WHY VIRTUAL REALITY WORKS!
Top-Down Vision in Humans and Robots

Prof. Lawrence W. Stark

Telerobotic and Neurology Units
School of Optometry, 485 Minor Hall
University of California at Berkeley
Berkeley, California 94720

TEL: 510-642-3621, FAX: 510-642-7196
NET: stark@pupil.berkeley.edu

I. Introduction

Virtual reality, 'VR', is the 'hype' name for virtual environments, but hype and public relations alone cannot drive the field of VR. So the question we ask ourselves is: why do displays work so well? The answer which I will present in this paper is that seeing is an illusion that hides the actual processes of vision. These illusions apply equally well to the worlds of VR as to the so-called 'real' world.

From primitive times humans have accepted a story-telling setting. As society has become more technological we have progressed from oral stories to books and radio; the theatre has been supplemented by movies and TV and now head-mounted displays. But throughout human history the human imagination has been captured by story-telling and the theatre.

Humans are also very adaptable; many careful studies have been carried out on eye-hand adaptation. The wearing of spectacles is another common example. We know from experiments that the human cerebellum plays an important role in this adaptation.

II. Description of Visual Processes that lead to VR

A number of informational and interactive components of the VR setting all work together to lead to the illusion that we are actually experiencing the artificial sensory world as it is being presented to us.

II-1.- IMMERSION

max screen
head mounted display
stereo -- also multisensory surround
interest as with good movie or absorbing novel

II-2.- INTERACTION

ocular manual control and congruency
image tracking of head motion -- rate limitation
computer graphics -- overlay model >> video
preview/prediction

II-3.- INTUITIVE

easy for naive beginners
training -- in a natural way
robust in case of emergencies

end-effector control vs joint-angle-control
glove >> joysticks
comfort vs ocularmotor strain

II-4.- INFORMATION FLOW

visual enhancements -- on-the-scene
on-the-screen
overload -- dual displays -- see-thru
looking without seeing
workplace

II-5.- ILLUSIONS

nature of human perception
normal 'seeing' is an illusion

III. Description of Some Visual Illusions

We have tried to collect anddocument a number of illusions that may play greater or lesser roles in various VR settings. These are listed in outline form below and will serve as a skeleton for our discussion.

III-1. ILLUSION of COMPLETENESS AND CLARITY

top-down cognitve models control
the perceptual process
visual lobes are not perceived
e.g.,- a page of reading material
use Necker cube illusion to demonstrate
rapid generation of models

III-2. ILLUSION of 3D WORLD

in spite of 2D retina reception
due to generation of models in 3D
cognitive and spatial models are 3D
again, use necker cube illusion

III-3.- ILLUSION of CONTINUITY IN TIME

eyelid blinks interrupt vision
saccadic suppression interrupts vision
also absence of grey out secondary to
microfixational eye movements

III-4.- ILLUSION of INSTANTANEOUS ACTION

sampled data delay in motor commands

III-5.- ILLUSION of SPACE CONSTANCY

retinal image motion with saccades

type II -- commanded motion but.....

corollary discharge and efferent copy

to the Helmholtz comparator

saccadic suppression of image displacement

III-6.- ILLUSION of SPACE STABILITY

unconscious of flow fields

unconscious of expansion fields

III-7.- ILLUSION of PRESENCE OR TELEPRESENSE

congruence of head and image motion

egocentric direction set by

gaze movements and space constancy

congruency of visual-motor action

reference frames

III-8.- ILLUSION of SEEING WHEN ONLY LOOKING

why HUD may be dangerous

"looking without seeing"

" one doesn't see and ALSO doesn't know one is not seeing!"

(consider --- seeing without looking)

IV. Looking Without Seeing

Some years ago, when HUD (head-up display) was being suggested for automobiles, a letter I wrote was published in the Forum section of the IEEE "Spectrum" (p. 8, April, 1989) suggesting that this might not be a very safe development. The text of the short letter is given below:

"Looking without seeing"

I was very interested to read Ronald K. Jurgen's "New Frontiers for Detroit's Big Three" [October 1988, p. 32]. In my opinion the head-up display (HUD) offered by General Motors Corp. is a very dangerous developent since people do not always see things, even when they are looking right at them.

Even more dangerous, with the HUD, people might not even be aware that they are not seeing. Thus a driver could be distracted from seeing a part of an obstacle on the road by virtue of the instrumentation display on the windshield, even though his gaze is straight forward.

I believe the HUD so common in military aircraft may in fact be a contributing cause to the large number of accidents that have occurred in the last few years. Of course, the military's high normal accident rate makes it difficult to be certain that the HUD is the major distracting factor.

Lawrence W. Stark, Berkeley, Calif.

V. The Scanpath Theory for Active Vision

The scanpath theory, put forward in 1971, suggests that a top-down internal cognitve model of what we "see" controls not only our vision, but also drives the sequences of rapid eye movements and fixations, or glances, that so efficiently travel over a scene or picture of interest.

The cognitive model of what we expect to see is what we actually 'see'. This internal model drives our eye movements in a "scanpath", a repetitive, sequential set of saccades and fixations over subfeatures of the picture or scene, so as to check out and confirm the mode (Figure 1). These scanpath sequences are idiosyncratic to the subject and to the picture.

VI. Further Experimental Evidence Supporting the Scanpath Theory

On the basis of this suggestive evidence, the scanpath theory was put forward asserting that an internal cognitive model controlled active looking, scanpath eye movements and the perceptual process (Figure 2).

At that time, most visual neurophysiologists, psychologists, and computer vision scientists believed that events in the external world controlled eye movement. Our internal cognitive model must fairly acccurately represent the external world scene or our species would have gone bottom-up like the dinosaurs. How, then, can we prove our assertion of the scanpath theory? Experimental answers came from studies of scanpaths of human subjects viewing fragmented figures, ambiguous figures (Figure 3), Necker cubes, from studies of eye movements during visual imagery and during visual search.

Mathematical methods to quantitate the evidence for experimental scanpaths developed along two lines: (i.) Markov Matrices of transition coefficients (Figure 4) documented that scanpaths were neither deterministic nor random, but were considerably constrained by probabilistic coefficients; and (ii.) String Editing Distances that appeared to be better suited to measuring the constraints over an entire scanpath string (Figure 5).

Experiments have shown that when we look at ambiguous pictures (Figure 3), patterns of eye movement change with the mental image we have of the ambiguous figure. When we engage in visual imagery, looking at a blank screen and visualizing a previously seen figure, our scanpath eye movements are similar whether viewing the figure or the blank screen (Figure 6B). This provides strong evidence that the internal cognitive model and not the external world (since this is absent in visual imagery) drives the scanpath. Recent evidence uses string editing distances to quantitate the similarity and dissimilarity between scanpaths. Also, studies of visual search indicate that a primitive form of pre-cognitive spatial model controls a 'searchpath' sequence of eye movements (Figure 6B).

VII. Robotic Vision

Buttressed by these new views of top-down human vision, we have applied the scanpath theory to robotic

vision. Here we use our knowledge of the spatial layout of the robotic working environment, including position and orientation of the video cameras, and the nature of the robots and the work-pieces to develop a computational "cognitive model" (Figures 7 and 8) .

This computer model then controls the image processing. Regions of interest, ROIs, are generated so that image processing, such as local thresholding and centroid calculations, can be carried out efficiently and robustly (Figure 9). Only those subfeatures essential for identification and control are processed, reducing the computational task greatly. The model not only controls image processing, as in human vision in the scanpath mode, but can also control the robots, the cameras, and displays for the supervisory human teleoperators. The model also serves to reduce communication bandwidth requirements since only commanded and correction model parameters are transmitted. Thus a top-down visual scheme satisfies a visual feedback control system for robots.

Of course, this application to robotic vision and autonomous control of robotic motion does not prove that the scanpath theory is operative in man; the experiments reported earlier, hopefully, provided that evidence. The use of the scanpath sceme in robotic vision does explicate and make concrete the workings of this theory of active vision and its efficacy.

VIII. Overview of Human Vision

VIII-1. Perception and Philosophy and Eye Movements

Perceptual processes may be divided into four stages, as by Kant (Figure 10). The chaotic world of Appearances, "stuff", has energies that impinge on our organs of Sensation, "bottom-up physiology without space and time." Representation, called ideals by Plato and "top-down cognitive models" by us, activates Perception per se to actively seek for, and interpret, confirmation or denial of hypothesized models. Here we consider the "active looking scanpath as the operational phase of perception per se."

An open question often raised at my lectures is "how are cognitive models formed"? Again, naive realists, bottom-up as always, believe without much reflection that external experience flows inward to form a cognitive model. The centroid in some feature space of all the tables I have seen provides me with the type notion of a table. Alternatively, Plato felt that each infant was born with a complete set of ideals that needed to be "awakened" by some argument or experience. Our concept rests upon the inate ability of the brain to synthesize models by means of analogic reasoning or propositional construction. These models can be brought up for consideration rapidly, and can be then discarded, modified, etc.to match the perceptual situation.

The scanpath theory, put forward in 1971, suggests that a top-down internal cognitve model of what we "see" controls not only our vision, but also drives the repetitive sequences of rapid eye movements and fixations, or glances, that so efficiently travel over a scene or picture of interest. Only 10% of the duration of a sequence of views of a target are taken up by the durations of the saccadic eye movements; the intervening fixations or foveations have 90% of the total viewing period.
Philosophers have speculated that we "see in our mind's eye", but until the scanpath experiments, little evidence supported this conjecture. Eye movements are an essential part of vision because of the dual nature of the visual system -- i) the fovea, a narrow field, about one-half to one degree, of high resolution vision; and ii) the periphery, a very wide field, about 180 degrees, of low resolution vision, sensitive to motion and flicker. Eye movements must carry the fovea to each part of a scene or picture or page of reading matter to be processed

with high resolution. An illusion of clarity exists, that we 'see' the entire visual field with high resolution, but this cannot be true.

VIII-2. Vision

Understanding that visual processes can be bottom-up or top-down helps us to put in order our notions about vision. Keep in mind one of the most interesting questions for future neurophysiology (especially aided by active imaging experiments, functional MRI and PET): Where does bottom-up vision meet top-down vision?

Bottom-Up processes: Lower level vision is sometimes the name given to foveal vision wherein high resolution acquisition of information, as in reading a word, can occur. Recall, the fovea is only one-half to one degree in diameter. Wide angle peripheral vision, middle level vision, of the human retina, although low resolution, is ideally adapted for motion perception, flow field analysis, and pre-attentive "pop-up" parallel sensing.

Top-Down processes: Higher level vision includes perception occurring in the mind's eye. The cognitive model of a scene or a picture is the philosopher's "representation". Its operational phase is the active looking scanpath. Eye movements are also driven in a top-down fashion so that critical regions-of-interest (ROI's) determined from the cognitive model can be sampled with high resolution foveal vision.

VIII-3. VR Applications

We now return to the developing world of VR. How varied and exciting are the applications for VR! Entertainment has quickly moved into this area, moving from TV to head-mounted displays, video games and theme parks. Immersion in a wide, colorful, interesting visual scene and interactions with head movement and manual and locomotory controls of position and viewpoints enable the human subjects to participate in the multi-sensory surrounding environment. Although I have focused on vision, sound, smell and other senses can strongly reinforce the visual illusions.

Technical expertise from and to the field of simulation has played a pioneering role in VR. Flight simulators are, of course, the most developed, but there are also auto, ship, and train simulators. Telerobotics, or the control of distant robots and vehicles, is likely an important area for enabling people to work through, and in, a VR system. Indeed, this was our laboratory's entree into VR. See-through head-mounted displays, HMD's, permit not only a VR environment, but also an extra outside view onto a real scene or another display device. Of course with this "augmented" reality, there is the problem of sharing human attention between two tasks, of "looking without seeing."

The aim of this paper was to discuss VR in the light of what we know about human vision; I hope we have succeeded.

Acknowledgement: We are happy to acknowledge partial support from TRADOC, USA, White Sands, New Mexico (Drs. David Dixon and Fernando Payan, Technical Monitors) and from NASA-Ames Research Center (Drs. Stephen Ellis and Robert Welch, Technical Monitors). Also for enthusiastic discussions over the past years with our many colleagues and students; and, especially, to Professor Michitaka Hirose, Tokyo University.

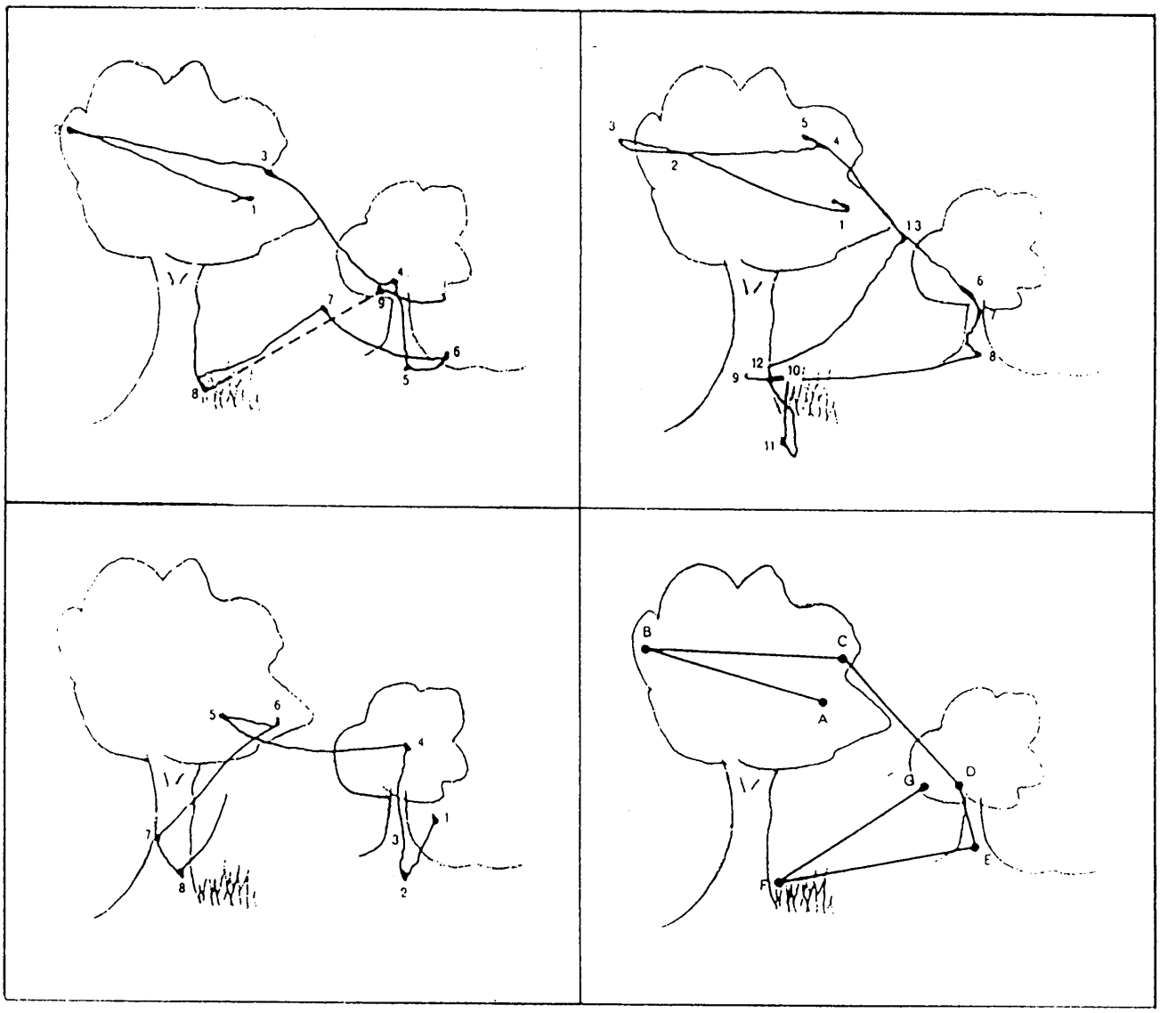
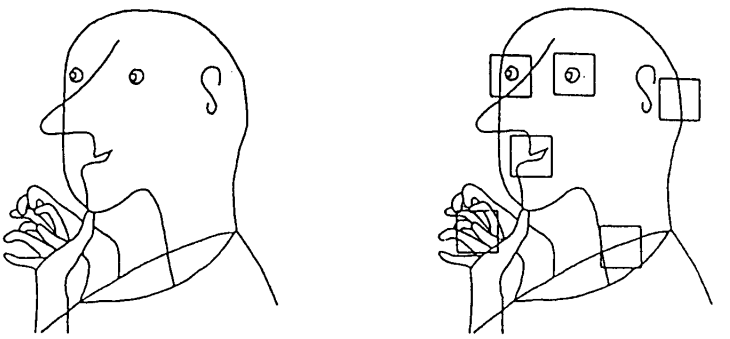
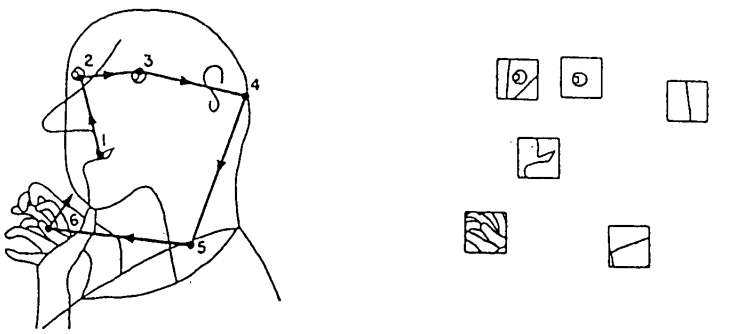


Figure 1 Experimental Scanpath Examples

The scanpath consists of the repetitive sequences of eye movements and fixations while a subject viewed the outline drawings of two trees (upper left and upper right); these are idealized in the diagram (lower right). Note also a quite different eye movement pattern (lower left). (From Noton and Stark, 1971)





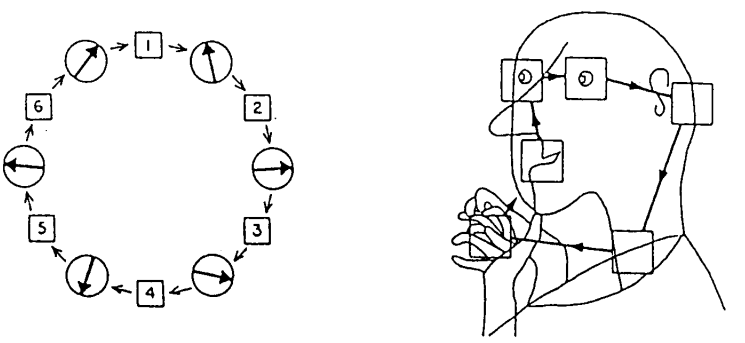


Figure 2 Scanpath Theory
Human visual perception is largely a top-down process with a cognitive model actively driving foveated vision in a repetitive 'scanpath' over subfeatures of the scene or picture of interest to check on and modify or change the working hypothesis. (From Stark and Krischer, 1989 [324])

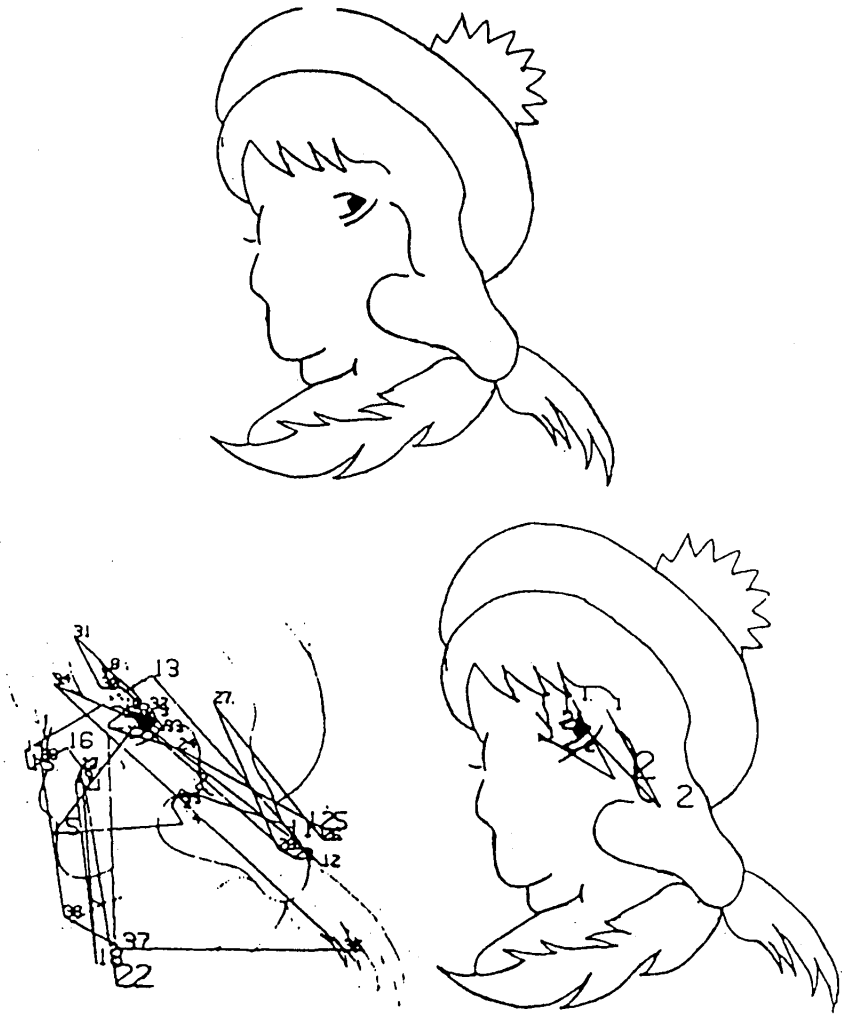


Figure 3 Triply Ambiguous Figure
Upper picture shows old man with moustache, old woman with gnarled chin and nose, and young woman seen in profile with eyelash extending from silhouette (Fisher, 1972). Lower left figure shows eye movements and fixations during experimental run. Lower right figure shows eye movements during four successive occurrences with subject in that cognitive state wherein he saw old man. (From Stark and Ellis, 1981 [211])

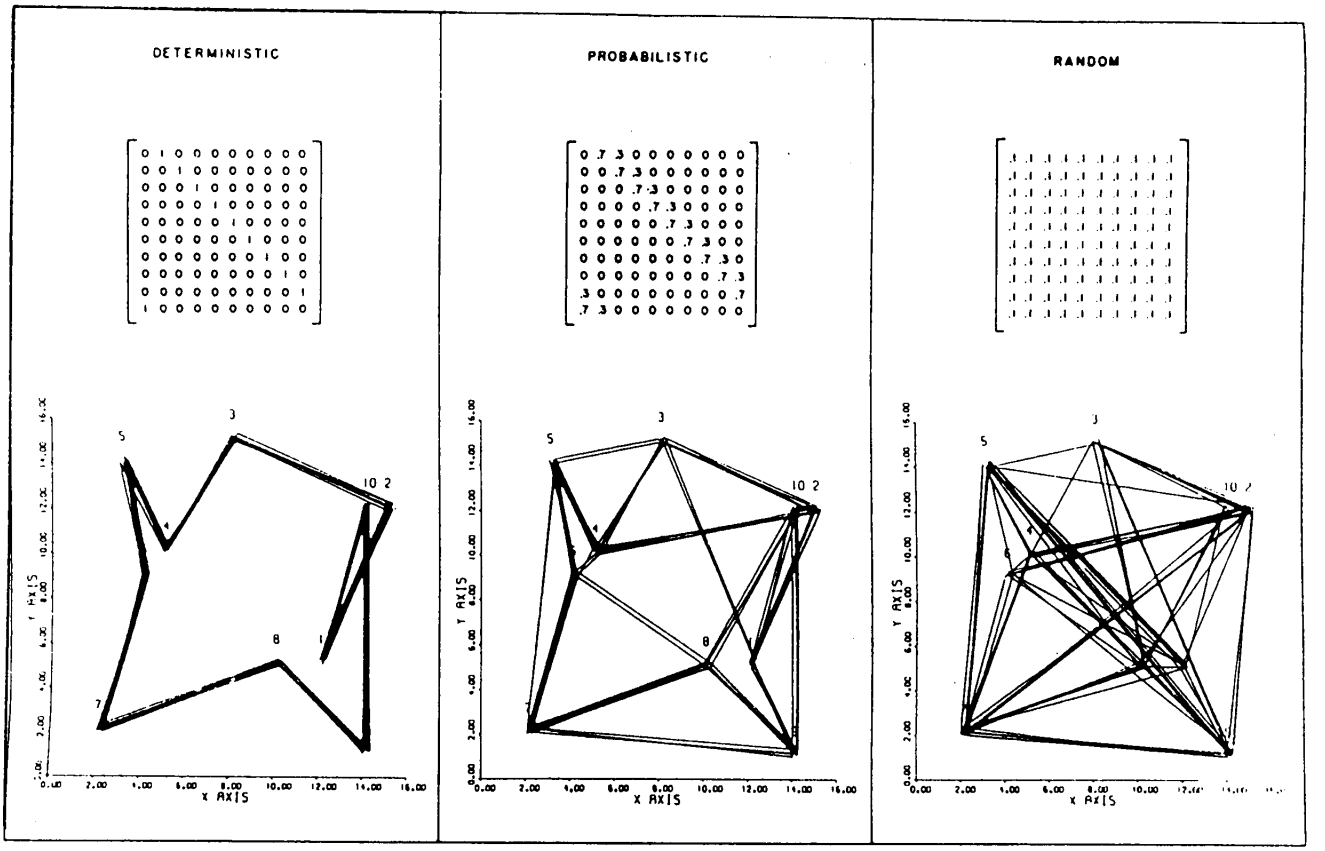


Figure 4 Markov Model for Generating Scanpath

It is possible to simulate scanpaths from markov matrix coefficients. Left matrix determines that in state n, simulated eye moves with probability 1 to state n+1 yielding deterministic simulated scanpath below; some fuzzyness in fixation within fovea is introduced and prevents line superimposition. Middle probabilistic matrix provides for transition probabilities as indicated and produces scanpath below showing some order and some randomness. Random matrix on right allows equi-probability of transitions from any state to any other state and results in completely disordered eye movement sequential pattern; note the coefficients. (From Stark and Ellis, 1981 [211])

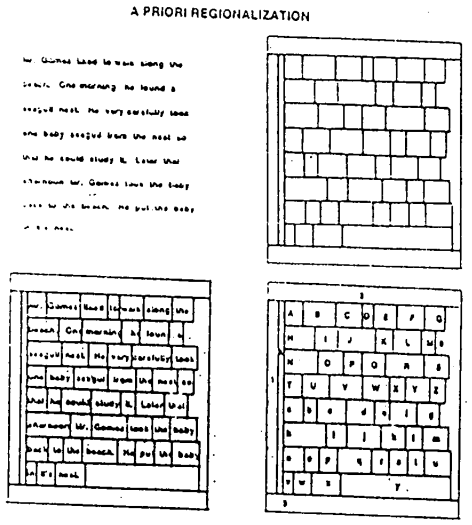


Figure 5a Regionalization for String-editing

For study of eye movements while text is read it is possible to carry out "priori" regionalization. Text (upper left), regions (upper right), text superimposed on regions (lower left), alphabetic code labelling regions (lower right). (From Choi and Stark, in preparation, 1992)

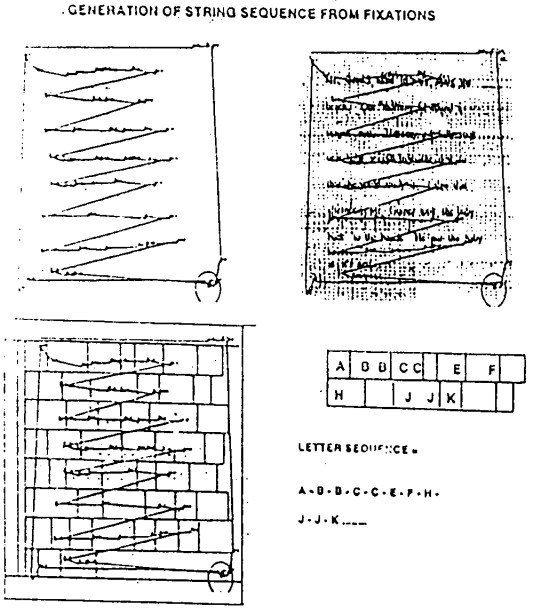


Figure 5b Generation of String Sequence from Fixations

Eye movement superimposed on text (upper right). Eye movement fixation labeled in sequential order and by duration (upper left); note peripheral calibrating saccades. Eye movements and fixations superimposed upon "a priori" regions (lower left). Example of fixations labeled as per appropriate regions (lower right). Finally (also lower right) a letter sequence established as A-B-B..... (From Choi and Stark, in preparation, 1992)

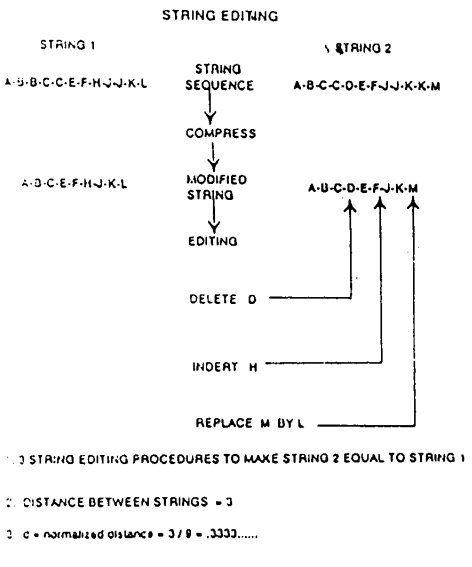


Figure 5c Algorithm for String-Editing Distances

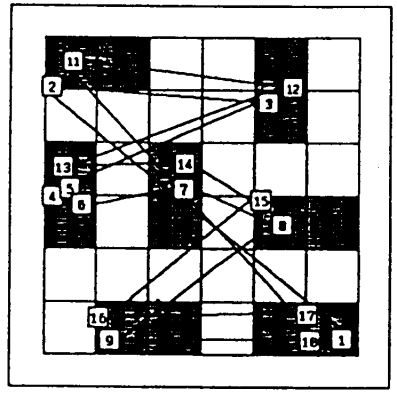
Example of two strings produced in this case by one subject reading the same text on two occasions. The string sequence is compressed to remove refixations within a single region leaving modified string. Editing includes deletions, insertions and replacements; all three processes are illustrated. When string-editing has occurred string two equals string one. Thus the distance equals three and normalized distance equals three divided by string length. (From Choi and Stark, in preparation, 1992)

Figure 6a Visual Imagery Experiments
Showing similarity of scanpaths for looking (upper left) and for imagery (upper right). Two repetitive scanpaths occurred during the 10 seconds for looking and also during the 10 seconds for visual imagery for one pattern, #2, of four patterns of quasi-random simplified checkerboards. In this experiment two trials were also carried out approximately one week apart on the same subject. Scanpaths that 'should' be similar (between looking and imagery of the same target or between looking and imagery of the same target one week apart) are underlined so that the string editing distances of these 24 related scanpaths can be seen to be much less (0.25 + / – 0.15; n=24) than distances between unrelated views (0.78 +/ – 0.08; n=96) , (right middle). (From Brandt and Stark, in preparation, 1994)

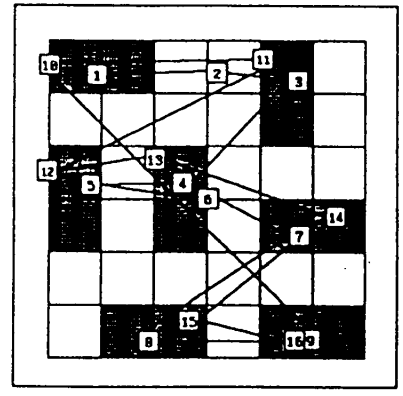
Figure 6b Searchpaths
The eye movement pattern developed while the subject was repetitively tasked to find about 7 +/– 2 target trucks intermixed with vans and cars serving as decoys in this stereographic scene. The repetitive sequences of eye movements, here called searchpaths controlled by a spatial model may be a primitive form of scanpath, controlled by a cognitive model. (From Choi and Stark, in preparation, 1994)

BRANDT EXPERIMENT

Looking 2



Imagery 2



Subject JE : All Matches in Two Trials

I.	Look 1	Look 2	Look 3	Look 4	Img 1	Img 2	Img 3	Img 4
Look 1	0	-	-	-	-	-	-	-
Look 2	0.67	0	-	-	-	-	-	-
Look 3	0.88	0.73	0	-	-	-	-	-
Look 4	0.63	0.80	0.56	0	-	-	-	-
Img 1	0.28	0.73	0.71	0.90	0	-	-	-
Img 2	0.71	0.13	0.78	0.72	0.86	0	-	-
Img 3	0.71	0.73	0.28	0.63	0.86	0.87	0	-
Img 4	0.71	0.80	0.78	0.36	0.86	0.82	0.91	0
II.								
Look 1	0.00	0.73	0.78	0.63	0.29	0.73	0.73	0.91
Look 2	0.75	0.13	0.65	0.82	0.86	0.18	0.87	0.91
Look 3	0.81	0.80	0.33	0.70	0.86	0.87	0.33	0.73
Look 4	0.63	0.73	0.72	0.42	0.86	0.71	0.73	0.45
Img 1	0.07	0.71	0.79	0.54	0.00	0.77	0.85	0.73
Img 2	0.80	0.27	0.80	0.73	1.00	0.27	0.79	0.91
Img 3	0.85	0.77	0.46	0.85	0.86	0.77	0.31	0.91
Img 4	0.75	0.80	0.72	0.53	0.86	0.87	0.73	0.55

	M	sd	n	ml
D	0.25	0.15	24	
O	0.78	0.08	96	

II.	Look 1	Look 2	Look 3	Look 4	Img 1	Img 2	Img 3	Img 4
Look 1	0.00	-	-	-	-	-	-	-
Look 2	0.80	0.00	-	-	-	-	-	-
Look 3	0.80	0.76	0.00	-	-	-	-	-
Look 4	0.67	0.76	0.79	0.00	-	-	-	-
Img 1	0.15	0.84	0.84	0.61	0.00	-	-	-
Img 2	0.80	0.15	0.86	0.73	0.85	0.00	-	-
Img 3	0.84	0.76	0.30	0.76	0.77	0.69	0.00	-
Img 4	0.73	0.86	0.69	0.30	0.85	0.80	0.77	0.00

Figure 6a Visual Imagery Experiments

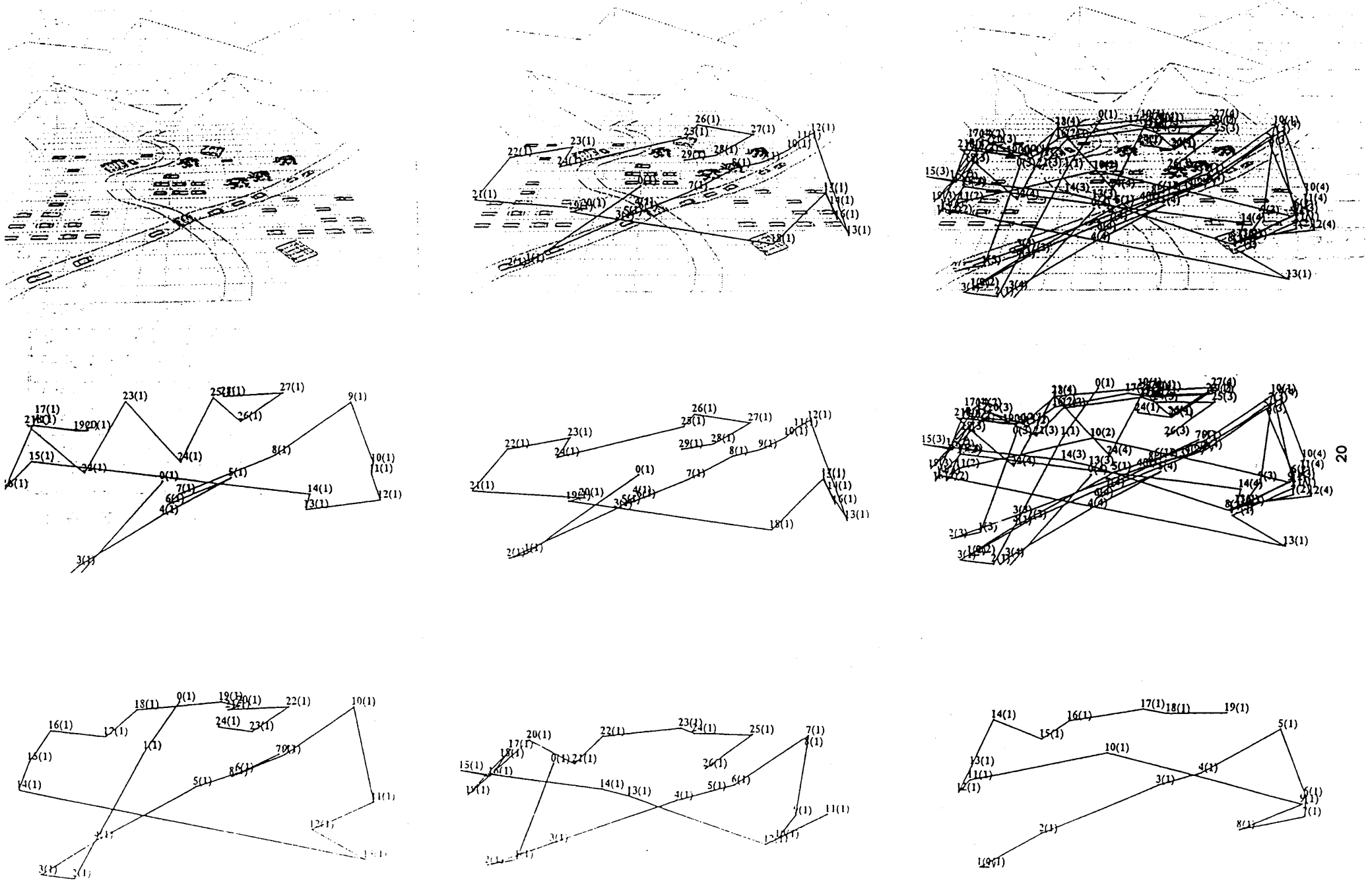


Figure 6b Searchpaths

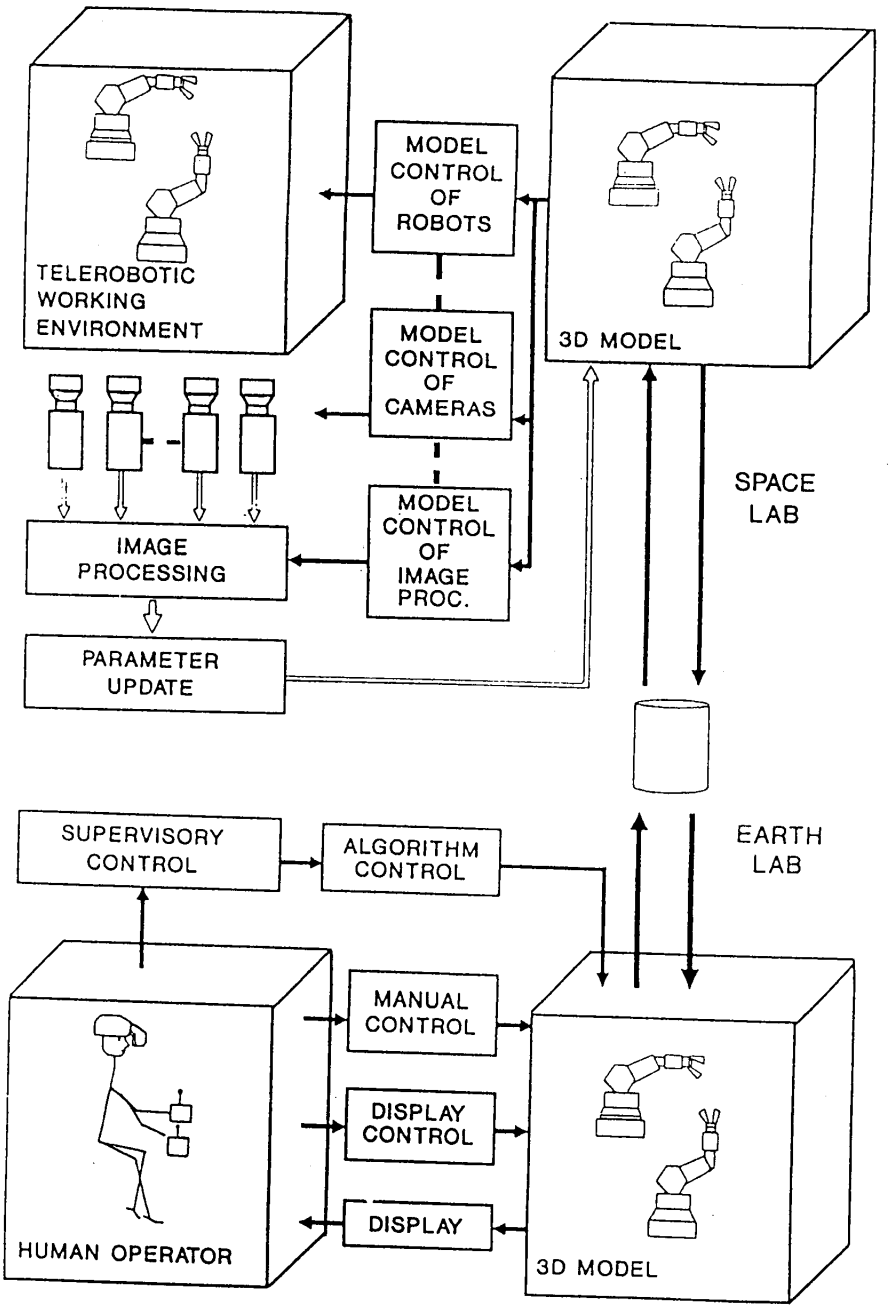


Figure 7 Cooperative Control in Telerobotics

Overview of scheme for model control of image processing. 3D Model (upper right) in space controls the TRWE, the telerobotic working envionment (upper left), including robots, cameras, and image processing. Defining numerical parameters, abstracted by image processing algorithms working only within ROI's, regions-of-interest, are able to correct and update the 3D model. Communication channel can be narrow-bandwidth since only sensed and control parameters are exchanged between the space and the earthlab 3D models (right). The human operator, H.O. (lower left), can view a display, partly controlled by him and partly by the lower 3D model. He can also manually control the robots and cameras by controlling the 3D model with immediate (or delayed as would be the actual case) feedback, or he may be in a supervisory command mode --- approving of suggested paths or task-segments or setting into motion emergency interruptions and reinitializing and recalibrating. procedures. Here, we have redundant control pathways and modes of operation. (From Stark et al., 1987 [296])

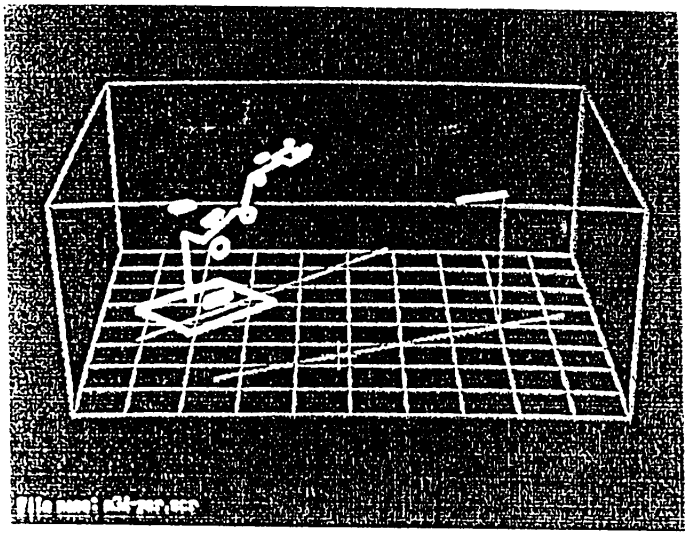


Figure 8a Forward Model for Robotic Control

Kinematic structures of Armatron indicated together with open circles and open oblongs indicating commanded and thus expected position of on-the-scene-enhancements. Also shown is 3D schematic robot working environment with grid floor. Note path planning lines indicated together with a critical point marked by a cross; also workpiece to be grasped with a reference line, an on-the-screen-enhancement to grid floor; height is thus made easily perceivable for human supervisory control. (From Nguyen and Stark, in preparation, 1992)

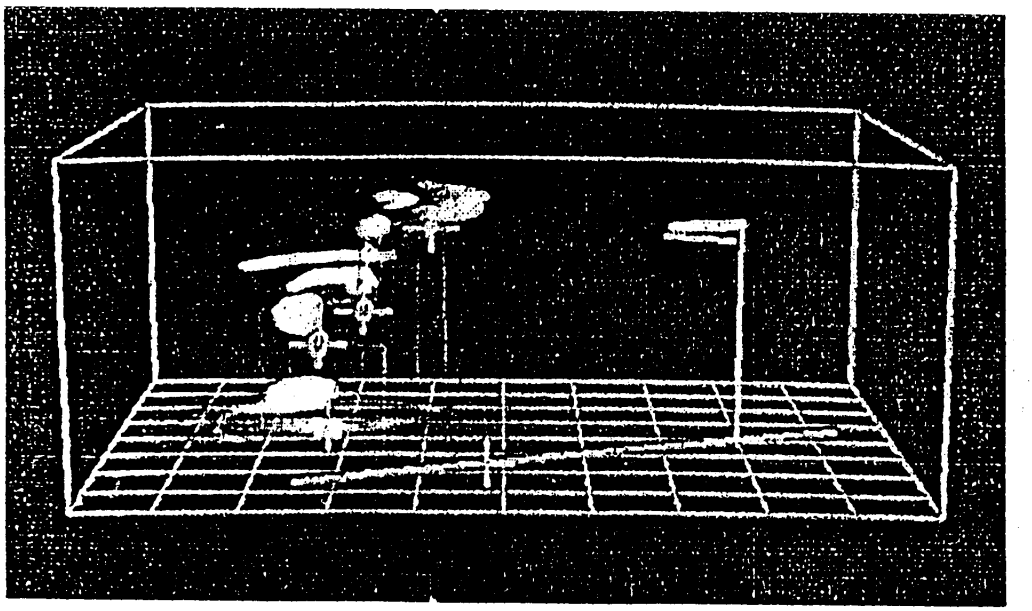


Figure 8b Feedback Model Monitoring Dynamic Performance

Computer model of Armatron robot rotating under autonomous path-planning control; about 20 successive frames indicate dynamic rotation. Plus markings indicate results of centroid calculation; thus this is a feedback model of robot. (From Nguyen and Stark, in preparation, 1992)

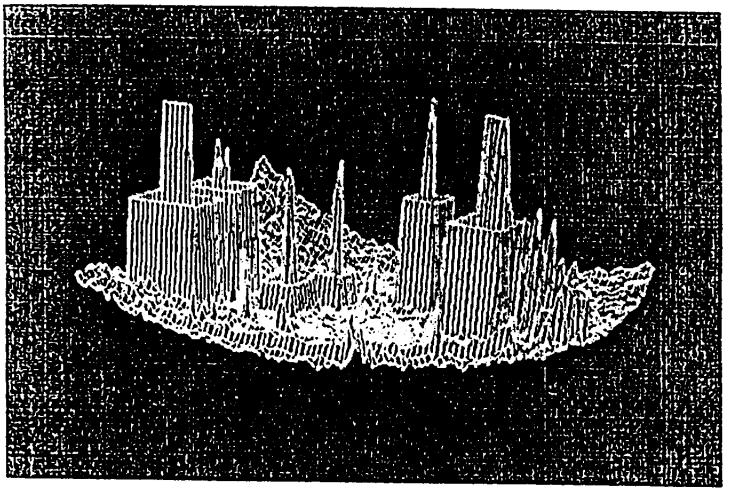


Figure 9a Bottom-Up Threshold Algorithm
Pixel intensity plot shows pixel intensity as a function of x and y coordinates of video picture with intensity in the third vertical axis. ROI outlines are picked up as a vertical open top boxes with on-the-scene enhancement as peaks within the ROIs. ROI heights are proportional to adaptive threshold, with each local region having a different threshold level; this advantage of top-down control of image processing makes for a very robust and autonomous scheme. (From Nguyen and Stark, in preparation, 1992)

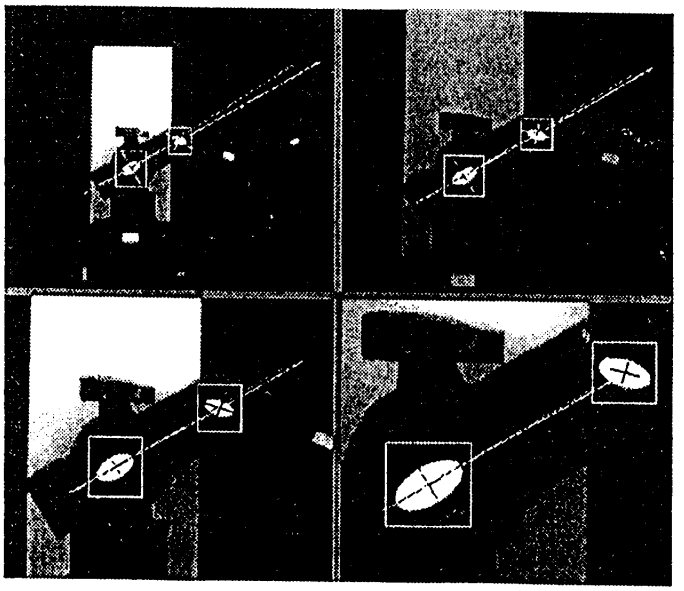
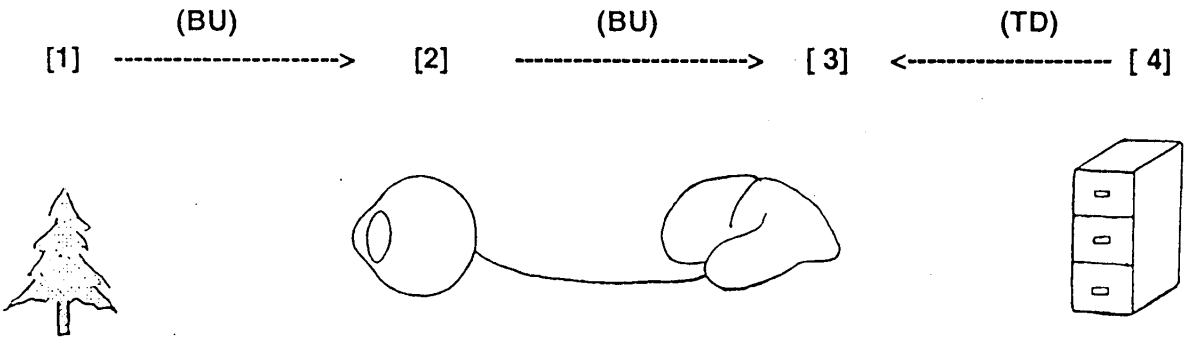


Figure 9b Centroid Calculation
Video images of robot have ROIs superimposed. In these examples the ROIs controlled by the feedforward model are excellent predictors of dots. Centroid calculations are very rapid and enable feedback processing of kinematic parameters of robots, showing accuracy in response to feedforward command. This example studied comparison of two centroids versus one local ellipsoid whose major angle could equally generate kinematic parameters. (From Nguyen and Stark, in preparation, 1992)

PERCEPTUAL PROCESSES



Kantian Definitions

APPEARANCE	SENSATION	PERCEPTION	REPRESENTATION
Erscheinung	*Empfindung*	*Anschauung*	*Vorstellung*

Other Philosophers

phenomenon (Leibnitz)	impression (Mach)	perception (Leibnitz)	noumenon (Leibnitz)
class of appearances (Russell)		intuition (Descartes)	ideal (Plato)
			notion (Berkeley)

Our Terms

"stuff" (not 'things'!)	"bottom up physiology without space and time"	"active looking scanpath as the operational phase of perception per se"	"top down cognitive model"
	doctrine of specific nerve 'endings'	a more planned, forceful, determined activity	

Figure 10 Bottom-up and top-down components of overall perception

The active looking scanpath is the operational phase of "perception per se".