers are not yet robust enough to be truly called "automatic." The fix may not lie in further refinements of correlation-based methods. We should worry about speed *after* a robust method is developed. It is time to do some more basic research on this problem. There are alternatives to correlation.

Bibliography

Arnold, R. D., and T. O. Binford, 1980. Geometric Constraints in Stereo Vision, *Proceedings, SPIE*, San Diego, California.

Baker, H. H., 1980. Edge Based Stereo Correlation, *Proceedings ARPA Image Understanding Workshop*, University of Maryland, pp. 168-175.

Baker, H. H., and T. O. Binford, 1981. Depth from Edge and Intensity Based Stereo, *Proceedings, Seventh Joint Conference on Artificial Intelligence*, Vancouver, B.C.

Gennery, D. B., 1977. A Stereo Vision System for an Autonomous Vehicle, *Proceedings, Fifth Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts.

Grimson, W. E. L., 1981a. A Computer Implementation of a Theory of Human Stereo Vision, *Phil. Trans. Roy. Soc. of London*, Ser B292, pp. 217-253.

———, 1981b. *From Images to Surfaces*, MIT Press, Cambridge, Massachusetts.

Levine, M. D., D. A. O'Handley, and G. M. Yagi, 1973. Computer Determination of Depth Maps, *Computer Graphics and Image Processing*, Vol. 2, pp. 131-150.

Marr, D. and T. Poggio, 1976. Cooperative Computation of Stereo Disparity, *Science*, Vol. 194, pp. 283-287.

———, 1979. A Computational Theory of Human Stereopsis, *Proc. Roy. Soc. of London*, Ser B204, pp. 301-328.

Mayhew, J. E. W. and J. P. Frisby, 1981. Psychophysical and Computational Studies Towards a Theory of Human Stereopsis, *Artificial Intelligence*, Vol. 17, No. 1-3, pp. 379-386.

Moravec, H. P., 1980. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*, Stanford A. I. Memo AIM-340.

Movi, K., M. Kidode, and H. Asada, 1973. An Interative Prediction and Corrction Method for Automatic Stereo Compassion, *Computer Graphics and Image Processing*, Vol. 2, pp. 393-401.

Nagel, H. H. and W. Enkelman, 1982. *Investigation of Second Order Grey Variations to Estimate Corner Point Displacements*, ICPR.