"Making a Picture Fit the Eye: Human Engineering for Computer Graphics"

Lawrence W. Stark, Stephen Ellis*, Michitaka Hirose**, Won Soo Kim ***, Andrew Liu, Charles Neveu, An H. Nguyen, Greg Tharp and Iris Hitomi Yamashita

Telerobotics and Neurology Units,
University of California, Berkeley

INTRODUCTION

Human visual perception is largely a top-down process with a cognitive model actively driving foveated vision in a repetitive "scanpath" over subfeatures of the scene or picture of interest to check on and modify or change the working hypothesis. An image processing scheme for robotic vision has been developed based upon this metaphor.

Visual search has random elements but these are often minimal in real work cases so that searchpaths or scanpaths can be demonstrated here under controlled visual display conditions. Again, artificial computer vision, based upon a multiple adaptive matched filter algorithm, appears to be a reasonable engineering design. Use of computer search aids in defining human visual search processes.

Spatial displays for human controllers acting in manual control modes, in supervisory control modes, or as interpreters of visual direction need to be designed with careful attention to human visual processes. Visual enhancements with symbolic and analog elements can be shown to be of practical use as documented by quantitative performance tests.

Underlying neural mechanisms for visual processing are of considerable interest as scientific objects in their own right, as metaphors for artificial neural networks for artificial intelligence connectionist get-rich-quick schemes, and to illuminate behavior studies that depend upon these lower level processes. Motion detection was stimulated by the early Reichardt-Hassenstein work in the first cybernetic revolution in the 50's and now has spawned a set of expanded reichardt models. The founders of this area, Warren McCulloch and Walter Pitts, propounded not only the formal neuron and the neuro-anatomically and neurophysiolocally based model for the control of eye movements. We are now studying possible networks for blur detection and the control of accommodation, the focusing mechanism of the human eye.

The following abstract of our panel discussion area is largely depicted in the figures and the figure legends. We look forward to a stimulating interaction among the panelists and between the panelists and the audience.

* also with NASA-Ames Research Center
** now at the University of Tokyo
*** now at Jet Propulsion Labratory

CH2913-2/90/0000/0411/$01.00 – 1990 IEEE

1. SCANPATH VISION AS A METAPHOR FOR MODEL CONTROL OF IMAGE PROCESSING

Evidence for repetitive sequences of saccades was presented by Noton and Stark (1971) who used this serial process to develop a top-down theory that postualtes a sensory-motor schema or cognitive model that controls active looking. Further studies by Stark and Ellis (1981) showed that when only the mental states of a subject changed, as while looking at an ambiguous figure, the scanpath changed; Brandt and Stark (1990, in preparation) used imagined figures to demonstrate that similar scanpaths occurred even when a blank screen was observed.

FIGURE 1: Scanpath Theory for Top-Down Human Vision

Human visual perception is largely a top-down process with a cognitive model actively driving foveated vision in a repetitive "scanpath" over subfeatures of the scene or picture of interest to check on and modify or change the working hypothesis. (Stark and Krischer, 1988).

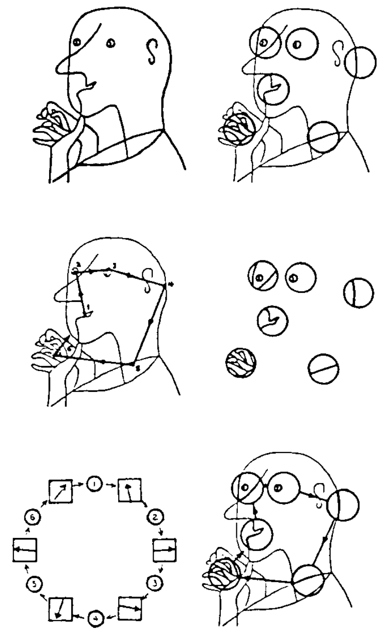

FIGURE 1: SCANPATH MECHANISM FOR TOP-DOWN VISION

2. MATCHED FILTERS AS A MODEL FOR VISUAL SEARCH

FIGURE 2: Top-Down Image Processing for Robotic Control

2D projection of robot model with superimposed ROI's to direct image processing (upper left); frame-grabbber view of robot showing 2 on-the-scene visual enhancements to ease thresholoding operation (lower right); several image processing algorithms operating within the ROI's, including sobel operators, thresholds, these combined and centroids with heavy crosses superimposed (upper right); sequencing of successive and repetitive scanning of ROI's, updating model in similar fashion to human scanpath (lower left). (Stark, Mills, Nguyen, Ngo, 1988)

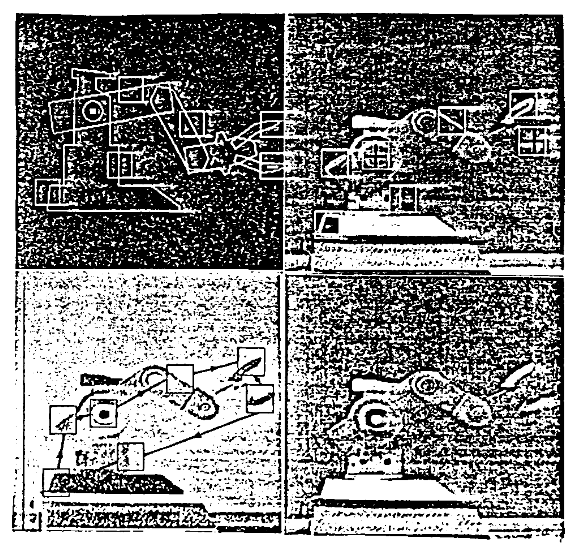

FIGURE 2

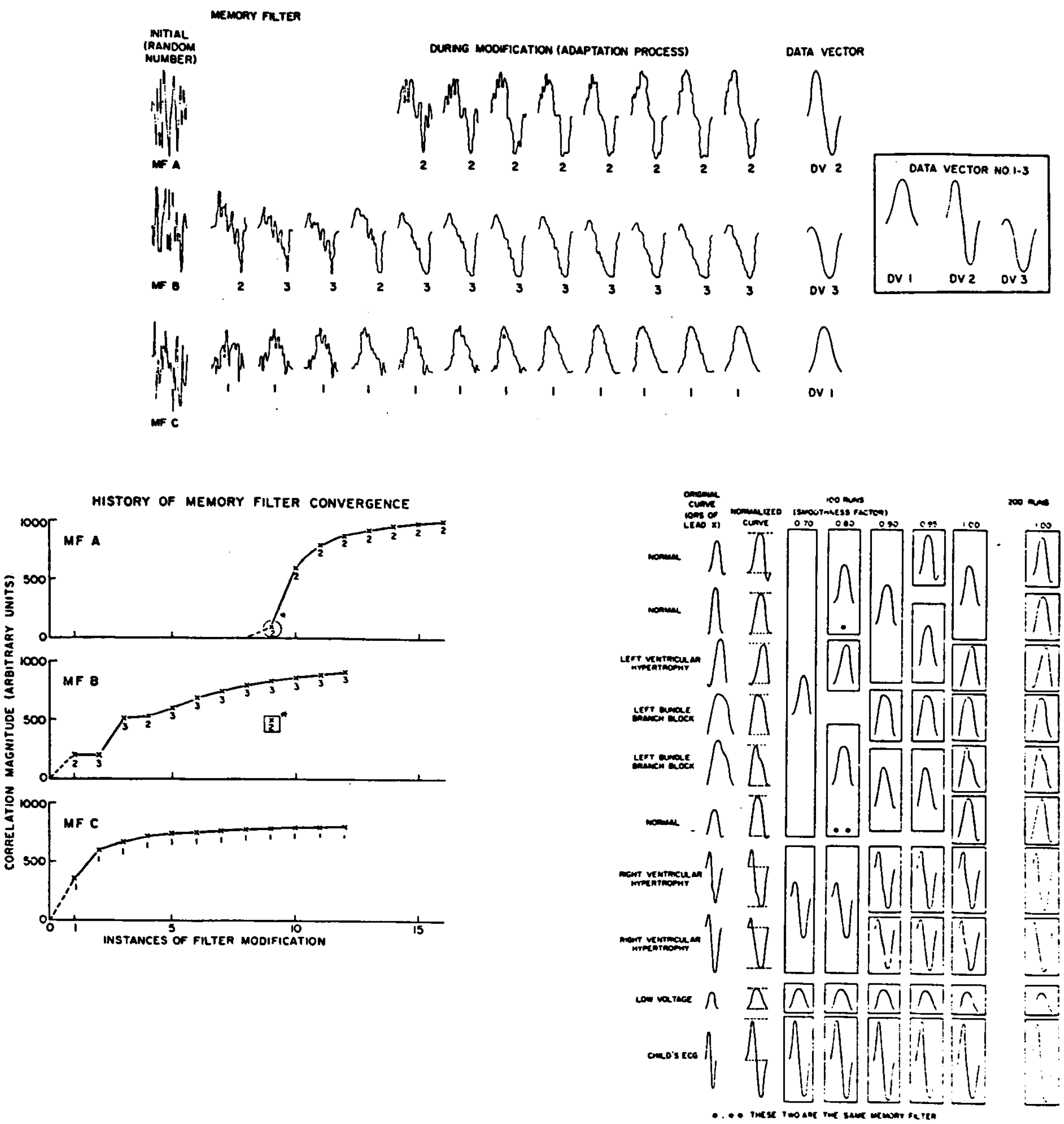

FIGURE 3: Adaptive Matched Filters

Shaping by modification of filters by accepted waveforms with filters closely approaching data vectors (upper panel); threshold for acceptance increases as filter forms so as to protect memory; note exclusion of data vector 2 from MF B and new MF A forming (left lower panel); results after presentation of real ECG data; note here noise is only collection of other data vectors rather than gaussian white noise above (lower right); as threshold parameters change so also does final classification in this adaptive phase and consequently identification or diagnosis in non-adaptive recognition phase. (Stark, Okajima, Whipple, 1962)

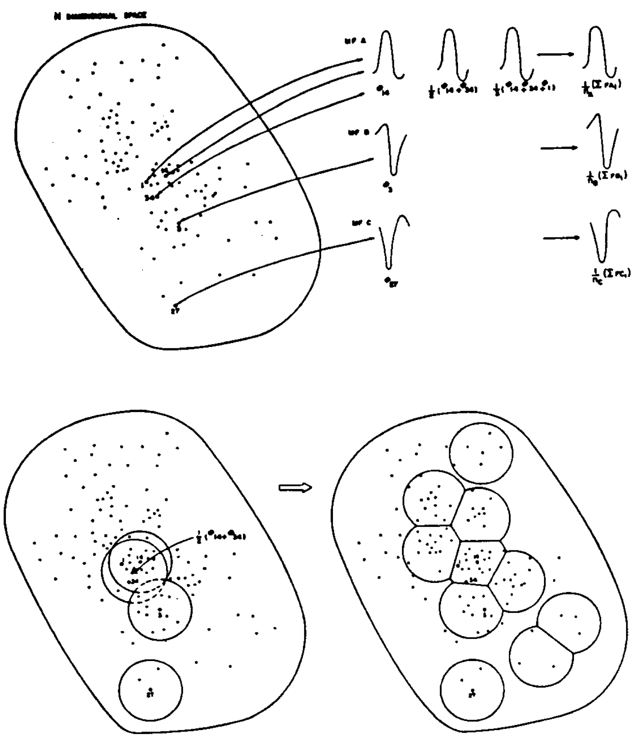

FIGURE 4: Learning Without a Teacher

Adaptive matched filters forming in feature space, a 31 dimensional space (upper panel); after full division of space adaptive phase switches to fixed recognition phase (lower panel); compare important parameters representing thresholds, convergence limits, etc. in this space vis-a-vis descriptions of same adaptive processes depicted in Figure 3. (Okajima, Stark, Whipple, and Yasui, 1963)

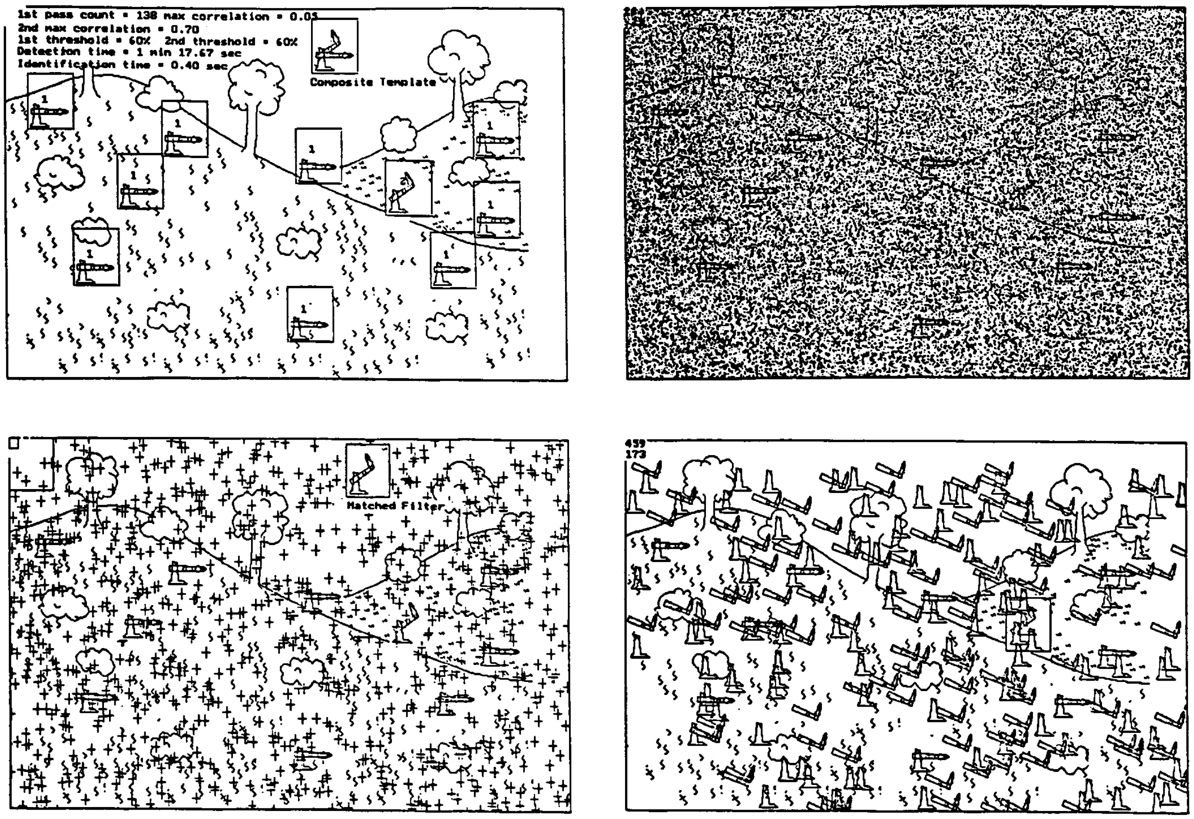

FIGURE 5: 2D Matched Filter Search.

Composite template first recognizes all instances, and then later individual templates-matched filters identifies particular postures of toy robots; additive salt-and-pepper noise makes correlation task more difficult (upper right); cross-noise (lower left) places an even greater burden on cross-correlation matched filter discrimination process; indeed we have developed a pre-filter to help visual search in spite of this obstacle; clutter-noise, defined by close resemblance to target, is of course the severest burden (lower right), so that simple signal processing filters cannot aid in the discrimination task; however matched filter scheme manages to overcome difficulty. (Huy and Stark, 1990, in preparation)

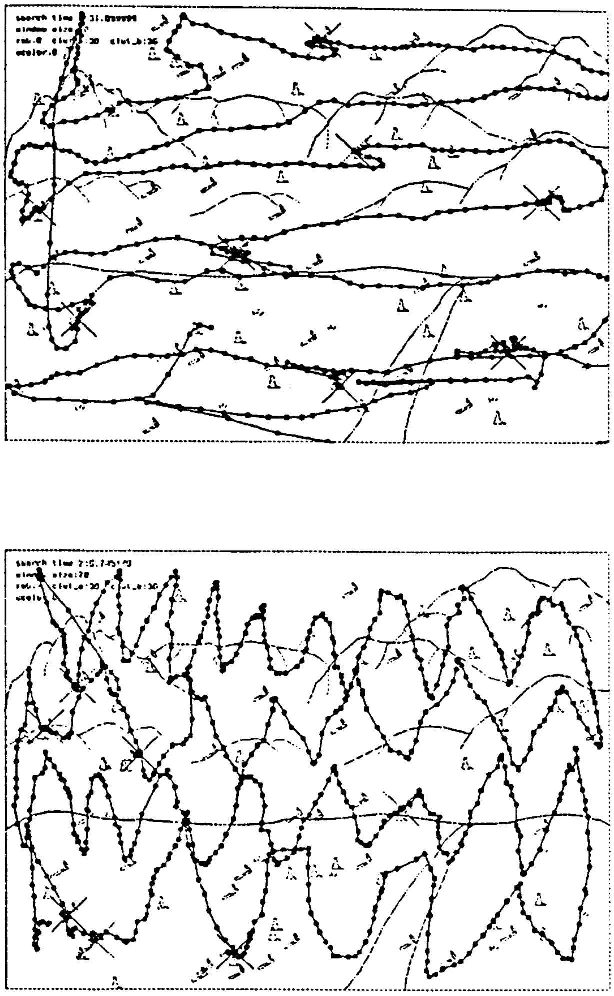

FIGURE 6: Search patterns found in experimental human search of 2D non-naturalistic scenes.

Note that unguided patterns in naive observers followed rather regular patterns often reminiscent of reading em patterns. Social evolution of printed page likely followed natural em patterns. Also resemblance to formal search theoretic patterns developed in beginning of operations research in airplane hunts for submarines. P[d[was generally fairly high. See also Kraiss with suggestion from modeling of experimental studies that if P[d] was low then random search performed as well as patterned search. (Yamishita, Tharp and Stark, 1990, in preparation)

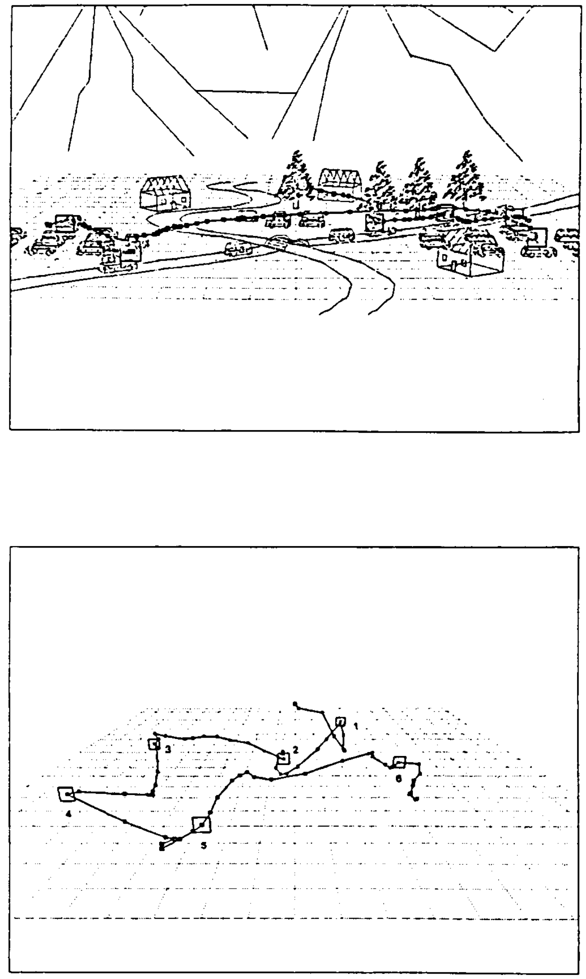

FIGURE 7: Search paths found in experimental human search of 3D naturalistic scenes.

Note the scanpath -like quality of the path and also the repetitive nature. Operational definition of naturalistic is suggested by the evidence for familiarization and memory effects with repetitious presentation of the same scenes and of the same locations of the targets, respectively.

3. ENHANCEMENTS FOR SPATIAL DISPLAYS AND PERFORMANCE IN VIRTUAL ENVIRONMENTS

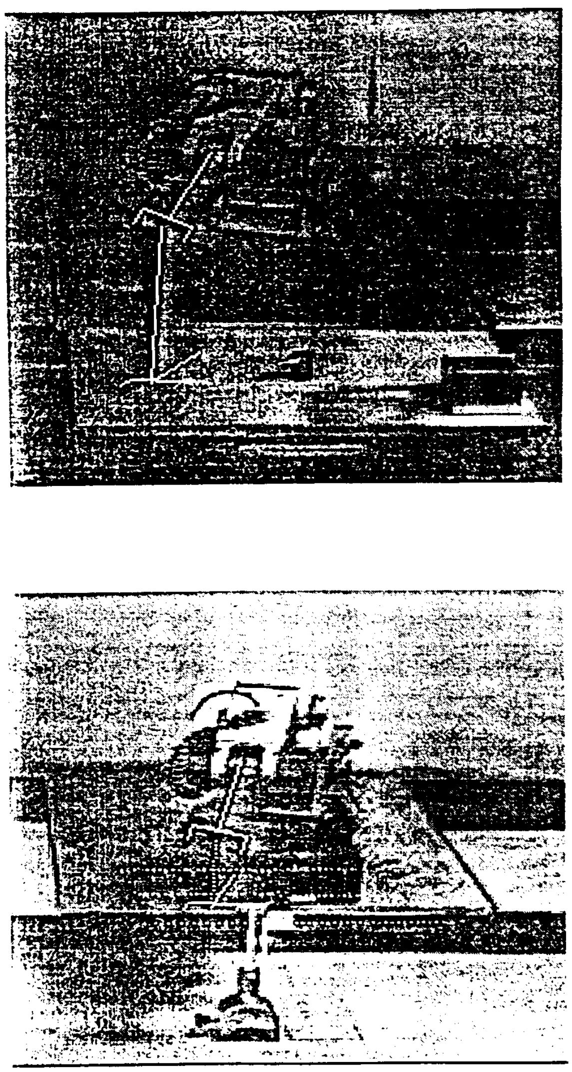

FIGURE 8: On-the-Screen Visual Enhancements.

In our robotic control scheme we have a model of the robot and its commanded position and attitude; thus it is possible to superimpose a computer graphics outline of an important part of the robot, forearm and gripper (upper panel) together with its projection on the horizontal floor; the reference line connecting these has been demonstrated by us to be a most useful visual enhancement in a number of performance studies. The horizontal projection can be elevated off the floor to the level of the work piece (lower panel) and a plan grid added as an additional visual enhancement. Indeed these computer graphics enhancements not only importantly supplement the video picture, but are actually sufficient and efficient without the video picture for adequate control; consider the extreme reduction in communication bandwidth when model parameters are transmitted rather than video images! (Kim and Stark, 1989)

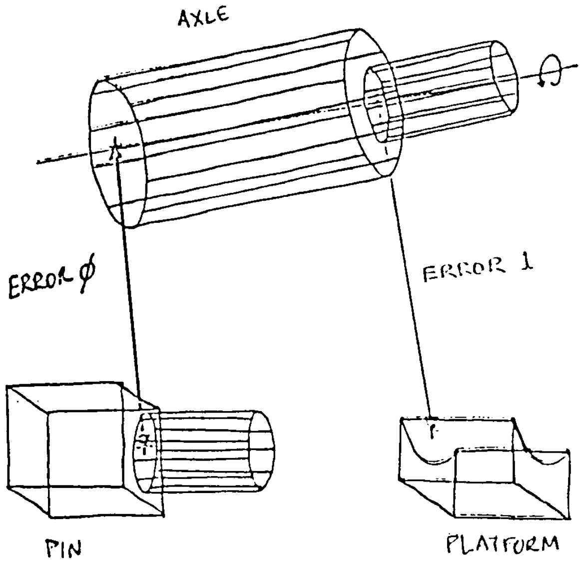

FIGURE 9: 3D Positioning Task

Cylindrical object (upper) has to be oriented and inserted over large pin (lower left) and then also fitted onto curved platform (lower right) by a series of careful repositioning and reorientation maneuvers; necessary 3D depth information can be provided by a true stereo image generated by computer graphics and displayed to right and left eyes on human operator; it may also be possible to provide depth illusions by rotation of the cylinder or wobbling of the task setting. Stereo has been demonstrated to be a robust and important aid to tele-operations; the rotational motion parallax has not improved performance with or without co-existing stereo. (Hirose, Liu, and Stark, 1990, in preparation)

4. UNDERLYING NEURAL MECHANISMS FOR VISUAL PROCESSING

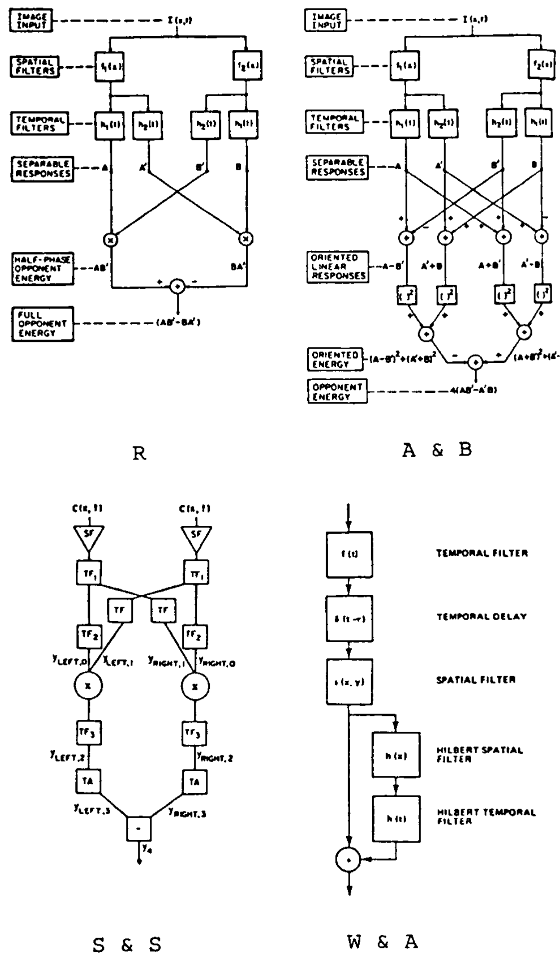

FIGURE 10: Reichardt and elaborated Reichardt models for velocity detection.

R is original Reichardt model (1961), a key feature is the non-linear element; elaborated models, A & B by Adelson and Bergen (1985), S & S by van Santen and Sperling (1985), and W & A by Watson and Ahumada, Jr. (1985), modify by extension for particular purposes various input, output and intermediate elements to satisfy modeling requirements of actual experimental data; the core principles remain the same.

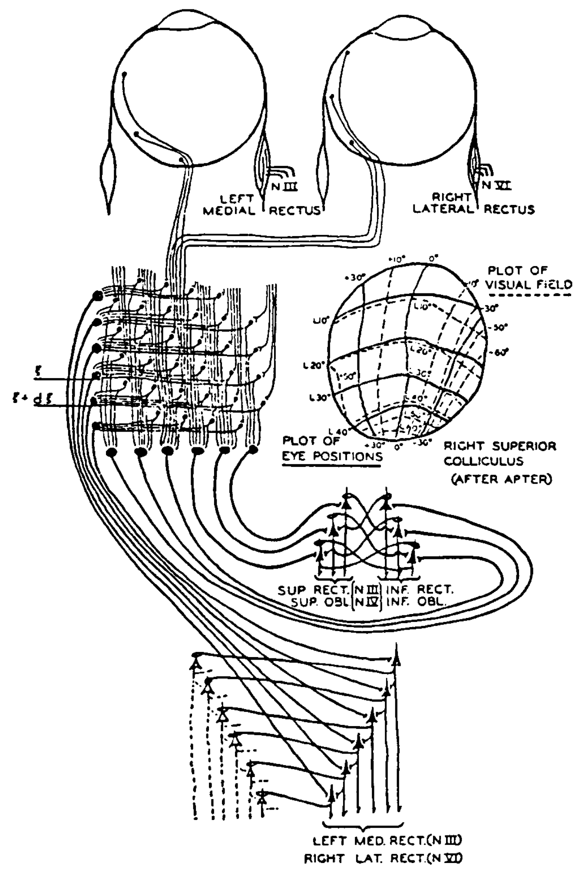

FIGURE 11: Non-Random Neural Model

Note severe constraints placed upon model by knowledge of connectivity, neuro-anatomy, and functional relations of various local and more global elements of model. It may be that constraints in real central nervous systems enable function; absence may make random generation of function unlikely in modern-day connectionist networks. (McCulloch and Pitts, 1947)

PANELISTS* AND THEIR CO-AUTHORS

*Lawrence W. Stark is a Professor of Engineering Science and of Physiological Optics at Berkeley; his interests range from neurological control of movement and human vision to engineering autonomous robots requiring supervisory man-machine monitoring via virtual environments.

*Stephen Ellis both teaches at Berkeley and at NASA-Ames Research Center as a Group Leader in the design of Spatial Displays and Spatial Instruments based upon knowledge of human visual performance.

Michitaka Hirose, a Professor of Mechanical and Information Engineering at the University of Tokyo, has just completed a sabbatical year exploring virtual environments and some real environments at Berkeley.

Won Soo Kim has contributed to our knowledge about appropriate visual display enhancements to make possible human control of robots and is now developing these methods in space robotic applications.

*Andrew Liu is exploring use of stereo and motion parallax clues in quantitative experiments based upon actual performance measures.

*Charles Neveu is interested in application of neuronal models to understand visual mechanism such as blur detection, used to control complex accommodative focusing functions.

Huy X. Ngo has recently concluded studies in 2D matched filter visual search algorithms that enabled definition of such elements and operations as "clutter-noise" and hierarchical search with detection preceding recognition and identification.

*An H. Nguyen has developed an ingenious "model controlled image processing scheme" based upon the metaphor of the top-down human scanpath theory. It has been demonstrated to be efficient in practical control of toy-robots as well as providing efficient image compression for low band-width communication. He has also studied human accommodation from a control theoretics point of view.

Greg Tharp studied human stereo performance and developed a model for directional errors in artificial displays. More recently he has led a group in developing a computer graphics matrix for both autonomous control of robots with ccd camera visual feedback and as well providing for ease of display to supervisory human operators.

Iris Hitomi Yamashita has recently completed an exciting study of human visual search that has provided evidence for search patterns related to reading eye movement patterns, but more free of the structure of the display pattern itself and search patterns related to scanpaths, a top-down repetitive sequence of localized gaze movements to check upon details of guessed scene, figure or object.