

DTIC FILE COPY

2

**COMPUTATION OF STEREO AND VISUAL MOTION:
FROM BIOPHYSICS TO PSYCHOPHYSICS**

Principal Investigators: Tomaso Poggio and Ellen Hildreth

Other Primary Investigators: Norberto Grzywacz and Heinrich Bülthoff

Center for Biological Information Processing
Massachusetts Institute of Technology

PROGRESS REPORT, FOR PERIOD 4/1/88 — 9/30/88

DTIC
ELECTRIC
OCT 24 1988
D

AD-A201 873

I. Overview

During the second and third quarter of this "stretchout" year of funding, we have continued to explore a number of problems in motion analysis, including the parallel detection of motion using a correlation-based mechanism, motion correspondence, neural mechanisms for motion detection and measurement, and the recovery of 3-D structure and motion. We are also starting to focus more deeply on the integration of multiple visual cues. Described here is some work on the interaction between surface shape, albedo, and the illuminant direction. Finally, we summarize our continuing theoretical and experimental work on the analysis of color.

II. Scientific Accomplishments*Network Models for the Velocity Field and its Discontinuities*

As we noted in our previous report, we are developing and testing some variations on a parallel network model recently proposed by Hutchinson, Koch, Luo and Mead for combining the computation of the smoothest velocity field with line processes (suggested by Geman and Geman) for handling motion discontinuities. Our modified network derives the initial motion measurements only at the locations of significant intensity changes, allows greater flexibility in the placement of the discontinuities and considers variations on the energy function being minimized to implement the smoothness constraint. The network is also designed in a way that more closely parallels physiological properties of motion-sensitive neurons in area MT of monkey visual cortex. We are running simulations of this modified network model, with encouraging results. [Ando and Hildreth]

Effects of Past Motion on Motion Measurements

We extended our theoretical studies on how to include past motion information on the measurement of present motion. It was shown that if one applies past motion as if Newtonian inertia applies to the image features, incorrect results are obtained. The

DISTRIBUTION STATEMENT 4

Approved for public release

Distribution Unlimited

88 10 4 03

reason is that constant velocity in the visual world does not correspond to constant velocity in the image. A better alternative is to assume that image features tend to maintain a constant direction of motion. (This alternative is consistent with our previous psychophysical experiments.) Calculations show that constant velocity in the visual world leads to motion with constant direction in the image. We proceeded to generalize our recent theory of motion coherence to include constant motion direction. [Grzywacz, Smith, and Yuille]

Effects of Spatial Coherence on Motion Measurements

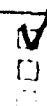
A new problem to which we applied our recent motion coherence theory is the aperture problem. An example that illustrates the rationale is the motion of a rigid contour. Local motion detectors can only measure motions that are perpendicular to points in the contour. But the contour is one object, and as such, its points are moving coherently. Thus, if the visual system imposes coherence, it may find the real image motion from the "wrong" measurements of the local detectors. We found that this is indeed the case for a large range of the model's parameters. It was also shown that the solution is not fully correct, but its mistakes can be made arbitrarily small by proper choice of parameters. Our method to solve the aperture problem is different than the one proposed by Hildreth, which imposes smoothness along contours. The new method imposes smoothness (coherence) across the entire space. [Grzywacz and Yuille]

Using Time-to-Collision Estimates to Recover 3-D Motion and Structure

We are testing two models for recovering the 3-D trajectories of moving objects from their changing 2-D projection. The first is an instantaneous model that builds up the 3-D trajectory incrementally as time progresses. The second also recovers the global parameters of the trajectory, assuming that the object's motion obeys the simple laws of physics of free-falling objects. The latter scheme allows us to predict future trajectory positions. We are currently running simulations to test whether these models can correctly recover object trajectories under ideal conditions, and are just beginning to implement a version of the first model that will allow us to analyze natural imagery. We are conducting perceptual studies to test the validity for the human visual system of both models for recovering 3-D object trajectories. Our experiments use a simple computer game to test whether this changing size information is used in a quantitative way, and whether the human system also makes explicit use of the laws of physics of free-falling objects to recover these trajectories (for example, the direction and magnitude of acceleration due to gravity). [Hildreth and Mistle]

Computing Observer Heading

Psychophysical experiments by B. Warren at Brown University indicate that human observers can recover their 3-D direction of heading with an accuracy of 1-2 degrees.



per lti

A-1

We first conducted simulations of an algorithm that uses time-to-collision estimates to recover rough 3-D heading, and found that the algorithm is very accurate when the observer's heading is close to the direction toward a target in the scene around which the heading calculation is performed, but the algorithm degrades with deviation away from this heading. Although not viable for recovering absolute 3-D heading in general, this algorithm may be very reliable at judging whether the observer is moving to the left or right (or above or below) of a target in the scene. It may then be possible to embed this scheme in a larger algorithm that incrementally improves the computation of heading direction through the repeated use of this scheme that can accurately determine the rough direction of motion of the observer relative to targets in the scene. We are currently formulating such an algorithm. [Hildreth]

Structure-from-Motion and Surface Interpolation

Recent psychophysical studies by R. Andersen at MIT suggest that when presented with sparse image features in motion, the human visual system interpolates a smooth 3-D surface consistent with the derived relative depths of the features, and that this filled-in surface further constrains the subsequent recovery of the 3-D structure of the discrete features at a later time. We are examining a structure-from-motion algorithm that combines Ullman's incremental rigidity scheme for recovering the structure of the discrete features (such as intensity edges) with a surface interpolation mechanism. The reconstructed surface at each moment provides the initial depths for newly appearing image features at a later moment. We expect this model to account for Andersen's experimental data. [Hildreth]

The Perception of Transparency and the Coupling of Stereo and Motion Information

It is well-known that the human visual system can reconstruct depth from simple random-dot displays given disparity or motion parallax information. This fact has lent support to the notion that the recovery of structure from stereo and motion relies on low-level primitives or tokens (e.g., edges) derived from image intensities. In contrast, the judgement of surface attributes such as transparency or opacity is often considered to be a higher-level visual process that would make use of low-level stereo or motion information, and later perhaps recognition to tease apart transparent and opaque parts. This is exemplified by the notorious lack of success of computational studies to deal with transparency, compared to the success of a number of algorithms to solve structure from motion or stereo. We provide demonstrations that question the above view by showing that apparent opacity can veto both the rigidity assumption in structure from motion, and prevent either fusion, or seeing the correct depth relations in stereo. More generally, this suggests that the correct classification of edges according to their physical causes may have to be done either before, or simultaneous with the computation of structure from motion and stereo.

In one demonstration we made a stereogram of a transparent "Mondrian" in front of an opaque rectangle. It is difficult for most observers to see the Mondrian as transparent. This inability prevents the observer from seeing the proper depth relations – the Mondrian looks either behind the opaque rectangle, or appears to be in the same plane. Using the Symbolics color system we have produced an analogous display in motion. Instead of two views, the transparent Mondrian and the opaque rectangle are rocked back and forth in time, introducing motion parallax information. The perception is bistable – when the observer sees the Mondrian incorrectly as opaque, it appears in the same plane as the opaque rectangle and appears to slip over it; when the observer sees the Mondrian as transparent, the depth relations are seen veridically. [Bülthoff and Kersten]

Shape from Shading and Specularities

Andrew Blake from Oxford University visited the Center during June and July to collaborate with Heinrich Bülthoff on the problem of shape from shading and specularities. In particular, we explored the problem of whether the appearance of highlights in stereoscopic views influences perceived stereoscopic shape. We used the Symbolics S-Geometry package to generate surfaces of revolution with realistic shading and texture generated by Fourier methods. Initially, while looking for geometric effects, we found instead strong effects involving perceived transparency and glossiness, occurring when the physics of specular reflection is violated. We are currently devising forced-choice procedures to test these effects. Influence of highlights on geometry seems to be more elusive. We are trying to quantify it by using "concave/convex inversion" as a measuring probe. [Blake and Bülthoff]

Qualitative Depth and Shape from Stereo

The motivation to consider more qualitative information that can be obtained from stereo vision is two-fold: First, once the correspondence problem (that is, matching the two images and obtaining disparities) is solved, camera calibration leaves the computation of depth from disparity still a difficult problem. Most of the research on early stereo vision concentrate on this difficult problem, which is not yet closed. Second, human vision seems to suggest the use of some qualitative depth information which is easier to obtain and more robust. In the present research we show that qualitative depth and shape information can be obtained from stereo disparities with almost no computations, and with no prior knowledge (or computation) of camera parameters. First, we derive two expressions that order all matched points in the images in two distinct depth-consistent ways from image coordinates only. Using one for tilt estimation and point separation (in depth) demonstrates some anomalies and unusual characteristics that have been observed in psychophysical experiments, most notably the "induced size effect". To explain this effect, no specific depth computation scheme or specific

correspondence algorithm need be assumed, no asymmetric convergence is implied, opposite induced-effects in neighboring spatial regions are allowed for, and deterioration is expected near the meridians, as observed in humans. Other predictions are the following: (1) The observed deterioration in depth separation for points coplanar with the fixation point; (2) The observed increase in stereo-acuity thresholds with the increase in standing disparity or elevation angle of an oblique line (the "cosine-rule"); (3) An anisotropy between horizontal and vertical dimensions, as observed in experiments of depth-acuity threshold estimation; and finally, (4) the fact that exact tilt and relative depth can be readily obtained from the above orders, depending only on the angle of convergence of the eyes, agrees with the empirical evidence for the dependence of relative depth perception on extraretinal perception of this angle. The above model is *not* the only computational explanation for these last effects, though. Second, we apply the same approach to estimate some qualitative behavior of the normal to the surface of any object in the field of view. More specifically, we follow changes in the curvature of a contour on the surface of an object, with either x - or y -coordinate fixed. In a similar way we develop an algorithm to compute axes of zero-curvature from disparities alone. The algorithm is shown to be quite robust against violations of the basic assumptions of the computation on synthetic images, with relatively large controlled deviations. It performs almost as well on real images, as demonstrated on an image of four cans at different orientations. [Weinshall]

Integration of Vision Modules and Labelling of Surface Discontinuities

We assume that a major goal of the early vision modules and their integration is to deliver a *cartoon* of the discontinuities in the properties of the surfaces around the viewer and to label them in terms of their physical origin. The output of each of the vision modules is noisy, possibly sparse and sometimes not unique. A restoration stage is needed to counteract the noise and fill-in sparse data, while at the same time detecting discontinuities. In previous work (Gamble and Poggio, 1987), we suggested that a coupled MRF at the output of each module – stereo, motion, color, texture – can be used for this dual goal. Integration of the image data within each MRF helps to find discontinuities and align them. In the present work we explore the problem of how to label the discontinuities in terms of depth discontinuities, orientation discontinuities, specular discontinuities, albedo discontinuities and illumination discontinuities, i.e., shadows. We have obtained labeling results using a simple linear classifier operating on the output of the MRF associated with each vision module and to the image data. The classifier has been trained on a small set of synthetic images. Presently we are working on the elaboration of the classifier to improve labeling results. We still have to develop the theory and implementation of a single system that will find discontinuities and label them with mutual interactions. [Geiger, Poggio, and Weinshall]

The Psychophysics of Color Vision

We are continuing research on the psychophysics of color vision using a color-matching setup designed for both monkeys and humans, described in the last report. Our experiments test whether the basic assumptions underlying our computational models of color constancy are indeed valid for the human visual system in the real world. [Hurlbert and Poggio, in collaboration with Dr. Nikos Logothetis of MIT]

Learning Color Algorithms

As we reported previously, we have used optimal linear estimation techniques to synthesize a lightness algorithm from examples. (Lightness algorithms provide a solution to the problem of color constancy under restricted conditions.) In ongoing work, we are exploring several adaptive algorithms in which a segmentation operator works in concert with a lightness operator. In particular, we have developed a cooperative color algorithm based on the cooperative stereo algorithm of Marr and Poggio. The algorithm arrives at a unique reflectance value for each pixel in an image by successive iterations, helped by input from a luminance edge detector that defines the regions within the reflectance values as constant. The algorithm performs successfully on synthetic images (two-dimensional Mondrians with added noise) and on real images (images obtained with a three-color-channel calibrated CCD camera). [Hurlbert and Poggio]

Biophysical Experiments on Nonlinear Mechanisms for Retinal Direction Selectivity

We continued to study the hyperpolarizing and shunting components of the directionally selective cells' inhibition. To achieve this goal, further data analysis was done on the experiment described in the last report. The experimental stimuli were two-slit apparent motions in the null direction. The measurements were of the cell response dependence on the second slit contrast, parametric on the first slit contrast. We compared four different synaptic models with the data. These models contain an excitatory and an inhibitory synapse in different positions of the ganglion cells' dendritic tree. The inhibitory synapse may work through shunting or through hyperpolarization. A nonlinear fit from these models to the data chooses to set a significant shunting component and a zero hyperpolarizing component. This result held for ten cells irrespective of the model used. Thus, we provide strong evidence that shunting, but not hyperpolarizing inhibition underlies retinal direction selectivity. [Grzywacz and Amthor]

Theoretical studies on Nonlinear Biophysical Mechanisms for Direction Selectivity

We started to construct a mathematical model that incorporates the relevant features from the data collected recently from the rabbit retina. Our model, which is one dimensional, postulates that a directionally selective cell receives excitatory and inhibitory inputs through independent compartments. The model's assumptions on these inputs are: Each of the compartments is excited by a corresponding receptive

field point and inhibited by points to one side of it. A band-pass filter followed by a rectification mediates the excitation, since it is transient in the retina. (As suggested by anatomy, the cell has two independent receptive fields, one for ON and one for OFF stimuli. This feature is incorporated in the model as was done by Grzywacz and Koch, 1987.) On the other hand, the inhibition is sustained, and thus better modeled by a low-pass filter. The inhibition may also sometimes be rectified, since small stimulus intensity changes lead to large changes of inhibitory strength. (We modeled this high gain sustained inhibition with a negative feedforward mechanism that appears to account for photoreceptor adaptation). Also, in the receptive field side that inhibits a compartment, the inhibition spatial distribution is broad and starts at the point that excites the compartment. The directionally selective inhibition acts on the excitation as if the latter is a current and the former a resistance shunting. [Grzywacz]

Neural Implementation of the Parallel Motion Algorithm

It is natural to ask whether some version of the Parallel Motion algorithm described in previous reports may have a natural implementation in terms of cortical physiology. There is some psychophysical evidence that motion computation is done in two stages, with the first stage computing the perpendicular components of motion and the second stage combining these measurements over an extended area into a coherent motion pattern (Adelson and Movshon, 1982). There is also physiological evidence for a two-stage motion computation in primate visual cortex. Movshon *et al.* found that motion sensitive neurons in area V1 could only compute the component of motion in the direction perpendicular to the orientation of image features. These neurons only responded to one component of two superimposed sine wave gratings moving in different directions (same stimulus as in Adelson and Movshon). In area MT, however, cells have been found that appear to respond to the direction of motion of the combined pattern (pattern cells). These psychophysical and physiological observations suggest a physiological implementation of the parallel motion algorithm. We propose a scheme based on a layered structure. The neurons in these layers are fine-tuned to a specific direction and velocity (probably area V1 of visual cortex). Motion could be computed by local interactions between pairs of LGN inputs on the dendritic tree (Torre and Poggio, 1978). The "voting step" that finds the support in a neighborhood takes place in the mapping from these layers to another set of layers of neurons (possible in area MT) with a relatively large receptive field corresponding to the voting neighborhood. The final choice of the velocity with the largest support at each spatial location could also take place in MT. The algorithm seems to have much simpler physiological implementations than other schemes recently discussed (Hildreth, 1984; Yuille and Grzywacz 1988). [Bülthoff, Little and Poggio]

Parallel Motion Algorithm and the Aperture Problem

The parallel motion algorithm suffers from the "aperture problem" only in a mild form. More precisely, it will find the correct velocity field except in pathological cases (a very long straight contour with no visible terminations or features). The comparison stage is equivalent to patchwise cross-correlation, which exploits local constancy of the optical flow. As matching features, we have used zero-crossings, the Laplacian of Gaussian filtered images, their sign, and the smoothed brightness values, with similar results. It is interesting, but not surprising, that methods *superficially* so different (edge-based and intensity-based) give such similar results, since there are theoretical arguments that support, for instance, the equivalence of cross-correlating the sign bit of the Laplacian filtered image and the Laplacian filtered image itself.

The algorithm makes a general point that is of some interest given several recent papers in the area: phenomena such as motion capture are to be expected by any algorithm that integrates information about motion over local spatial neighborhoods. [Little, Bülthoff and Poggio]

III. Personnel/Financial Update

Although several major personnel changes have occurred, their effect will not actually be felt until the next quarter. Professor Tomaso Poggio has been named to the Uncas and Helen Whitaker Chair at MIT. Besides being a great honor, this appointment also brings its own funding for the recipient's salary. Dr. Heinrich Bülthoff has accepted a position as Assistant Professor in the Department of Cognitive and Linguistic Sciences at Brown University. His appointment began September 1, 1988. He will, however, continue to work with the Center as a consultant through the end of this contract. We would like to use some of the money originally budgeted for Dr. Bülthoff's salary to pay him as a consultant. Dr. Norberto Grzywacz has received a grant from the Biological, Behavioral and Social Sciences Division of the National Science Foundation. Subsequently, he has been appointed a Research Scientist at MIT, effective August 1, 1988. The contract with NSF will not interfere with Dr. Grzywacz's work under this contract, as the NSF contract will only demand part of Dr. Grzywacz's effort. As Dr. Grzywacz is no longer a Postdoctoral Fellow, he can no longer be supported through a fellowship fund. Consequently, we expect that his salary requirements will absorb the funds that would have been used for the salaries of Professor Poggio and Dr. Bülthoff.