MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Vision Memo. No. 121
Artificial Intelligence

February 1967

Estimating Stereo Disparities

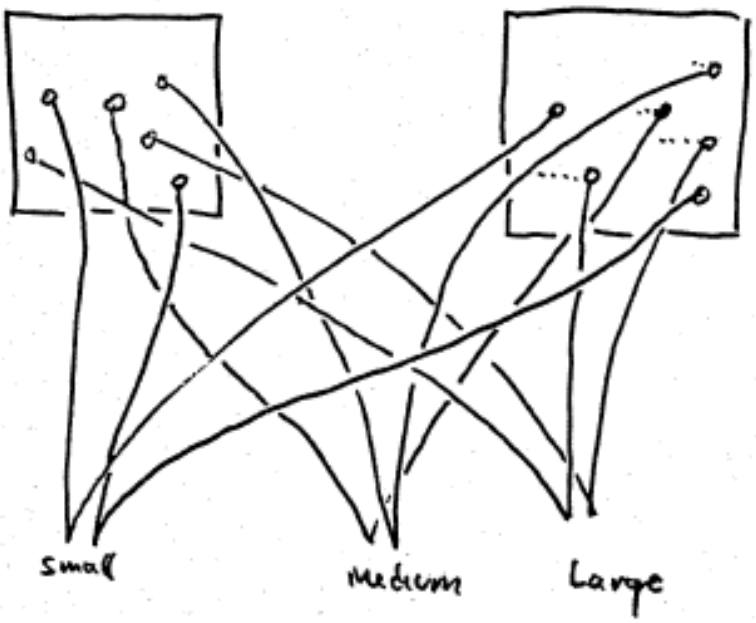
Marvin L. Minsky

Estimating Stereo Disparities

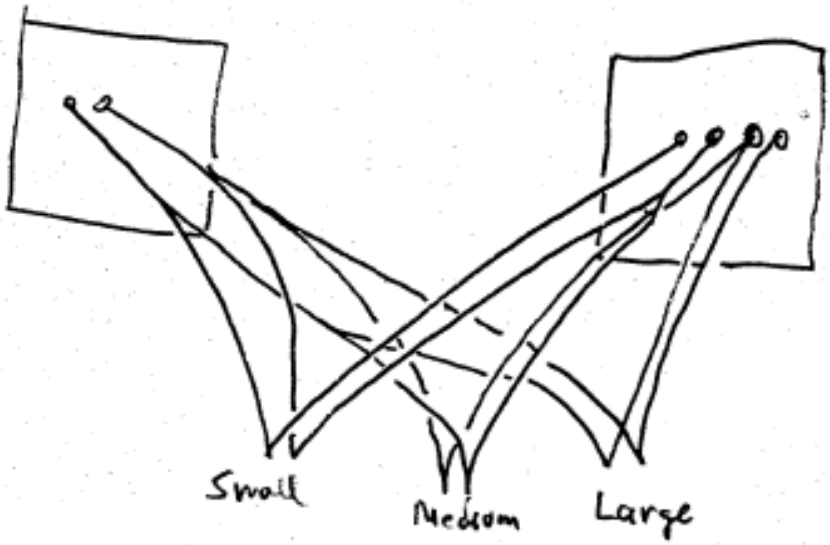
An interesting practical and theoretical problem is putting bounds on how much computation one needs to find the stereo-disparity between two narrow-angle stereo scenes. By narrow angle I mean situations wherein the angle subtended by the eyes is a very few degrees: the kind of correlation-disparity method discussed here probably isn't applicable to the wide-angle stereo we'll usually use for scene-analysis in the Project.

1. The method we consider is to find the local maximum of local correlation between the left and right scenes, over a range of displacements along the eye-eye axis. Obviously this is a simple-minded method that will fail in certain situtaions: here we are not interested in bad cases so much as in getting estimates of the minimal computation in the favorable situations.

A correlation can be considered as a properly-normalized sum of pairwise products of intensities (or other surface functions). The correlation, for each disparity d, is obtained by using pairs that are d units apart in visual angle, referred to a standard azimuth scale in each eye. One can imagine a scheme in which the pairs are all different in the retinas:



or schemes in which fewer points are used more effectively:



2. Now one kind of theory might take a "physiological" form, and postulate that some kind of global correlation is performed that cross-correlates the whole foveal area, or a large part of it. Perhaps there is an anatomical brain region where the two visual fields are mapped and compared, and the correlations emerge from millions, or many thousands, of products. But it is interesting to ask whether one can do well with fewer. Let the goal be to get just information enough to tell an oculomotor servo which way to move. This suggests dividing the correlated pairs into groups with disparity ranges increasing with distance from zero:

The measurements of large disparities could be made with cells of larger "receptive fields." Perhaps ten such ranges, with perhaps twenty or thirty cell pairs in each, would produce reasonable servo-convergence. The servo is operated, presumably, by the group with the highest product sum. One could also try linear summation to the servo, but then I would expect much noisier performance.

Proposal:

1. Study it experimentally, on a variety of scenes.

2. Study it mathematically. The scene-characterizing problem might be a mathematical stumbling-block.