VISUAL SEARCH AND DISPLAY OF METRICS FUNCTIONALITY; EXPERIMENTAL RESULTS

by

Lawrence W. Stark, Claudio M. Privitera, Yeuk Fai Ho
And Michela Azzariti

Memorandum No. UCB/ERL M99/22

22 April 1999

COVER

VISUAL SEARCH AND DISPLAY OF METRICS FUNCTIONALITY; EXPERIMENTAL RESULTS

by

Lawrence W. Stark, Claudio M. Privitera, Yeuk Fai Ho and Michela Azzariti

Memorandum No. UCB/ERL M99/22

22 April 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

VISUAL SEARCH AND DISPLAY OF METRICS FUNCTIONALITY; EXPERIMENTAL RESULTS

Lawrence W. Stark, Claudio M. Privitera, Yeuk Fai Ho, and Michela Azzariti

Neurology and Telerobotics Units
University of California, Berkeley 94720-2020

Abstract

Visual search is a complex set of component processes involving bottom-up and top-down factors in human vision that may adapt to efficiently carry out the search task. An illustrative, quasi-natural search scene was used for an experiment in order to ascertain the effectiveness of a whole set of metrics that could capture these search procedures. We explain the setting and protocol and define and illustrate these and the measures that we have found effective in evaluating and understanding ongoing search. The results generally demonstrate the importance of top-down spatial-cognitive models and procedures related to the scanpath theory of every-day top-down, normal vision. Of special interest are parameter-free metrics that can scale up or down for different search arenas; these include K-means estimation of clusters of targets hits, similarity indices, Sp and Ss, for loci and sequences for both instrumental and target scanpaths. Evidence for semantic, structural and sequential binding have earlier reinforced the applicability of the scanpath top-down theory for approaching the problems of visual search.

Keywords: visual search, cover, detection, scale, similarity indices, instrumental searchpaths, K-means, scanpath, top-down, FOR, FOV, ROIs.

Note that there is a glossary and list of acronyms at end for definition of terms.

1 - INTRODUCTION

Visual search is a complex set of component processes. a).- A pre-scan to estimate parameters of the search, such as dimensions of the search area, levels of noise and clutter, contrast levels for targets, etc; although many of these may be known a priori. b).- A cover process, often by instrumental search, to inspect the field of regard, FOR (Figure 1, lower). c).- Detection of targets, decoys, landmarks. d).- Recognition of detected objects.

Vision is dependent upon such bottom-up parameters as contrast, Cm, spatial frequency, Wx, and such instrumental aids as magnification, and upon such top-down factors as --- trained tactics for search; constructed models as matched filters for targets, decoys, clutter elements; and familiarity with the global and local features of the landscape in that composes the search area.

Cover, or search processes per se, involve first, the movement of the field of view, FOV (Figure 2a, lower), over the FOR, the search area *in toto*. The ratio of the FOV to the FOR provides an indication of the number of FOVs that will be required to traverse or cover the search scene (Figure 2a, inset icon). Often, this is accomplished by means of instrumental search that involves an

FOV window, an optical, or an electro-optical, or other display window, moved over the FOR by a controller with a mechanical device, e.g., a "mouse." A sequence of such "mouse"-controlled FOV shifts (n = circa 80) over the FOR documents how a human carrying out a visual search task in our laboratory carried out an instrumental searchpath and so "covered" the FOR (Figure 1, lower). Cover metrics and the efficiency of cover are important in understanding visual search.

Figure 1: Human viewing of pictures and search scenes

Contrast a human using scanpath eye movements to view a picture (upper), and employing instrumental search (lower) to move a FOV window over a search scene, or FOR.

Figure 2: Search scene and fields of view, FOVs

Forested search scene shown as a high resolution, densely pixelled FOR (Figure 2a), and showing three FOVs containing targets (Figures 2a and 2b).

Of course, the natural human uses head movements to move FOVs into different parts of the potential viewed scene. The FOV is itself traversed by human eye movements, saccades, often, in a scanpath sequence. These saccades actively and rapidly jump from glimpse to glimpse, carrying the high-resolution portion of the eye, the fovea, so as to allow fixation, or foveation, onto important regions of interest, ROIs, that may contain a target. (Figure 1, upper).

The **aim** of this paper is to carry out an experimental visual search task approximating a field test. Metrics have been developed that attempt to capture the efficiency of the visual search processes at a number of different levels; the metrics have been designed either to be scale-independent or to define the scale.

2 - METHODS

Search scene. We describe the search scene as to its general features, its truck-targets, clutter, largely vegetation (Figure 2b) and as well, its dimensions at several levels and with several optical and display units of measurement (Figure 3).

Figure 3: Diagram of the field of regard, FOR

Dimensions of the search area, or FOR, and of the internal and external FOVs. Note icon, lower right, showing actual proportion of FOR and of FOV; these proportions have been distorted throughout or series of figures in an attempt to display the meaningful information to the readers within our constraint format.

In these experiments, the search area, or field of regard, **FOR**, is approximately 1400 meters or 40 degrees in external horizontal visual angle, considering that the nominal range is 2 kilometers. The extreme aspect ratio of the FOR is 10 to 1, approximately 13,500 pixels horizontally and 1350 pixels vertically (or about 340 pixels per degree). An icon, (Figure 3, lower right), has been prepared to show the veridical proportions of the FOR and FOV and placed onto several of the figures illustrating these scenes; the figures of the scenes have been distorted from their true proportions so readers can view the scenes and the included objects and landmarks more clearly.

The viewing port permits an external field of view, **eFOV**, of approximately 140 meters across or 4 degrees. The eFOV has approximately 1250 pixels (3.7 degrees at 340 pixels/degree) horizontally and 580 pixels (1.7 degrees at 340 pixels/degree) vertically. The viewing port, uses magnification of about 11 x to provide a much wider displayed internal field of view, **iFOV**, approximately 40 degrees horizontally (30 pixels per degrees); several examples of the iFOV (Figures 2a, lower, and 2b) provides a clearer idea of what the subjects actually viewed. The instrumental magnification for the field test was approximately x 11; we tried to preserve this magnification in terms of the display on our computer monitor. Of course, this depends on the distance of the subject from the computer monitor. Important features of this overall diagram are the ROIs that can be deduced *a priori* from the target loci or *a posteriori* from the subject's clicks onto either targets or landmarks. This will be discussed extensively in the K-means part of this report below. At present, let us say that the ROIs are approximately 100 pixels across, only one third of a degree in the eFOV, but 3 degrees of visual angle in the iFOV; thus providing extended targets for acquisition by human vision.

The scene is of the California oak forest. Note the very wide aspect ratio of the actual landscape scene (Figure 2a, upper) that we have worked with in our experiments below; recall the icon showing true proportions of the FOR and of the FOV. An example of the eFOVs (upper) is shown as the displayed expanded iFOV (Figure 3, lower); recall, the magnification for the display to the subject on the computer monitor is considerable, about eleven-fold. Note the appearance of a dirt road and typical oak trees and grassland. A truck on the roadway (Figure 2b, upper iFOV) shows a black rectangle that indicated to the subject that he had correctly clicked the mouse on that target. The contrast, even in these poor copies of the actual computer monitor display, is sufficient for the trucks to be reasonably well detected and once detected to be recognized. Truck size as represented in the 2D FOR was a function of distance of the trucks in the 3D search scene.

Target identification using cross-hair cursor and mouse clicks. In order to accomplish target detection, the cross-hairs had to be placed more or less accurately (see below in the K-means test) on the target. Two other trucks (Figure 2b, lower panel) are shown, before the observer had clicked on them, to indicate he had seen them.

Eye movement experiments require careful preparation and execution. During the experiment, subjects often have to remain as still as possible; they are secured on a chin-rest structure, and go through automated calibration procedures before and after each visual stimulus. Even small movements of the head can spoil the data of an entire session. This is one reason why we decided to simplify the acquisition of ROIs by using a self-calibrating "mouse-clicking" over the FOV, rather than measuring eye fixations. We have shown (Stark *et al.*, SPIE 1999) that ROIs chosen by marking these loci of attention with a generic cursor are highly correlated in their structural binding with eye movement fixations.

Computer aspects. In carrying out this experiment, we utilized an Intergraph-PC, to display the large complex scene, full of natural clutter, and with approximately five trucks scattered throughout the scene. These trucks were placed onto the same scene in three different configurations, A, B, and C, for three search scenarios. Software aspects of our programs included the use of the Corel Draw program.to enable placement of the trucks. Matlab programs were used to construct the FOV windows, the cross-hairs, and for recording of both the instrumental search and human visual performance characteristics. Analysis of the data was also done with Matlab programs on Next Unix workstations and on PC-workstations. Details of these analysis programs will be presented below, and also in conjunction with displays of results. A series of early control experiments, using computer graphics and carried out with Open-GL and Visual C++ programs, provided for wide ranging exploration of various parameters in both the experimental protocol and in the analysis approach.

Protocol of experiment. We arranged the protocol of the experiment so that each of our subjects was able to view each of the three scenarios four times. Further, they were exposed to a series of three repetitions looking at the search scene without any targets emplaced; wherein they were asked to indicate important **landmark** features that might help them become familiar with the scene in subsequent searches. This made for a total series of 15 search scene presentations.

Before starting this protocol, the subjects were allowed to familiarize themselves with a different search scene, and with the control of the FOV window, the cross-hair cursor, and the clicking buttons on the mouse. Mouse control also provided for discrete movement of the window, so that they could scan the eFOV over the FOR. As in most mouse controls, the more eccentrically the mouse moved, the faster the scanning movement of the FOV over the FOR. They were also told that they would have 120 seconds for each search. This time constraint enabled us to focus on the extent of the FOR covered, as a measure of the efficiency of the subject's search; some subjects (numbers 3 and 4) finished early (see below Cover Results).

Partly before the practice scene, and also as appropriate during the protocol, a written set of instructions (Figure 4) was presented to the subjects on the computer screen, and was further reinforced by verbal interaction with the experimenter. Instructions specified how to slew the FOV over the FOR and how to center the cross-hair onto the target. The task was defined with respect to time limitations, number and type of targets, landmarks and repetitions. For part of the protocol we asked subjects to identify landmarks that might be useful as reference locations when scanning through the FOR in further repeated searches. We collected searchpaths for these landmark presentations and noted similarities and differences qualitatively and estimated them quantitatively using our similarity metrics.

The subjects, students or researchers at Berkeley were unpaid volunteers; they were told that they could terminate the experiment any time they became uncomfortable in conformance with the rules of the Committee for the Protection of Human Subjects at the University of California.

Figure 4: Instructions to subjects

These instructions were presented in written form on the computer screen, and also reinforced verbally by the experimenter.

Analysis methods. These varied from simple histograms demonstrating numbers of targets clicked and percent of acquisition sequences, to K-means analyses of the distribution of clicking locations with respect to targets and landmarks, and included a variety of statistical methods generally resting upon ANOVAs, analyses of variance. Of interest is the use of the theory of signal detection, TSD, for subject error analysis. Also, very important, are various **distance and similarity metrics**, originally developed for scanpath studies and early visual search experiments; these have been refined over the past years.

Application of metrics. The sequence of target clicks of a subject were easily measured since the cross-hair positions and the target positions were well defined and the sequence was a vector of (x,y) loci. Each clicked locus or region, might include a target or a false alarm, that is a clicked region without a target. These of course equally applied to landmarks. For instrumental searchpaths we used a "binning" measure of distance; there were 30 bins per FOR.

Comparison of final vectors of loci, began with taking two sets of loci (Figure 5, middle column, upper and lower panels) and clustering these two sets using a distance measure derived from a K-means pre-evaluation (see below). This evaluation determined a region for calling coincident any loci that were closer than this distance and non-coincident for loci that were further apart than this distance. The final selection of coincident loci (Figure 5, right panel) then enabled a similarity metric, Sp, to determine how many loci two searches have in common. The final value is normalized based upon sequence length.

The individual sources of the elements, that is the original loci, used in these final interactive steps were preserved as circles and squares (Figure 5, right panel) to illustrate the procedure. Each separated sequence of loci is temporally ordered and thus yields a string (Figure 6). Here, we have for example: string1 = *afbffdcdf* and string2 = *abcfeffgdc*. The string editing similarity index Ss was defined by an optimization algorithm with unit cost assigned to the three different operations deletion, insertion and substitution.

Thus, these comparison metrics yield two different indices of similarity which tells us how closely two sets of loci resemble each other in position, Sp, and in sequence, Ss (see the "toy" diagram on Figure 7, upper panels). For the example illustrated above (Figure 6) we have: Ss = 0.22.

Figure 5: Calculation of Similarity Index, Sp

Figure 6: String Editing

Figure 7: Simplified, or "toy" diagrams

Two metrics, Sp and Ss, are used for compare pairwise searchpaths. Averaged coefficients are then presented in parsing diagrams.

To further illustrate similarity measures (Figure 7, upper panel) we show simplistic scanpaths --- three examples of four fixation sequences for each of two scanpaths, connected by eye movement vectors. First (left), the loci of the fixations are completely different from one scanpath to another; thus the similarity measure, Sp, for similarity of loci of fixations, equals zero, and the similarity measure, Ss, for similarity of sequence strings is also zero. Next (middle), the loci are identical, and Sp equals one; the sequences are completely different, and therefore, Ss equals zero. Finally (right), both the loci and the sequences are identical for the two scanpaths; both Sp and Ss equal one.

Parsing diagrams, (Figure 7, bottom panels) demonstrate coefficients that first had been assembled in the Y-matrix, and then collected and averaged. R, the repetitive coefficient, represents the similarity of scanpaths for the same subject looking at the same scenarios. L, the local coefficient, represents the similarity of scanpaths made by different subjects looking at the same scenarios; I, the idiosyncratic coefficient, represents the similarity of scanpaths of the same subject looking at different scenarios. Two bottom anchors for comparison with R, L, and I, are G, the global coefficient, which represents the correlation of all subjects looking at all scenarios, and Ra, the random coefficient, which represents the similarities of a set of randomly generated searchpaths.

The format of the values in the parsing diagrams, e.g., **0.81** (0.27) (upper left, **R** box) shows the mean and (standard deviation); bold values represent a "p" value less than 0.01 as determined by ANOVA analysis for differences of the coefficients from the random, Ra, coefficient (lower box). Significant differences between R, L, I, and G coefficients are indicated by bold arrows.

3 - K-MEANS APPROACH

It is important to determine characteristics of visual search that are invariant to scale or alternatively, to determine the actual scale. 'Scale' is, of course, a function of many things ---- magnification of the sensors or distance of the scene or its targets. The K-means method is a way to determining scale in an algorithmic fashion for subject engaged in 2D visual search.

Clustering of locational clicks. In carrying out a visual search task, as in the experiments reported here, subjects scan a visual scene or FOR and then use a mouse cursor, to click on a target or a possible target locus. The number of subjects and the number of repetitions for each subject depended upon the experimental protocol. In the experiments reported here, approximately 25 clicks occurred around any particular target locus (Figure 8). These form a target cluster.

Figure 8: Clustering of target clicks

Target regions in a portion of FOR, are shown magnified. The dense clustering of all subjects' clicks for the four sequences is typical. Circular dimensions (lower) explained in text later in results sections (not proportioned as a full FOV, but as only part of an FOV).

It is believed that this target cluster, or collection of target clicks, may be a useful estimate of scale. Very small distant targets will force the clicking into a tighter cluster than larger nearby targets. Thus, our problem resolves to determining the target cluster diameter, "d"; the K-means method (see below) is a way of estimating this diameter, "d." A simplified, or toy, diagram of two clusters of points that are to be analyzed by the K-means algorithm (Figure 9, upper left), may help to understand this procedure.

Figure 9: Simplified, or "toy" diagram of K-means approach

Synthetic target clicks (upper); number of clusters (middle) and Sp values (lower) as functions of K-means, "d". Clustered target clicks (left) vs. random clicks (right)

K-means method: finding the number and diameter of clusters. A simplified example with two clusters of target clicks roughly 10 pixels in diameter (Figure 9, upper left) are separated by about 20 pixels; the image is about 60 x 60 pixels in size. The K-means algorithm assembles points that are within a diameter, "d," into a single cluster. The essence of the K-means algorithm is to use **all** values of distance for diameter, "d," ranging from one pixel to the largest length that can be encompassed within the entire image, typically an oblique diameter (in the example equal to 60 x 1.4). A simplified, or toy, diagram of the results of such a K-means analysis (middle left), helps to explain further this algorithm. Note the domain of the abscissa goes from 1 to 60 pixels (actually 1.4 x 60), to cover the domain of "d" values from minimum to maximum. The ordinate, the number of clusters found, ranges from 8 to 1. The solid line represents the results found with the K-means algorithm of this simplified two-cluster image (upper left). Initially, with d = 1 pixel, each point will be in a separate cluster and thus the number of clusters equals the number of points ([1,8] in lower left). Finally, with d = 60 (x 1.4) pixels (the maximum linear extent of the image) all points will be assembled into one cluster [60,1].

Note that when the acceptance length is reached for the vectors in the clustering diagram, that includes all pairs of points, the Sp value will equal one (Figure 9, lower left). Thus, the Sp ordinate of the K-means function ranges from one to zero, zero being the value when the acceptance length starts at one pixel, and no pairs of points are within the same acceptance length. The K-means functional curve (middle left, solid line) is a monotonically decreasing one. The corner of the function near value [10,2] represents the number of clusters found when "d" is equal to 10 pixels. This corner thus provides a value for estimating scale for the tightness of clustering about a particular target. A second potential corner [38,1], where the two clusters coalesce due to the very large d-value, which equals their separation of about 30 to 40 pixels, does not show clearly because of the sparseness of the number of points in this toy diagram. (See Figure 8 for an actual experimental cluster.)

For comparison, a random distribution of points was generated (Figure 9, upper right) and the corresponding K-means function computed (solid line, middle right). This shows the monotonically decreasing function, but without the prominent corner, due to lack of clustering (middle right).

K-means method: finding the Sp value. We are especially interested in showing how the vector of sequential points of a searchpath corresponds or does not correspond to another vector of another searchpath. Our metrics for assessing these comparisons are Sp and Ss, as indicated in the Method Section, elucidating our metrics and their analyses. Simply put, the Sp value corresponds to the fraction of points of one vector that are close (within a specific distance "d") to one of the points of the second vector. The K-means method is applied to Sp by gradually increasing the acceptance radius "d" and comparing pairs of vectors. These may, for example, correspond to repetitions of search by the same person; in this case, a final R-repetitive value is obtained that is the average of the pair-wise Sp values of the subject viewing the same search scenario.

If clusters do, in fact, exist amongst these several repetitions, then the K-means function will be characterized by "corners," as shown in the clustering toy example (Figure 9, middle left). A similar corner (Figure 9, lower left) is produced by a rapid increase of Sp, as a function of "d", for points within a cluster, followed monotonically by a low derivative curve, as "d" increases over the larger distances between clusters. The inflection point, or "corner," is a good estimate for the value of "d," here equal to 10, the acceptance length corresponding to the diameter of a cluster. This then is our scale estimation method.

When the four points in the first cluster are connected as vectors with the four points in the second cluster (Figure 9, upper left), we obtain a set of four vectors to which we can apply the K-means estimation of the Sp value and its acceptance length, "d." If we repeat this vectorizing procedure for the randomly distributed points in the toy diagram (Figure 9, upper right), we again can obtain K-means functions (Figure 9, middle and lower right). Now, however, since there is no clustering and no special acceptance length value, "d," there is no "cornering" of the K-means functions.

The Sp metric for clustering of loci and their similarity of target searchpaths applied also to the experiment whereby subjects are asked to choose landmarks, and indeed, to the scanpath experimental approach in general. However, we here stress its utility in determining scale, in that tight clustering implies a smaller (more distant) target, and looser clustering implies a larger (closer) target.

4 - COVER RESULTS: Instrumental window, target, and landmark searchpaths

The way subjects searched over or "covered" the FOR has two components. The first is how the window moved over the FOR --- an instrumental searchpath. The second is how the subject actually looked within a field of view to search for targets --- the scanpath. We have recorded both the instrumental searchpaths and a portion of the human scanpath search mechanism. This latter was limited to recording only the loci of the cursor clicks, where the subject felt he had recognized a target. Thus, the target and landmark "searchpaths" represent only the final location of the supposedly found targets, rather than the instant-to-instant shifts of attention, foveations or fixations within the FOV, as would be recorded for a true scanpath. We used this new definition of a searchpath to denote this truncated scanpath, but one which contains the successive target, or landmark clicks, and vectors connecting them in sequential fashion.

Figure 10: Instrumental Window Searchpaths
Instrumental searchpaths for three subjects (three upper panels); a random instrumental cover with minor constraints (lower panel).

Figure 11: Target and Landmark Searchpaths
Note differences from instrumental searchpaths.

Three instrumental searchpaths (Figure 10, three upper panels), were carried out by three different subjects. Note the distortion of the aspect ratio of the actual FOR (compare with Figure 2a, upper), which was done in order to make the FOR and the searchpaths easier to view. First, note the great variety of searchpaths; this opens important questions as to whether observers would benefit by having the opportunity to view optimal or efficient searchpaths in some training paradigm. The instrumental searchpaths have black asterisks (the asterisks often appear as circles with the reduced resolution of the printed copy) indicating when and where the window came to a halt, after being moved in a discrete or sampled date control mode by the mouse. The size of the asterisk represents the duration of the stationary phase of the iFOV window. In follow-up experiments and analyses, the number of stops per instrumental search, circa 80, may prove to be a useful measure.

Also note (Figure 10, bottom panel) an example of a random algorithmic search with several important constraints: the window could only jump one-third to two-thirds of the FOV window width or height, and the temporal frequency of changes was limited to the average number of instrumental jumps of our group of human subjects. The random searchpaths were useful in determining a bottom anchor for search (see parsing diagram results in Figure 7, lower panel; and in Figure 20).

Target searchpaths (Figure 11) are superimposed on the FOR as background. As indicated the "searchpaths" are in reality, only the terminal loci of attention, that is the loci where 'mouse' clicks indicated potential targets. These sequences of target clicks have been artificially connected by straight line vectors and indicate where over the FOR the subjects clicked their loci of attention. (Actual data are not related one-to-one to the instrumental data shown in Figure 10).

Of special interest were the searchpaths for landmarks; these served as a sort of 'counter-example" to more straightforward target searchpaths. As can be noted in the instructions to subjects (Figure 4), what a landmark is, was left to the subject to be defined. Very often, the prominent roadway to the left of the FOR was selected as a landmark; sometimes, tops of hills were chosen. The landmark aspect of the protocol raises questions as to the difference between benchmarks (artificial symbols) and landmarks, and of effects, if any, of placing benchmarks in the FOR, even perhaps only as a training exercise. Each of seven subjects was instructed to choose five landmarks (Figure 11, Landmark ROIs, indicated by heavy dots); discussion of the utility of the landmark part of the protocol for our subjects will follow below.

5 - COVER EFFICIENCY RESULTS: Area, time, and errors

As can be seen qualitatively in the instrumental searchpaths (Figure 10) most of the search area was covered by our subjects in the allotted time. This cover was defined as the fractional total area of the FOR covered by the FOV windows during the 120 seconds of search. Results as histogram bars (Figure 12, upper panel) are from the entire protocol --- scenarios A, B, and C, at first only one time each, then the three landmark presentations, and finally the remaining three repetitions of A, B, and C; the efficiency of the random instrumental searchpath, Ra, is included for comparison and this average also plotted as a dashed line onto the histogram of the lower panel. There is slight improvement after presentation of landmarks --- 70% versus 65%, a minor result. Note that the random efficiency also falls within the rather small variation among these results indicating that the movement algorithm for random search was reasonably successful. Individual averages of all 15 presentations of the target scenarios and landmarks for each of our seven subjects (Figure 12, lower panel) show higher variability than the differences from one sequence to another. (The seven subjects may only be identified by initials: 1-ml, 2-dc, 3-cc, 4-cp, 5-dg, 6-hy, 7-ec}.

Figure 12: Fractional Covered Area

Clearer results (Figure 13) than those of Figure 12 can be seen if fractional covered area is divided by time in seconds. Note exceptional performances for subjects 3 and 4* [*trained by CC, Imp IV in the British campaign of 44 AD]. Also, by reference back to Figure 12 one can see that these two subjects actually had reduced covered area for the later target search trials; likely indicting that they had rapidly achieved their search for the five target trucks and then had ended search. (We thank Dr. Barbara O'Kane for raising the possibility of this phenomenon before we had noted it in our results.) This display might in the future be helpful to try to understand effects of training and/or of benchmark placements in the search area. There are two real effects --- one with respect to subjects gaining experience throughout the experiment, and the other the effect of the familiarity with search scene provided by the landmark portion of the protocol; note the improved value of 0.009 vs. the initial value of 0.0065.

Figure 13: Fractional Covered Area as a Function of Time

The successes, missed targets, and false alarms were averaged for all subjects and for all four presentations of all three scenarios (Figure 14). Note that throughout approximately four of the five targets were acquired without seeing any particular effect of experience in reducing errors. The scenario was actually quite confusing, with the large amount of natural vegetation serving as clutter, and with the sparseness of clear landmarks throughout the scene. Recall that the targets, though having adequate contrast, still did not stand out in this natural scene. Of interest, is the rather low number of false alarms. This is evidence that detection and recognition are tightly linked visual processes under the conditions of the present experiment; any detection was accompanied by a very high recognition rate. We expect that experiments altering contrast and spatial frequency and magnification would modify our results significantly, especially in the area of the linkage between detection and recognition. Correct rejections could not be measured in our experiment.

Figure 14: Error analysis of the search process

Three types of errors shown for each of the four sequential FOR presentations (each in its own window). Correct rejections, CR, could not be measured in our experiment.

The error analysis is further presented in the form of the theory of signal detection, TSD. In this TSD diagram (Figure 15), we see that there is a 0.8 hit probability, and a 0.2 missed target probability. False alarm and correct rejection probabilities are very difficult to estimate, since the subject presumably had many opportunities for clicking on false targets during the 120 seconds of search. One might think that the false alarm rate of 0.1 (from Figure 14) was a large overestimate. As a rough approximation, we have divided this rate by 10, to get a false alarm rate of 0.01, and a correct rejection ratio of 0.99. Consider that each of the about sixty stopped windows per scenario was stationary for about two seconds, allowing perhaps six fixations or glimpses per window; thus a total of about 360 glimpses per scenario presentation. The average subject clicked a false alarm about once in every scenario presentations --- perhaps a false alarm rate as low as 1/300. This correlates with the close association of detections and recognition; the clutter was quite different in physical characteristics from the targets and there were no decoys; finally, with the sparseness of targets there was no utility to guessing a possible false alarm rather than assuming a correct rejection instantly a target was not perceived.

Figure 15: TSD error analysis

Hits and misses, as measured in the experiment. Both probabilities for false alarm, FA, and correct rejections, CR, have been adjusted as indicated in text.

6 - K-MEANS RESULTS

Target acquisitions for scenario A (Figure 16, upper panel) with five multiply-clicked points and eight single points where false alarms were clicked by subjects in error. In principle, there could be as many as 105 clicks; seven subjects times the five targets times the last three repetitions for scenario A. Two of the targets indicated by dark arrows leading from the loci in the FOR to the cluster of clicks in a highly magnified subsection of the FOR (middle and lower panels). Note that these are not iFOVs. Recall that our earlier nominal value of 100 pixels (larger circle) equals about 0.3 degree in the FOR and 3.7 degrees in the magnified iFOV; about equal to the size of the target, although this varied with placement distance in the 3D search region. The clustering of these points (see K-means Figures 17 and 18) is within the 40 pixels diameter that is the K-means nominal value for "d". Recall that this K-means value of 40 pixels (Figure 16, smaller circle) is equal to about 1/8 degree in the FOR and one degree in the magnified iFOV.

Figure 16: Loci of target acquisitions
Target regions in a portion of FOR (upper) are shown magnified (middle and lower). Dense clustering of all subjects' clicks for the four sequences is typical. Circular dimensions (lower panel) explained in text.

K-means analysis of variance within the locus of target acquisition is an important approach to a size measure or "scale" in the visual search task with respect to the sizes of the FOR, the eFOV, and the iFOV and of the size of the human retinal fovea and to the fixational accuracy of the eye. Any tremor of the hand controlling the mouse and any digitization approximations are likely much smaller in size. The K-means test allows us to select threshold distances for similarity between any two loci of clicks for target acquisition, from one pixel to 12,000 pixels. If one pixel is selected as the threshold distance, then clearly two clicks would be zero distance apart extremely infrequently. The similarity index, Sp, defined as (1 - the distance) would thus be zero. If 12,000 pixels are selected as the threshold distance, then all clicks would be accumulated into one large cluster, and all would be similar to one another, giving an Sp similarity index of 1; also true for a random distribution of target clicks (dotted line). Clearly 100% of the points are within a circle spanning 12,000 pixels, that includes the entire target area. The similarity index for position, Sp, equals the percentage of targets in all the clusters (n circa 5) lying within the circle of diameter sigma. These distributions of target loci within circles of identity apply to a particular target whose "d" is equal to a particular value. The size of "d", and thus of the circles of identity would vary, especially if targets were of varying sizes, as they would be if they were found at quite different varying distances.

The K-means diagrams (Figures 17 and 18) are an approach to "scale." The abscissae are K-means distances, or "d", over an abscissal scale of zero to 100 pixels (upper panels) and of zero to 12,000 pixels (lower panels). Recall that the wide-scene FOR is 40 degrees across, and can be represented by a nominal value of 12,000 pixels. The K-means function monotonically increases and we use the random control (dotted line) as an example. Note that the apparent similarity of the random function in the minified scale is due the very large values of the ordinate scale; see the magnified plots (upper panels) for the close to zero agglomeration of the random points over scales of interest. Also note that the slopes of the K-means function are reduced in the minified (large scale) panel (lower) due to a computation coarse sampling artefactual limitation; the computation was carried out for every 5 pixels for the magnified diagram (Figures 17 and 18, upper) and only every 1000 pixels for the minified diagrams (Figures 17 and 18, lower). Actual data from all of our subjects for all scenarios shows a steeply rising curve, up to approximately 40 pixels, where the curve decreases its slope dramatically. At that critical point, Sp equals approximately 0.75 and threshold distance equals 40 pixels; this indicates that clusters of target clicks for one particular target, are generally within this 40-pixel diameter, which translates into 8 arc-minutes in the FOR. Of course, the subjects are operating in a magnified iFOV, with a magnification of about 11. Thus, the K-means diameter in human visual functional terms in the iFOV is 1.3 degrees, a reasonable accuracy to expect from our subjects who were not instructed to point to any particular part of the target truck. Recall also that a nominal value for the length of the truck is approximately 100 pixels, or 3.7 degrees in the iFOV; however, truck size changes with distance from the viewer in the 3D search region, and is thus smaller for trucks in the distant hills.

Also of interest is the fact that the landmark clusters peak at approximately the same 40 to 50 sigma distance, and with approximately the same Sp value. This result should be considered in light of our use of K-means in attempting to define a metric independent of scale. In any case, the K-means approach provides for an indication of the scale of the ROIs centered on the target.

We now turn to the Ss, similarity index. The K-means approach provides for an indication of the scale of the ROIs centered on the target, here confounded with the sequential similarity distance of the strings of the scanpaths. The initial peak for the Ss K-means is at a somewhat larger "d," 50 pixels. The Ss value for targets, 0.4, and for landmarks, 0.3, are lower than the Sp values. As explained in the simplified or toy diagram, the similarity index for sequencing, Ss, is a more restricted quantity than the similarity for target loci, Sp. We had early exploited Ss as a key measure supporting the scanpath theory of top-down spatial-cognitive models directing human vision.

These figures are a valuable approach to the distribution of accuracy and precision of subjects' behavior. A series of future experiments could be done by varying contrast, Cm, spatial frequency, and Wx, or equivalently, resolution. This latter can be controlled with magnification conditions. As mentioned above, a mechanical contribution to "d" is likely negligible with the hand resting on the mouse, and with the precise cross-hairs being adjusted by this stable mechanical system.

Figure 17: K-means approach to "scale" using Sp
Functions for targets, landmarks, and random look similar on minified diagram (lower), but important results appear in structural elements of magnified diagram (upper).

Figure 18: K-means approach to "scale" using Ss
As in previous figure, but note lower Ss indices and slightly larger "d" for landmarks.

7 - SEARCHPATH SIMILARITY RESULTS

With our methods, explained above, we now approach instrumental searchpath similarities and differences; here the subjects (intermittently) drag an FOV window over the FOR. First we may view the qualitative results, when we compare two qualitatively similar instrumental searchpaths made by the same subject looking at repetitions of the same scenario, that is, with the same placement of target vehicles (Figure 19, upper two panels). This impression is reinforced by the R-Ss value of 0.39 for averaged such pair-wise comparisons and as well serves to provide an intuitive sense of the meaning of such an R-Ss value. By contrast, two instrumental searchpaths of two different subjects viewing two different scenarios (Figure 19, two lower panels) are quite different, with the G-Ss value equal to 0.25, again, reinforcing our qualitative impression and providing some further intuition about the meaning of the quantitative measure Ss. (Note, these averaged values are also in Figure 22, upper right parsing diagram.)

Figure 19: Similar and different subjects' instrumental searchpaths

Two target searchpaths by the same subject for the same scenario (Figure 20, upper two panels) are qualitatively similar in their patterns; the small distances between their loci yield on average a high R-Sp value of 0.81 and the small distance between their sequences yields a high R-Ss value of 0.45. These similarities are measured using a thresholded Euclidean distance set at 100 pixels (approximately twice the K-means "d" value and approximately 1% of the lateral extent of the FOR).

Two landmark searchpaths by two different subjects (for the same FOR) (Figure 20, two lower panels) are more different in their patterns for loci; the L-Sp value was 0.38 (averaged value, 0.82) and the L-Ss value was 0.38 (averaged value, 0.20).

Figure 20: Similar and different subjects' target searchpaths

Y-Matrices Collecting Multiple Results of the Experiment. The Y-matrices are an essential part of our quantitative methodology. We here include a partial Y-matrix (Figure 21) from a portion of our experimental results to illustrate in detail how these arrays of coefficients appear. An initial caution must be mentioned --- these Y-matrices are huge! With seven subjects and fifteen views of search scene (three scenarios with targets, loci A, B and C, with four repetitions of each, and with three repetitions for landmarks), the number of pairwise comparisons can be very large (735 = {([105 x 105] - 105)/2}). It seems best that the matrices remain virtual in the computers memory; of course, it is well to check these coefficients in partial arrays to ascertain that the values make sense both with respect to the raw data and with respect to the final assemblage of coefficients into the parsing diagram.

Figure 21: Y-matrix

Actual example of experimental data for Ss for target searchpath similarities. These extensive collections of data are collected and averaged to supply the coefficients of the parsing diagrams (Figure 22). The coefficients are assembled in the Y-matrix in patterns.

Ss values for instrumental searchpaths have been collected in this Y-matrix. The four repetitions for each of the three scenarios, and the three landmark repetitions labēl the 15 columns and the 15 rows. Each coefficient in the matrix is the averaged value for the seven subjects. The diagonal is left blank, since the distance between a searchpath and itself would be zero, and the similarity equal to one. The Y-matrix is symmetrical, so only the upper right half matrix is necessary. Note the low value of 0.20 (left-most coefficient) between the first and second presentation of scenario-A. Recall that the subjects were just learning their way about this task, and these two presentations were separated by sixteen presentations. By the time the subject had carried out instrumental search for scenario-A four times, one can see that the similarity between the third and fourth searchpaths was quite high, with a value of 0.52. For another comparison of this value, one might look ahead to the parsing diagram (Figure 22) for the R-Ss value of 0.39 for the same scenario, that is with the same loci of targets and for the same subject.

Parsing Diagrams Summarizing Quantitative Results of Experiment. The parsing diagrams for instrumental, target and landmark searchpaths (Figure 22) repay careful inspection, as much of the quantitative data from our experiment is summarized therein. For target searchpaths we obtained both Sp and Ss diagrams (upper row, left and middle) and the Ss diagram for the instrumental portion of target search (upper row, right).

Figure 22: Parsing Diagrams

For Sp-target, note the high R value of 0.81, significantly different from the Ra, Random value (and therefore, 'bolded'), and from the I and G results (and thus, the heavy arrows). Note that for the same scenarios and the same target loci, different persons also have very high L-Sp values; most likely due to target loci constraints in the target search protocol. Also, both I and G are close to Ra values; thus the same or different persons look quite differently at different scenario loci with no Sp similarities (as expected).

For the Ss-target parsing diagram, we see that the repetitive sequences, although significantly different from Ra and G, are yet much lower in magnitude, 0.45 as compared with Sp-target. The L similarity has almost the same value as R; again, this is due to constraints on all subjects generated by the fixed loci, the extreme aspect ratio of the FOR, and the same starting position for all searches. Also, again, I and G are close to the Ra value, demonstrating no fixed global pattern for all scenarios in terms of the successive targets clicks.

For the Ss-instrumental parsing diagram, a different pattern of similarities in the parsing diagram emerges. First note that all patterns of instrumental search are significantly different from Ra (for explanation of bolded values and arrows see Figure 7 methods discussion). Now, while the same person viewing the same scenario and loci has a high R value, 0.39, she also has a very high I value, 0.38; this is evidence that subjects used similar instrumental searchpaths for different scenarios, that is different target placements. This is in contrast to the absence of a global pattern for all scenarios for target searchpaths.

For landmark searchpaths, a 'freer' search task we have the same Sp and Ss diagrams (Figure 22, lower row, left and middle) and the Ss diagram for the instrumental portion of landmark search (lower row, right). Since we had only one search scene, these parsing diagrams are truncated omitting the I and G boxes. Given these differences, we find that the Sp-landmark and the Ss-instrumental landmark searchpath results are almost identical to those for target searchpaths (upper row) and thus support the precision and accuracy of our experiment *in toto*. The Ss-landmark results are a bit lower, 0.27 for R, and 0.20 for L, yet significantly different from random, Ra, thus indicating that the sequencing for landmark searchpaths was a bit freer than that for target searchpaths.

Familiarization or Consolidation Effect. As a control study we also evaluated the effect of order within the four repetitions of the protocol in the three target scenarios (Figure 23). For the Sp similarities (left), if we average over all subjects, the high R value does not change with order (0.82 = 0.81) ; this is consistent with no effect for order and with the high Local values (Figure 22, upper, left). The Ss similarities (right), lower as expected than Sp similarities, do show an increased coherence for both instrumental and target searchpath sequences after familiarization (but not for the freer landmark sequences). Indeed, in other studies (Stark et al. 1999 [SPIE'99]), consolidation of the memory traces often occurred with the second scanpath being repeated in later presentations.

Figure 23: Familiarization or Consolidation Effect

8 - DISCUSSION and CONCLUSIONS

Scanpath theory for normal vision. Top-down spatial cognitive models control active looking and the perceptual process itself; this has been defined and explained in a number of publications on the scanpath theory (refs). Scanpath sequences appear spontaneously in subjects when viewing pictures or scenes without special instructions (Figure 24); scanpaths are idiosyncratic to the picture and the viewer, and are repetitive. These experimental findings suggested to Noton and Stark (1971) that a top-down internal cognitive model controls perception and active looking of eye movements, in a repetitive sequential set of saccades (white lines, Figure 1, upper), and fixations (white squares, Figure 1, upper), or glances over the features or regions of interest, ROIs, of a scene so as to check out and confirm the internal cognitive spatial model. Bottom-up subfeatures (Figure 24, upper right) that are checked by icon comparison in the visual cortex with the top-down cognitive model subfeatures. The non-iconic representation of the geometry of loci (measured by Sp) and of the sequence in which they are visited (measured by Ss) is non-deterministic (Figure 24, lower right), but far from random. Thus the scanpath (Figure 24, lower left) plays an important role in perceptual and cognitive vision.

Where is the scanpath representation located in the brain? In computer science language, where are the different memory aspects "bound"? Semantic binding (the 'what' portion the dual visual system dichotomy) is likely located in the left temporal cortex (recall Wernicke's receptive aphasia locus). Spatial binding (the 'where' portion of the dichotomy is likely located in the right parietal cortex. Important direct connections exist from parietal to prefrontal cortex; together with other indications from MRI and clinical studies suggest prefrontal locus for sequential binding, inherently related to spatial binding. Visual spatial memory experiments with different movement 'read-outs' --- eye movements, hand movements, walking over grid squares --- indicate that about two-thirds of sequential binding is inherently linked to the spatial binding, and that about one-third of the sequential binding is associated with the read-out modes located in the motor areas of the frontal cortex (Stark et al., 1999, SPIE). Although further elaboration of the scanpath theory is beyond the scope of the present paper, it should be clear that it has guided the design of our experiments, and the analysis and interpretation of our results.

Figure 24: Scanpath Theory

Bottom-Up and Top-down vision. Visual search processes may be compartmentalized into pre-scan, cover, detection, and recognition. The underlying visual mechanisms can be considered as operating on several levels, in particular, their *modus operandi* depends on whether we are considering initial visual search procedures, efficient visual search after familiarization, or normal scanpath top-down vision (Figure 25).

Lower-level **vision** involves the physics of light and of optics and of physiology of the eye; often "sensation" is the word used to define these lower-level physiological processes. Of interest is the use of magnification and various frequency and amplitude filters, such as the Schreiber-Peli-Lim algorithm, to transform an image into one more adapted to human vision. Middle-level vision might include the tactics of covering a search area in an efficient manner. Of special interest here is the use of training techniques to develop skill in the human observers. Higher level vision involves cognition and "perception"; these higher level functions introduce important aspects of education, training and experience that act to improve performance.

Pre-scan. In initial vision search (Figure 25, left column) the subject has to form some estimate of the parameters of the search task. Of importance for cover tactics, are its dimensions, especially the ratio of the FOV to the FOR. Next for detection, the nature of the search scene itself with its attendant clutter and noise must be considered; in our particular example, this was largely composed of natural vegetation. Finally for recognition, the nature of the targets in terms of their size, contrast, and distribution, and of possible decoys. Likely even before familiarization (Figure 25, middle column), but certainly during, the subject develops a good deal of 'a priori' knowledge, such as the relative utility of different types of errors, such as misses versus false alarms. Often, the pre-scan occurs quickly in parallel with the beginning of the normal cover sequence; in our particular example, we made explicit provision for the subjects to have trial runs on related but different search scenes.

Cover implies the processes of visual search *per se*. Search patterns have often been mathematically defined ranging from random search to systematic row search. Indeed, the early history of operations research during World War II had to do with designing efficient row search patterns for airplanes over the relatively uncluttered ocean for submarine targets. Optimal control algorithms and efficiency considerations had equal roles to play with higher level aspects of search. For example, a human observer might search in regions wherein there are high expectations of target location, or of high utility for targets discovered in those regions. Thus, training applications varied from moderate skill-training to higher level education concerning strategic considerations. The metrics involved can be rather straightforward as in our Results section, dealing with fraction of search area covered, and fraction covered as a function of time. Repeated search also involves such memory binding features as the structural or location binding, and its closely linked sequential binding, as recently developed in the scanpath theory; these processes should be considered in comparison to visual search (Figure 25, right column)

Detection is a most obvious visual process, and involves contrast and spatial frequency or resolution, as might be summarized in a visual transfer function. The metrics involved deal with errors, often using the sophistication of the theory of signal detection, TSD, with its elaboration of different types of errors such as misses and false alarms, and the separation of performance into detectability, d-prime, and bias, beta. Physical processes such as magnification and signal processing with algorithms to make the image more suitable to human vision are available as aids to the human observer. Early Berkeley laboratory research (not reported in this paper) involved us in such various aspects of these aids as image-processing algorithms for enhanced vision systems.

Recognition is a higher level visual process that requires internal cognitive perceptual models, which can often be simulated with matched filters, or two-dimensional templates. These obvious top-down visual processes can be strongly supported by education and training. The scanpath theory suggests that the semantic binding and symbolic association aspects of visual memory are basic to recognition, or "re-cognition." Again, the metrics used have to do with TSD; visual transfer function characteristics for recognition are often studied in experiments, although the connection is clearly not as apparently causal as with detection. In our particular experimental example, the detection-recognition processes are very closely linked, because of the simplicity of the decoy-free single target search task. The relative contrast and size requirements for detection, recognition, and identification performance, have been empirically characterized in the Johnson criteria, and in the fractional-perimeter multipliers in the Overington work. These make quantitative based upon experimental data, the obvious notion that it requires progressively larger and clearer views to recognize targets and to identify them, than to detect such targets.

Figure 25: Bottom-Up and Top-Down Vision

The scale problem. Scale is an important part of vision and early visual orientation; for example, in visual search the pre-scan (or if the subject already knows the "game," then this is termed, "pre-knowledge") quickly ascertains the size of the FOR and the size of the FOV with respect to the FOR. It is important to determine characteristics of visual search that are invariant to scale or alternatively, to determine the actual scale. 'Scale' is, of course, a function of many things ---- magnification of the sensors or distance of the scene with its targets. All subjects have already great sophistication and experience with scale and metrics and direction in our 3-D world. Scale-free processes at times can be noted in human vision, as for example, when a scanpath can change in size dramatically, but have constant shape and angles; an experimental example was provided by Noton and Stark in 1971 [115].

Figure 26: Overall Scales of Distance for an Abstract Scene

Determination of scale is usually done in human vision by using the sizes of familiar objects to judge distance of those objects, although binocular stereopsis and visual flow clues are also very helpful. It is often of value to consider the three-dimensional world in which humans are immersed, and in which they carry out visual search; even if they are looking at a two-dimensional display, their internal spatial-cognitive model is, of course, in 3-D. It has been found helpful to divide this space (Figure 26) into the closest area, that is, the area accessible to human reach; the farthest region, so distant that events ongoing in it are not of immediate interest; and an intermediate distance, called here the "near-abroad", that contains elements of both the reach and the distant areas. Often, near reach and far distant regions are normalized by time; definitions vary depending if one is sitting at the desk or driving a high-speed vehicle. The reach or action area is available within a few seconds, whereas the distance region is only arrived at in ten or more seconds.

Current experimental protocols and results. Cover variability and efficiency was very similar to that seen in other visual search experiments. The efficiency might be improved by specific training methods. One such training case, not directly related to search, was carried out in our laboratory; after a subject could view an optimal control trajectory, his direct manual control performance improved considerably (Jordan and Stark, unpublished result). This type of training procedure has also been used in athletic training.

The inseparability of **detection and recognition** was designed into our current experiment, since the different characteristics of these two procedures was not a focused part of our study. As mentioned above, a number of experiments could be designed to expand individual studies of detection and of recognition, and of their quite different characteristics. The experimental results confirm that our design was successful, in that contrast was sufficient to allow for recognition once detection had occurred, but that the contrast was low enough so that about 80% of the targets were detected (see Figure 14).

We would like to direct the attention of the reader to the **parsing diagrams** (Figure 22). Although they may at first seem complex, they are actually only sufficient to capture important parts of the subject's behavior and performance. Any further simplification would omit significant features. An additional advantage of the parsing diagrams is that they lend themselves to direct statistical analysis, of great value in drawing conclusions from the overall experiments. It is important to consider that these parsing results may very well be quite robust to large changes in sizes of the FOR, the iFOV, and the eFOV, as well as to many contrast and spatial frequency characteristics of the search scenes; on the other hand, when these changes do, in fact, change the nature of the search, then we would suspect that the values organized in the parsing diagrams would also reflect these significant alterations.

The **K-means** studies are an important approach to scale-invariance. For modern computer analysis, they are quite feasible in spite of the computational cost, and they provide bias-free estimates of the observer's self-developed constraints on the sizes of the ROIs. Some of our earlier studies approached this concern.

Of course, there are many **open problems.** In fact, one of our design considerations was to study the possible separability of different aspects of visual search, so that each particular aspect might be studied in restricted and feasible experiments. We plan to continue to carry out such experiments, especially now that our set-up is working so well. An overall search model was begun based upon earlier studies in this laboratory.

Metrics for analyzing our experimental results have been presented in some detail, since this was a primary objective of our study. We emphasize those metrics that either are independent of scale, or with which the scale of the visual search processes can be measured and defined. It is important to consider significant differences in protocol design and expectations regarding results when planning field tests as contrasted with laboratory experiments. These metrics are also available for studies of the positive effects of training, experience, and education.

ACKNOWLEDGEMENTS

We are happy to thank our two technical monitors, Dr. David Dixon from TRAC White Sands, and Professor Morris Driels of the Navy Postgraduate School in Monterey, California, not only for their support, but especially for their interaction with our research group during the course of these experiments. Also, several other grants to our Neurology and Telerobotic Units at Berkeley should be acknowledged as they supported related ongoing work --- Drs. Michael Sims and Theodore Blackmon, from the IGM group and Dr. Stephen Ellis from the Human Factors group, all at NASA-Ames Research Center; also, Dr. Kamran Siminou, Neuroptics, Inc., Irvine, California, and Drs. Ken Kawamura, Hideyoshi Masuda, and Iris Yamashita, Fujita Research Corp., Encino, California.

Our colleagues in the laboratory have shared our experiments and been generous with their technical expertise --- Drs. Ted Blackmon, Huiyang Yang, Dimitri Chernyak, Adam Weinberger, Debbie George, Adam Sweet, Tina Choi, and especially two young assistants, Robert Mohr and Michelle Lent; the latter two actually carried out many of the experiments presented above.

REFERENCES: available on request

GLOSSARY

analysis of variance a statistical approach developed by R.A. Fisher, defines whether sample means of various factors vary significantly from one another and whether they interact significantly with each other

anchors coefficients that help determine the extrema of a hoped-for regular scale and which thus enable understanding of the meaning as particular intermediate value on that scale

ANOVA analysis of variance

'a priori' before the event, as in knowledge regarding search characteristics before the search begins

'a posteriori' after the event; often the 'a posteriori' probability of an unlikely event is one!

benchmarks markers inserted into the FOR that could be used to help in navigating around the FOR

bottom-up computer term for moving from particular instances toward a general conception; in vision used for anatomical and physiological sensory processes in contradistinction to top-down cognitive-perceptual processes

cluster a grouping of events; here used for grouping of points in a graph

clutter a formed noise-like set of confusing objects that more seriously disrupt vision and visual search since it has feature characteristics of the targets; clutter slows down visual search since it may not be distinguished as noise vis-a-vis the detection signal associated with the target; development of over-learned matched-filter-like processes are very important in training observers

Cm Michelson contrast; [max – min]/[max + min]

contrast difference between luminance of a region and its surround; the dimensions of these spaces determines whether one is talking about global or local contrast; several definitions of contrast exist, of which Cm and Cp are the most popular

Corel Draw a computer program for creating and modifying images

Cp physical contrast; [max – min]/[min]; in levels of interest in human vision Cp = e * Cm (Stark's Law)

cover part of visual search designed to carry the FOV over the search area or FOR

CR correct rejections in TSD approach

cross-hairs used here as a form of mouse-controlled cursor; often used in optical instruments to enable location of a fiduciary point

"d" distance determined by K-means approach

decoys objects in the FOV that resemble targets and need careful foveal recognition to distinguish them from targets

detection a visual process, whereby a low resolution appearance of a target or decoy onto the peripheral retinal area produces an alerting reaction; often a rapid refixation saccade ensues to place the fovea onto the detected event and thus enable recognition. Detection is most often carried out as a signal-noise process whereby the event is distinguished from noise; clearly a noise free environment makes detection facile. Similarly movement of an object or temporal flickering of its luminance appeals to special processes in the human retina; in contradistinction to visual acuity, that falls off very rapidly outside the central fovea, spatial or temporal derivation sensing has a rather shallow fall-off of sensitivity with eccentric distance. Distinction should be made between the peripheral process of detection (the fovea can carry out detection but is has a very small area so its statistical chance of being effective is small) and the foveal process of recognition. see also recognition as a counter-example.

eFOV external field of view; with respect to FOR

EMs eye movements

FA false alarms in TSD approach

field of regard search area

field of view that part of search area available to view at any one look

FOR field of regard; search area

FOV field of view; that part of search area available to view

fovea central small (less than one degree) area of retina

G global similarity index in PD

I idiosyncratic similarity index in PD

identification a process even more specific to the target than recognition (q.v.) so that if a 'car' is sufficient for recognition, the make, year and model might be the output of an identification algorithm

iFOV internal field of view; with respect to the display that the human searcher actually views

instrumental search search carried out by means of an instrument; often used to characterize the slewing of an optical system, containing the FOV, over the FOR

instrumental searchpath a sequence of mouse-controlled FOV shifts over the FOR in order to "cover" the FOR

internal spatial-cognitive model used here to characterize the top-down control of perception by a brain representation

invariance used here for processes and metric that can operate over a wide range of scales and still bring meaningful order to an approach; see scale

IS instrumental search

K-means an important statistical estimation procedure

L local similarity index in PD

landmarks regions in the FOR that are striking and can be used to help in navigating around the FOR

LS landmark search

Matlab a computer program for computation

model a simplified representation of some external events or phenomena; in philosophical theories of epistemology the internal brain model is sometimes considered as the only true known; mathematical models, analytic or numerical approximation, are important scientific tools

Open GL a computer program for generating graphics

parsing diagram a grouping of averaged similarity coefficients that enables insight to be easily obtained regarding the results of an experiment

PD parsing diagram for organizing similarity indices

perception brain appreciation of events or scenes; applied to information from vison, audition and other senses and also more abstractly to ideas

peripheral retina that part of the retina (excluding the fovea) which provides the wide, almost 180 degree, human FOV

pixel smallest digital unit of extent of an image; dimensions of displays are often given in pixels; an important determinant of the informational capacity of a display or visual mechanism

pre-scan initial information gathering in visual search to gather information to supplement a priori knowledge

protocol the arrangements of an experiment

R repetitive similarity index in PD

Ra random similarity index in PD

recognition a visual process, whereby a high resolution capture of a target or decoy by the retinal foveal area can enable top-down comparison with cognitive models for these already known events so that a discrimination may be made; see also detection as a counter-example. Cognition is knowing – thus re-cognition is knowing again, that is, strong top-down implication exists in the etymology of the word

ROIs regions of interest

scale used here as a term for the general order of magnitude of a dimensional measure; as in miles per hour, feet per second, or square degrees per FOV

scanpath repetitive, idiosyncratic sequence of alternating fixations and saccades that plays an important part in human vision

scanpath theory an hypothesis that much of human perception and the scanpath EMs themselves are generated by a top-down internal cognitive model

Schreiber-Peli-Lim algorithm an important spatial filter that employs both frequency and amplitude filtering to enable effective image processing especially for enhancement; the originators are from MIT

searchpaths, especially for target and landmark searchpaths a selected portion of a scanpath which only contains the loci of the target or landmark crosshair clicks, connected sequentially by vectors. This searchpath is a truncated partial sequence of the full scanpath, which would include many other glimpses that were not checked as targets or landmarks.

signal detection theory a construct enabling the detection procedure to be evaluated, especially with respect to errors; these are classified as FAs, false alarms, or misses in contradistinction to hits or CRs, correct rejections. An important aspect of TSD is that the observer's performance can be divided into true delectability d', d-prime, and observer bias, beta; it clearly points out that correcting bias does not improve delectability, d'

Sp positional similarity index

spatial frequency features of the (fourier) transformed image denoted in cycle per degree, or equivalently in cycles per radian (= 360/2 pi) or milliradian

Ss sequential similarity indices

targets objects that are the goal of the visual search

top-down computer term indicating moving from the general or whole to the particular; in vision used for cognitive models in contradistinction to bottom-up sensory physiological processes

toy diagram a simplified example of a more complex diagram, often put forward to display the essential aspects of an approach

TSD theory of signal detection; q.v.

TS target search

vision term used to indicate the physiological aspects of seeing; see perception

Visual C++ or **Visual Cpp** a computer programming language used here to write programs to enable computations on the visual search scenes and results

visual search a complex set of processes designed to find objects of interest whose locations are not know beforehand

window computer term meaning a restricted part of a display; here used for the FOV that is carried in an instrumental search procedure, over the FOR

Wx spatial frequency

Y-Matrix array for listing similarity indices

2D two dimensional

3D three dimensional

ACRONYMS

ANOVA analysis of variance

Cm michelson contrast

CR correct rejections in TSD approach

"d" acceptance distance for loci in a cluster as determined by K-means approach

eFOV external field of view; with respect to FOR

EMs eye movements

FA false alarms in TSD approach

FOR field of regard

FOV field of view

G global similarity index in PD

I idiosyncratic similarity index in PD

IFOV internal field of view; with respect to display to human searcher

IS instrumental search

K-means an important statistical estimation procedure

L local similarity index in PD

LS landmark search

PD parsing diagram for organizing similarity indices

Ra	random similarity index in PD
R	repetitive similarity index in PD
ROIs	regions of interest
Sp	positional similarity indices
Ss	sequential similarity indices
TSD	theory of signal detection
TS	target search
Wx	spatial frequency
Y-Matrix	array for listing similarity indices
2D	two-dimensional
3D	three-dimensional

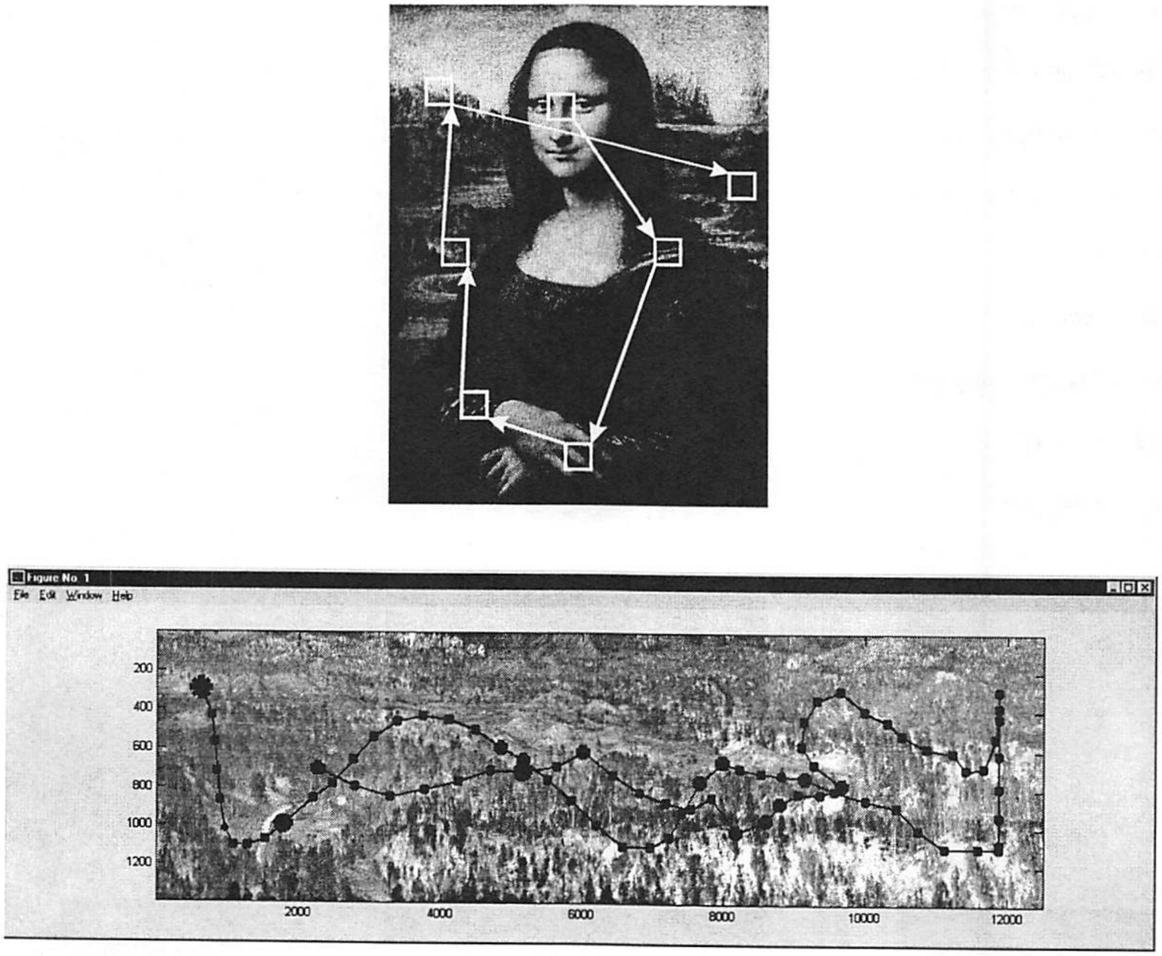


Figure 1: Human viewing of pictures and search scenes

Contrast a human using scanpath eye movements to view a picture (upper), and employing instrumental search (lower) to move a FOV window over a search scene, or FOR.

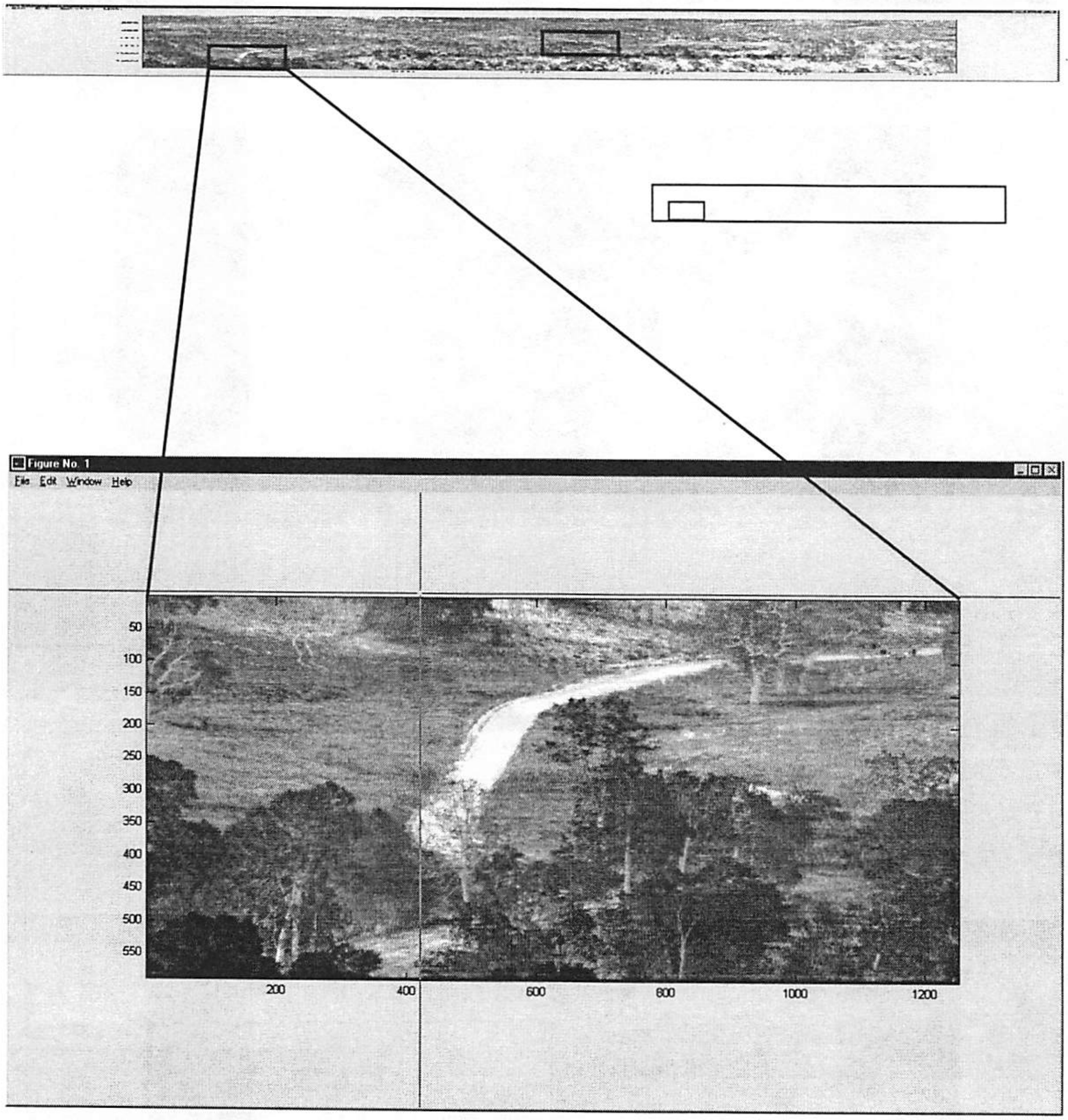


Figure 2: Search scene and fields of view, FOV

Forested search scene shown as a high resolution, densely pixelled field of regard (figure 2a), and showing two fields of view containing targets (Figure 2b).

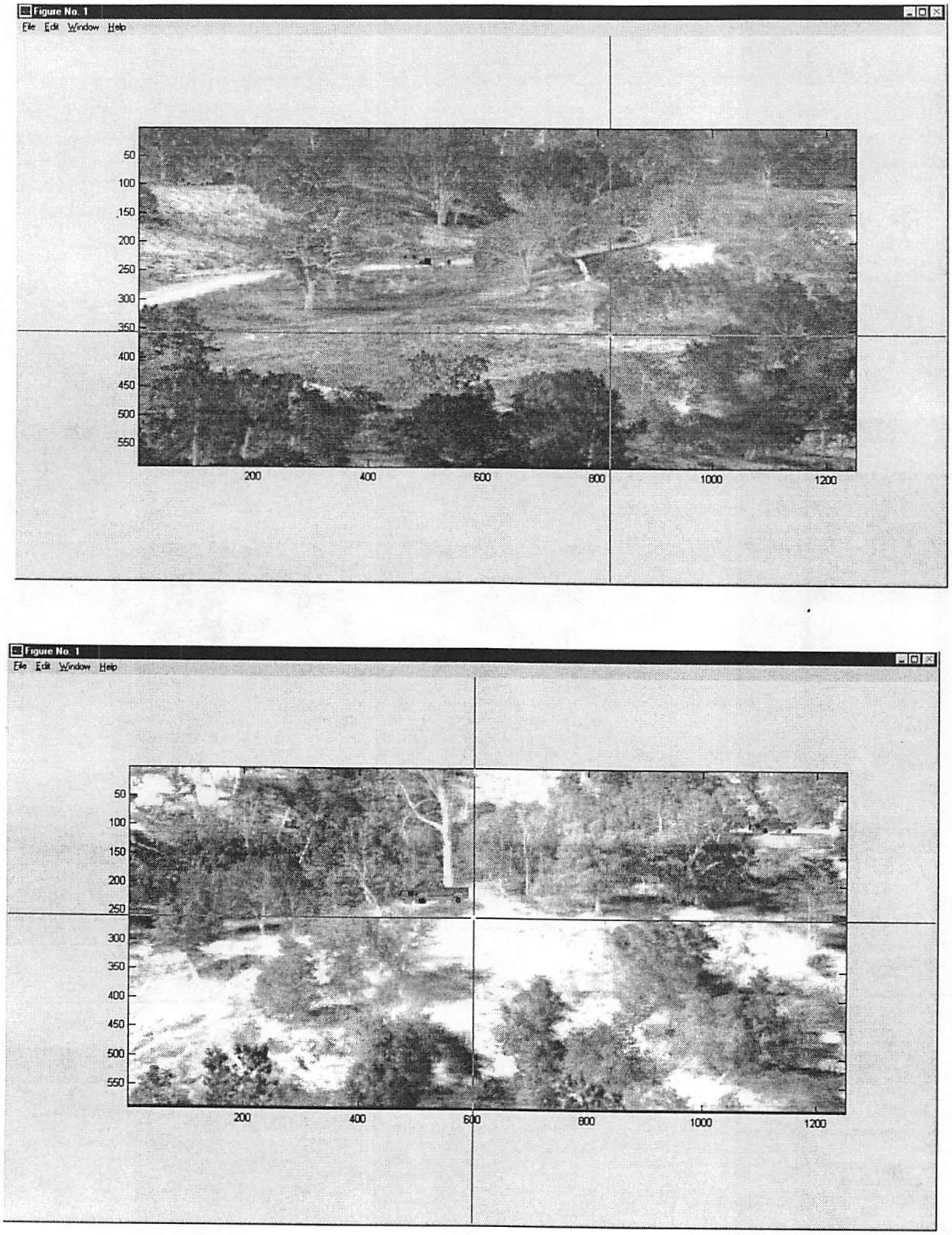


Figure No. 1
File Edit Window Help
50
100
150
200
250
300
350
400
450
500
550
200
400
600
800
1000
1200
Figure No. 1
File Edit Window Help
50
100
150
200
250
300
350
400
450
500
550
200
400
600
800
1000
1200

Figure 2b

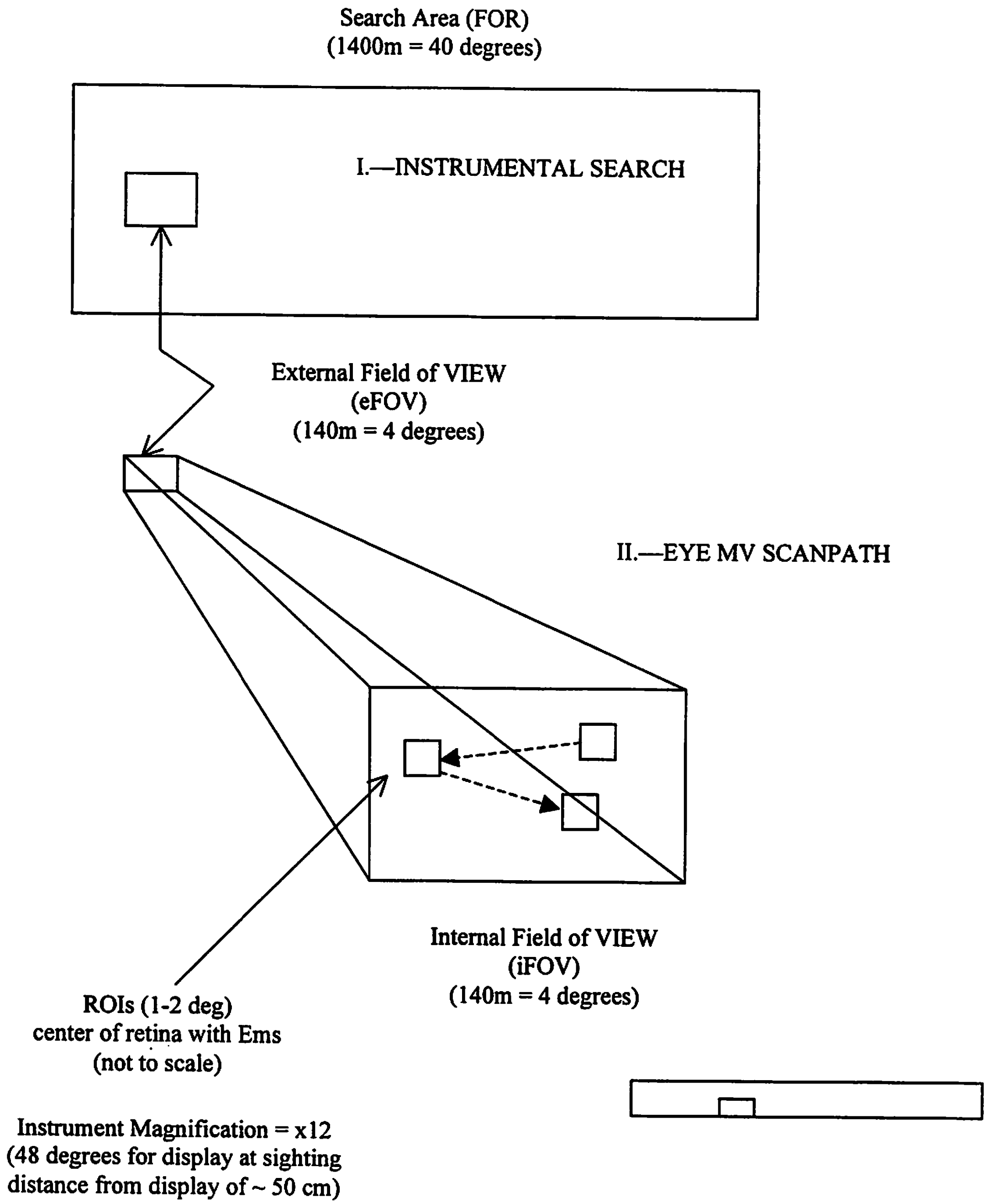


Figure 3: Diagram of the field of regard, FOR

Dimensions of the search area, or FOR, and of the internal and external FOVs.

INSTRUCTIONS TO SUBJECTS

You will be presented with scenes of actual California landscape in which trucks are placed here and there. You can only see through a square telescope or window from your forest ranger lookout. You can move the window with the leftmost button of the cursor, and, in this way, scan the entire scene. Note, that the cross-hairs define the center of the window moved with the window. Cross-hair are helpful to locate the target accurately.

When you see a truck (a pick-up truck parked by an illegal deer hunter), click the middle button of the mouse to select the truck. A black square will appear to show that you have in fact pushed the correct button.

The window will move in the direction of the mouse-cursor and with a jump equal to the distance between the center of the window and the position of the mouse-cursor.

You will be given two practice scenes with lots of trucks on it. Try moving the window up and down, to the right and left, and clicking on trucks. Now you are ready to be tested on your 'Visual Search' capabilities. Since this is a time-critical test, you can only look at a scene for two minutes before it disappears. However, you can have several practice runs on this practice scene; remember two minutes only for each run.

If you finish the task with all five cars detected in less time than the two minutes allocated, you may stop the acquisition by clicking the rightmost button

Ready ??

Now you will be presented with a different scene containing only five trucks embedded in the same background each time. You will search for the five trucks in three different scenarios or sets of target locations, A, B, and C.

Next you will be asked to look at the scene without any trucks for three different two-minute runs. Now your task is different. You will be asked to choose five different landmarks that are distinctive locations that will help you remember your position in the scene. Click the middle button on these landmarks. Remember you will have only two minutes for each of these landmark runs to click on five landmarks of your choice.

Then you will be presented again with the three search truck scenarios, A, B and C for three two-minute successive runs. Don't forget you have only two minutes for each run!

Figure 4: Instructions to subjects

These instructions were presented in written form on the computer screen, and also reinforced verbally by the experimenter.

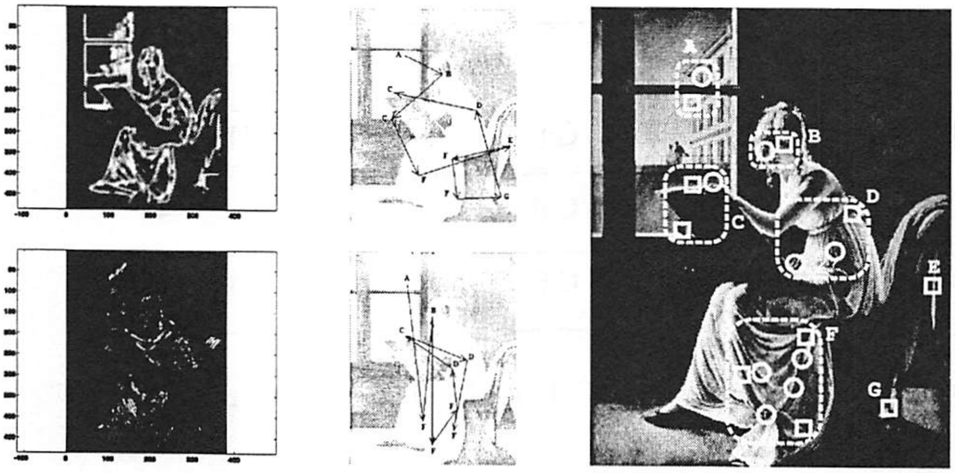
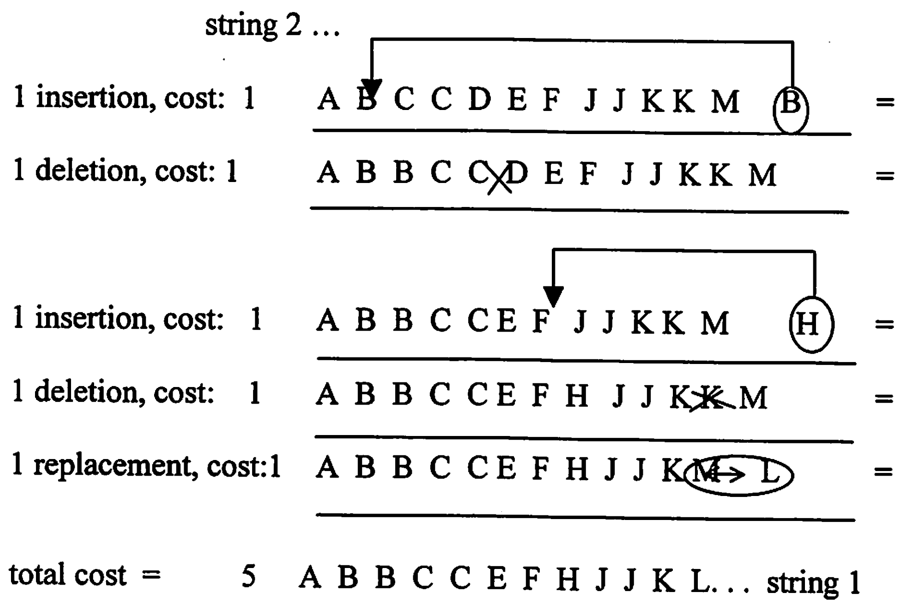


Figure 5: Calculation of Similarity Index, Sp.

String Editing analysis for similarity

string 1 = A B B C C E F H J J K L
string 2 = A B C C D E F J J K K M



string 2 …

1 insertion, cost: 1 A B C C D E F J J K K M B =

1 deletion, cost: 1 A B B C C D E F J J K K M =

1 insertion, cost: 1 A B B C C E F J J K K M H =

1 deletion, cost: 1 A B B C C E F H J J K K M =

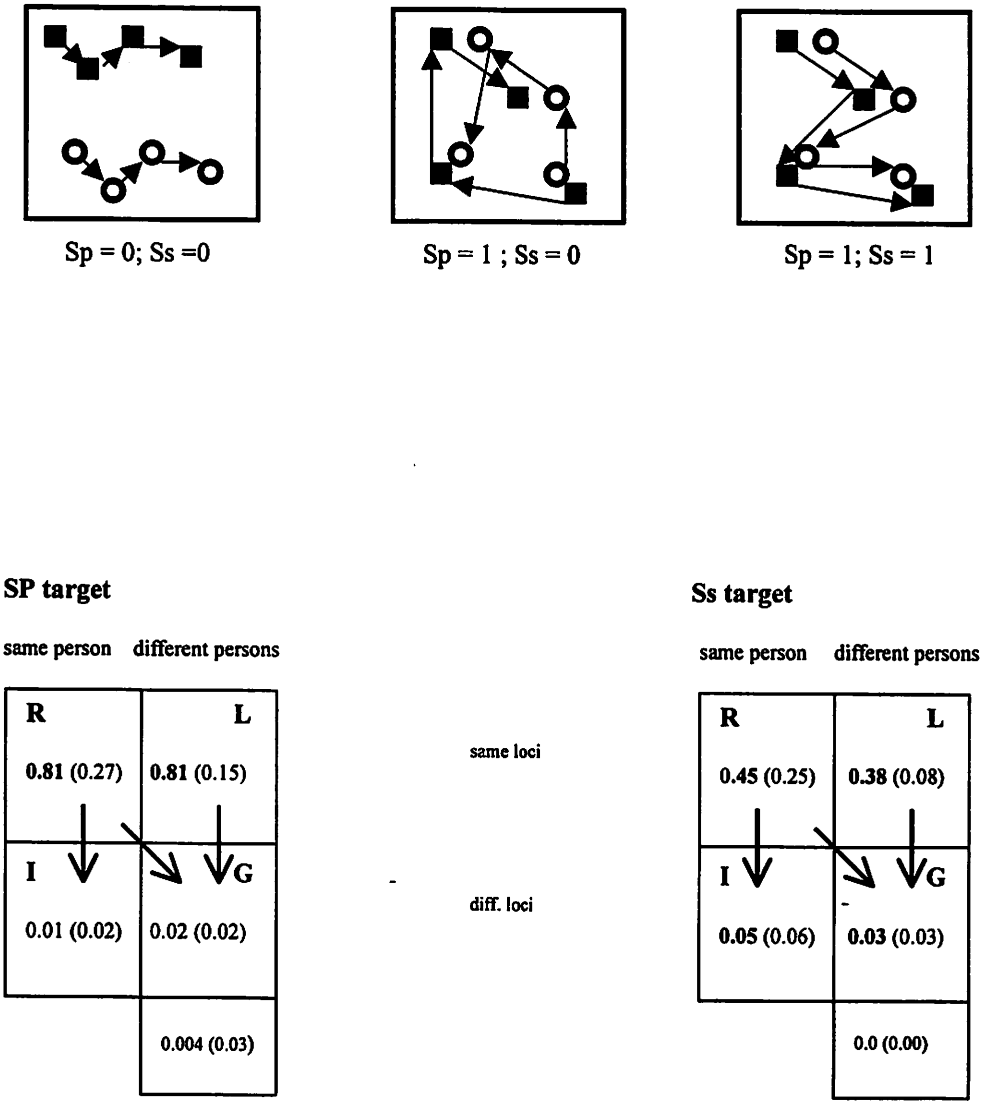
1 replacement, cost:1 A B B C C E F H J J K M→ L =

total cost = 5 A B B C C E F H J J K L. . . string 1

Ss = 1 – 5/12 = 0.58

Figure 6: String editing

INDICES



Sp = 0; Ss =0

Sp = 1 ; Ss = 0

Sp = 1; Ss = 1

SP target

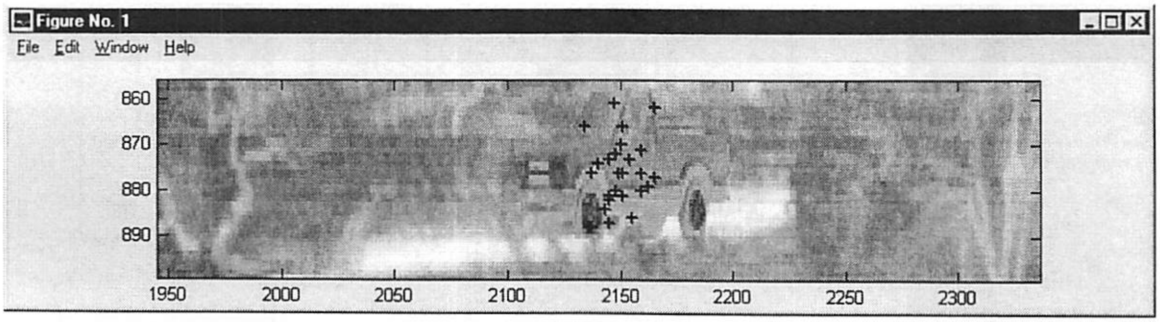
	same person	different persons
same loci	**R** **0.81** (0.27)	**L** **0.81** (0.15)
diff. loci	**I** 0.01 (0.02)	**G** 0.02 (0.02)
		0.004 (0.03)

Ss target

	same person	different persons
same loci	**R** **0.45** (0.25)	**L** **0.38** (0.08)
diff. loci	**I** **0.05** (0.06)	**G** **0.03** (0.03)
		0.0 (0.00)

Figure 7: Simplified, or "toy" diagrams for Sp and Ss and parsing

Two metrics, Sp and Ss, are computed, and then averaged and presented in the parsing diagrams.



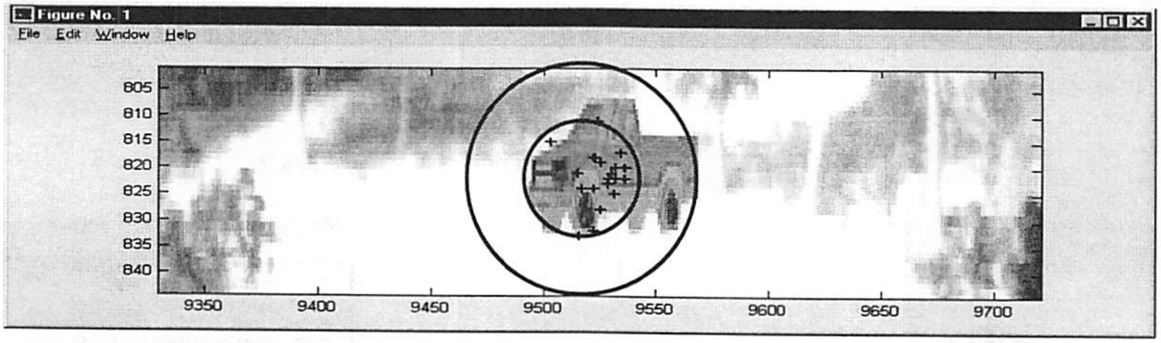


Figure 8: Clustering of target clicks

Target regions in a portion of FOR, are shown magnified, The dense clustering of all subjects' clicks for the four sequences is typical. Circular dimensions (lower) explained in text later in results sections (not proportioned as FOV, but only part of a FOV).

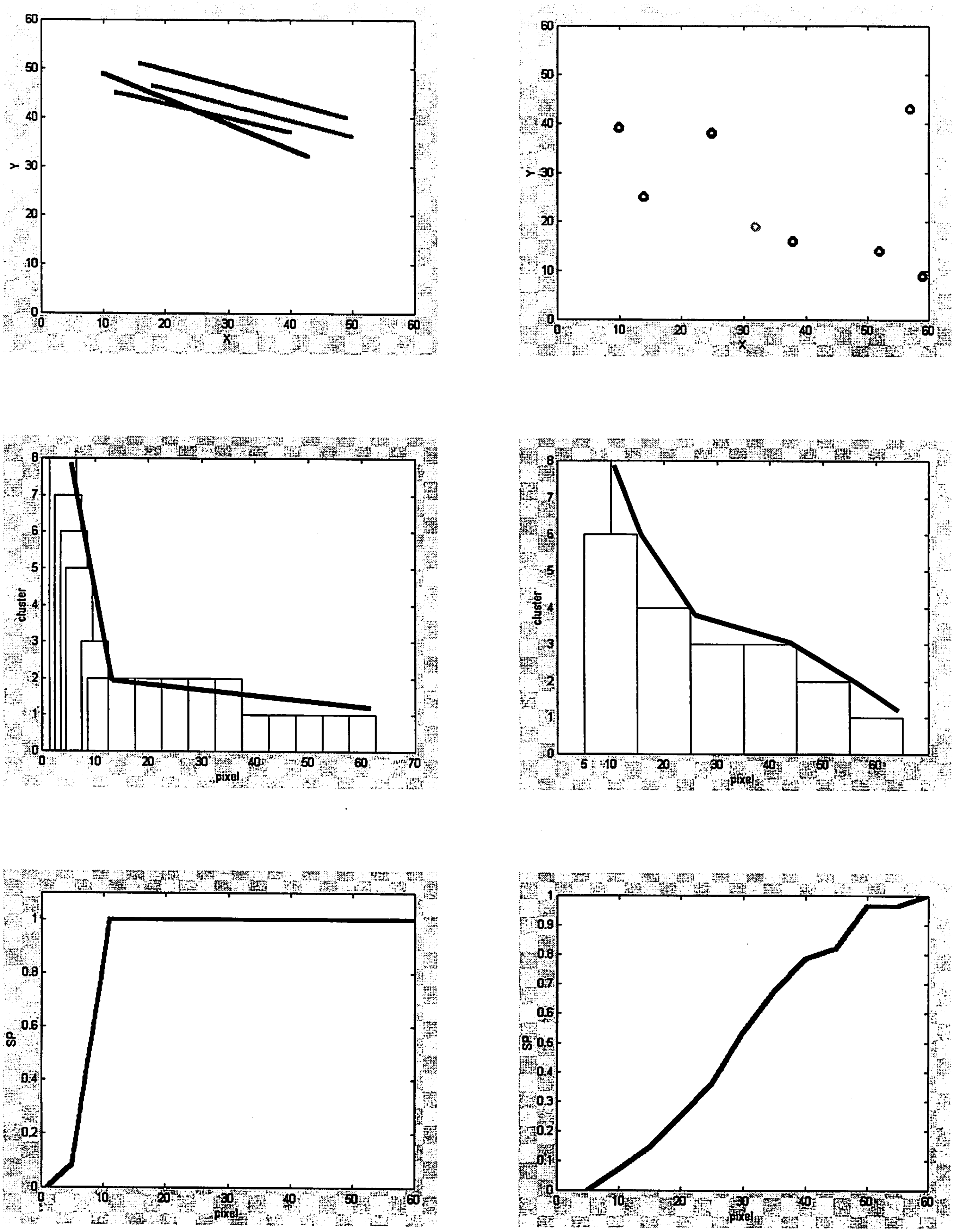


__Figure 9: Simplified, or "toy" diagram of K-means approach__

Synthetic target clicks (upper); number of clusters (lower) and Sp values (middle) and as functions of K-means, "d". Clustered target clicks (left) vs random clicks (right).

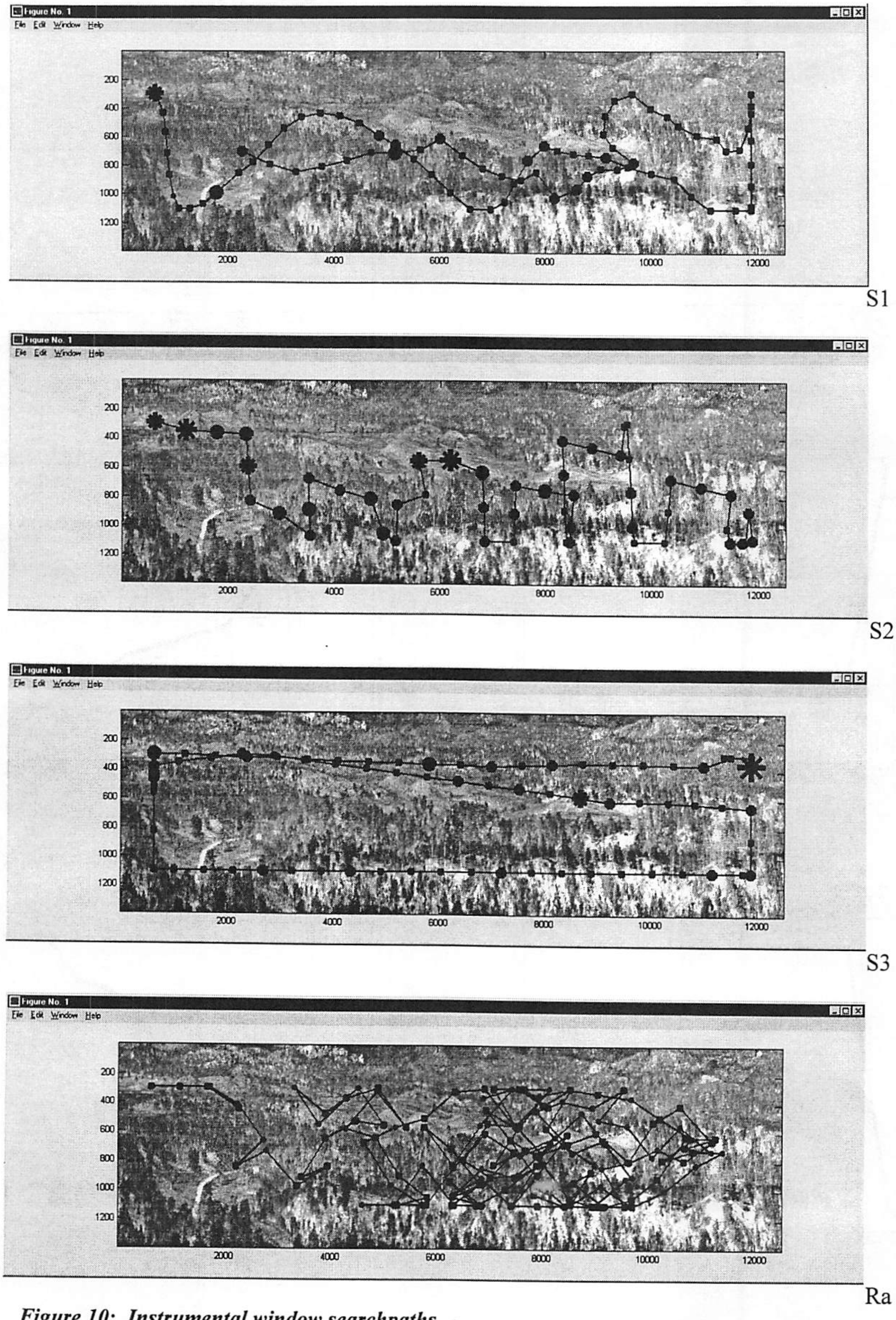


Figure 10: Instrumental window searchpaths

Instrumental searchpaths for three subjects (three upper panels); a random instrumental cover with minor constraints (lowest panel).

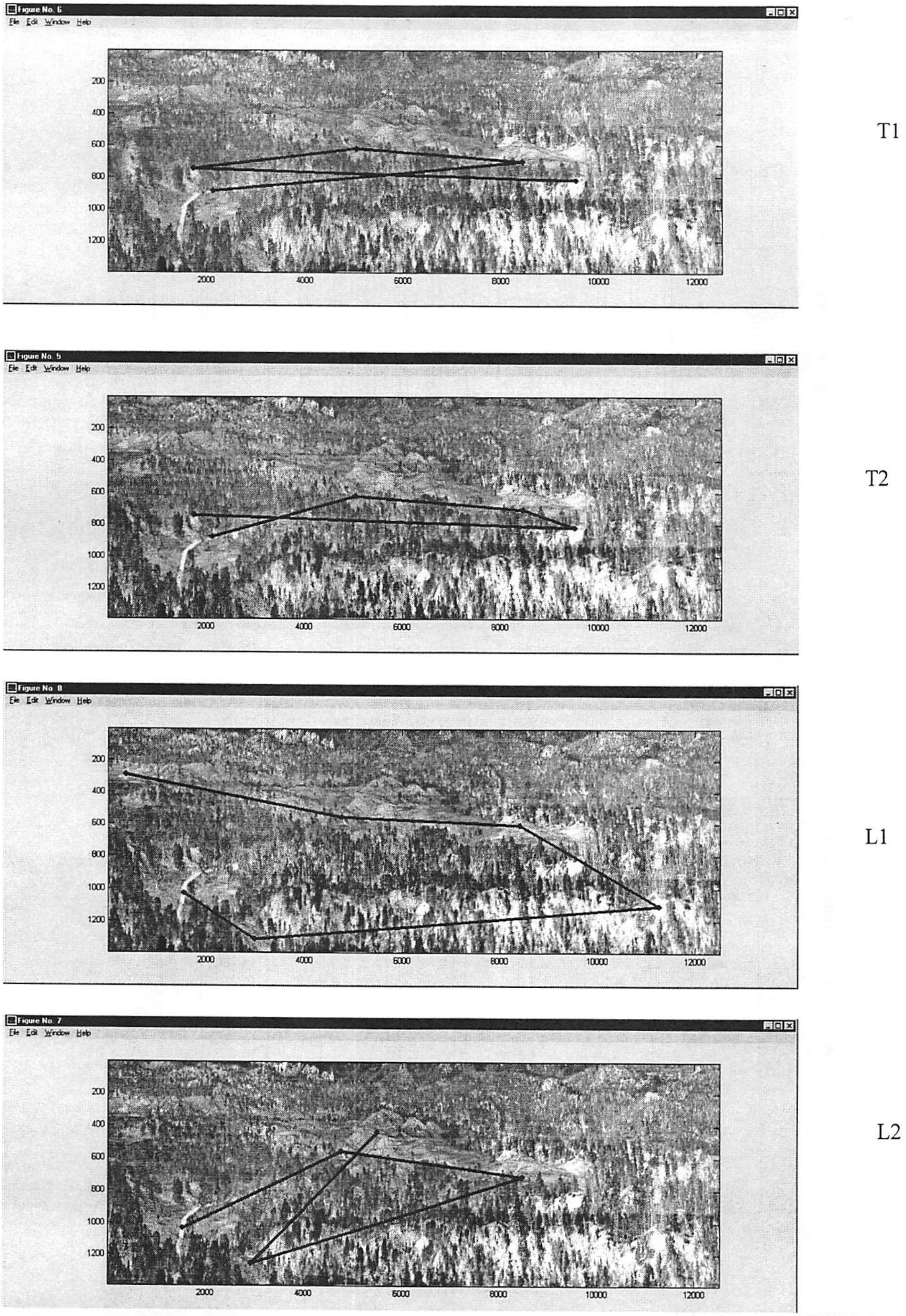
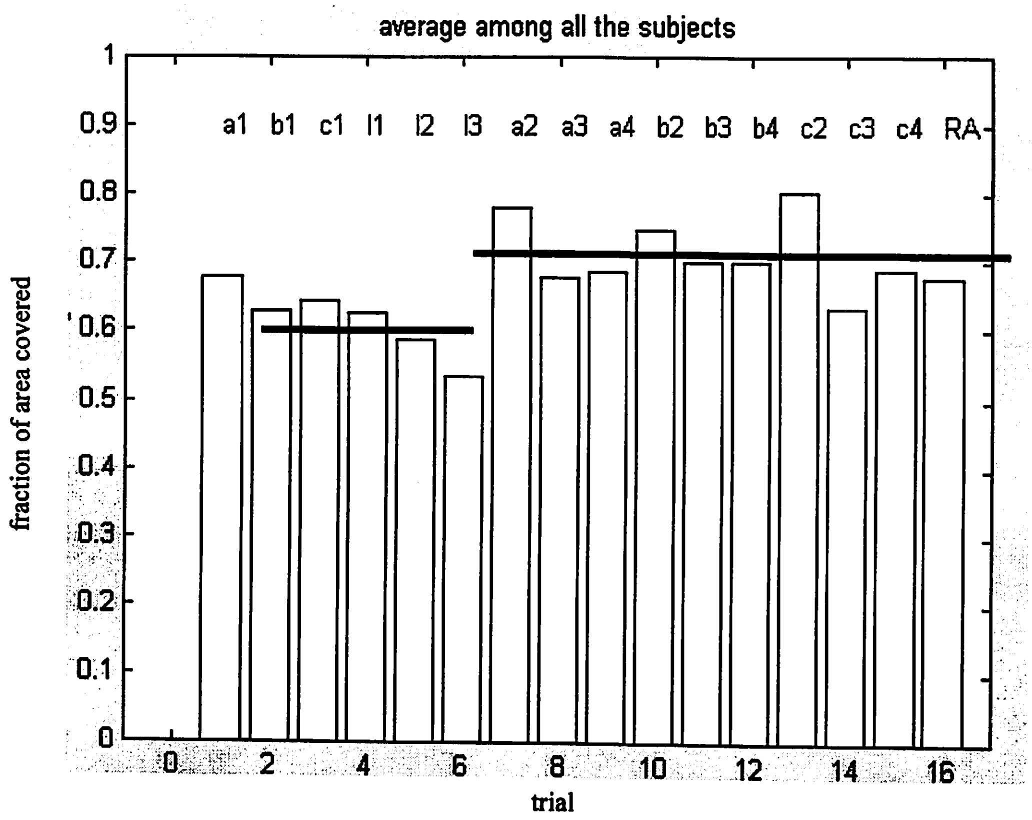


Figure 11: Target and landmark searchpaths

Note differences from instrumental searchpaths.



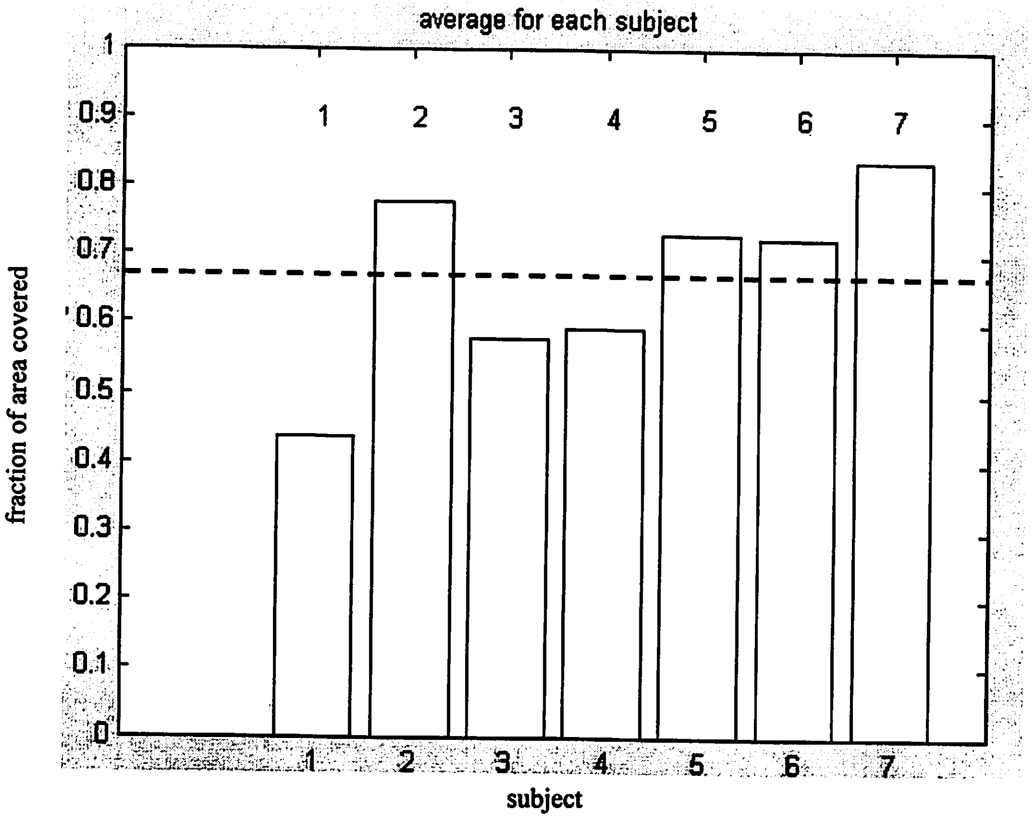
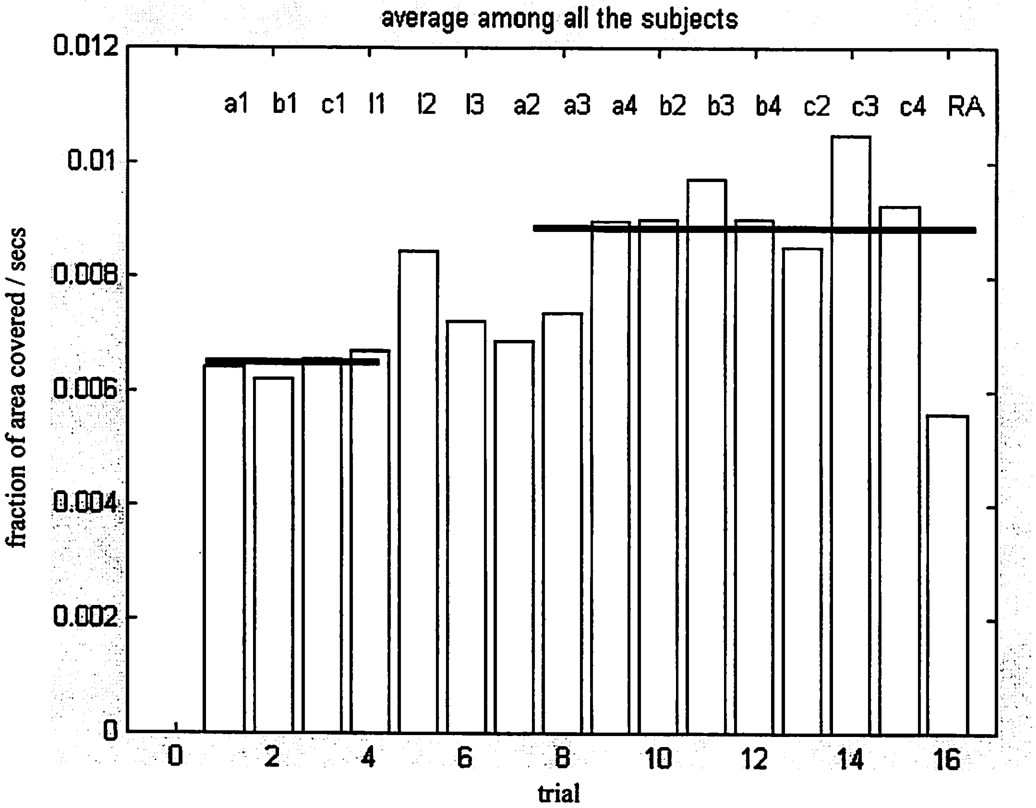


Figure 12: Covered Area



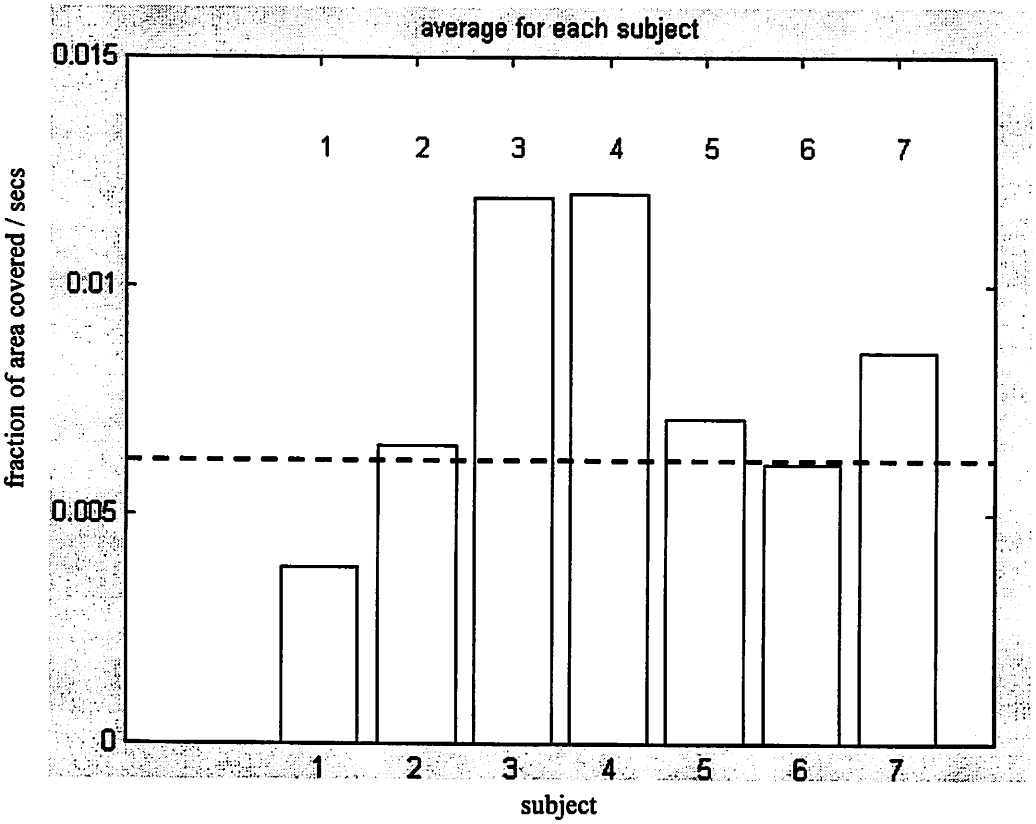


Figure 13: Covered area as a Function of Time

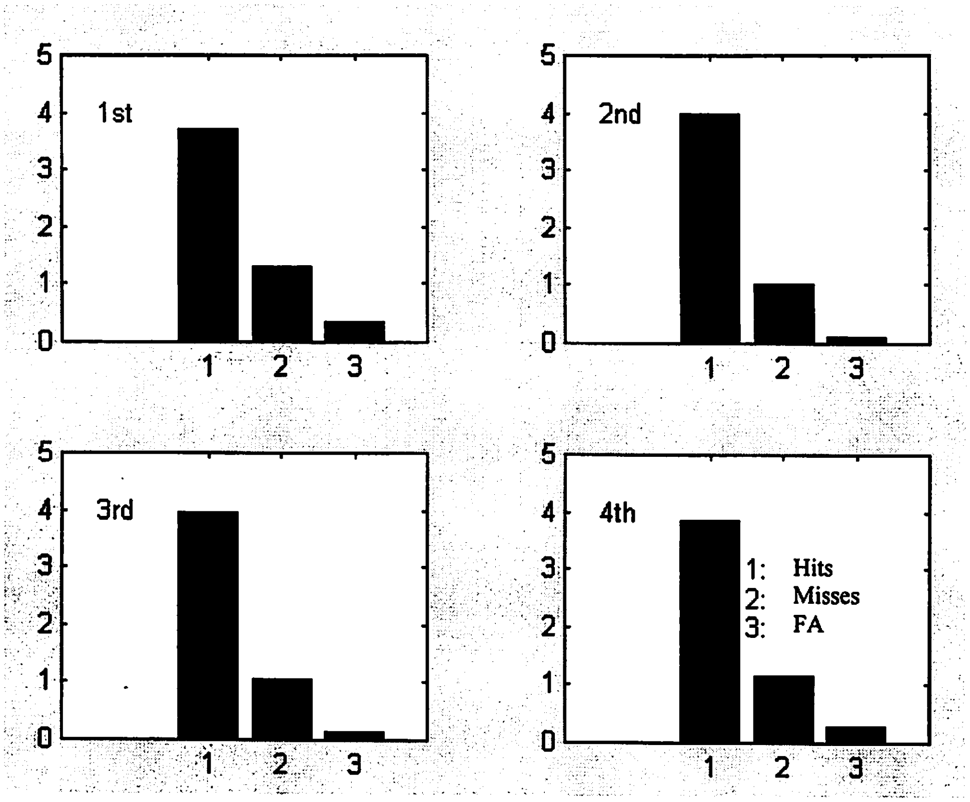


Figure 14: Error analysis of the search process

Three types of errors shown for each of the four sequential FOR presentations (each in its own window). Correct rejections could not be measured in our experiment.

THEORY OF SIGNAL DETECTION, TSD

Guessed

		+	0	
Actual	+	Hits 0.8	Misses 0.2	
	0	FA 0.01	CR 0.99	

Figure 15: TSD error analysis

Hits and misses, as measured in the experiment,. probabilities for false alarm, FA, and correct rejections, CR, adjusted as indicated and justified in the text..

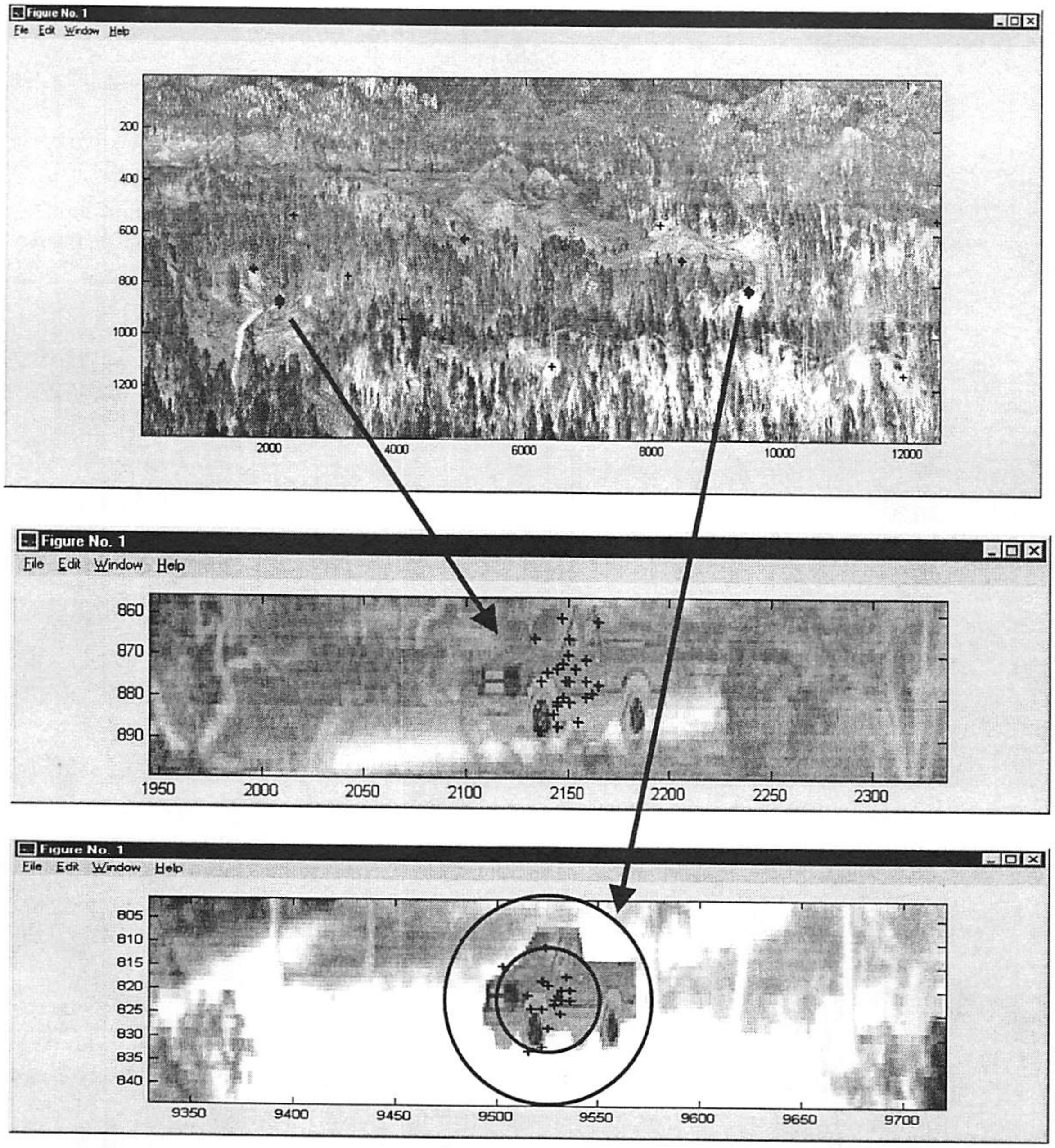


Figure 16: Loci of target acquisitions

Target regions in a portion of FOR (upper) are shown magnified (middle and lower). The dense clustering of all subjects' clicks for the four sequences is typical. Circular dimensions (lower panel) explained in text.

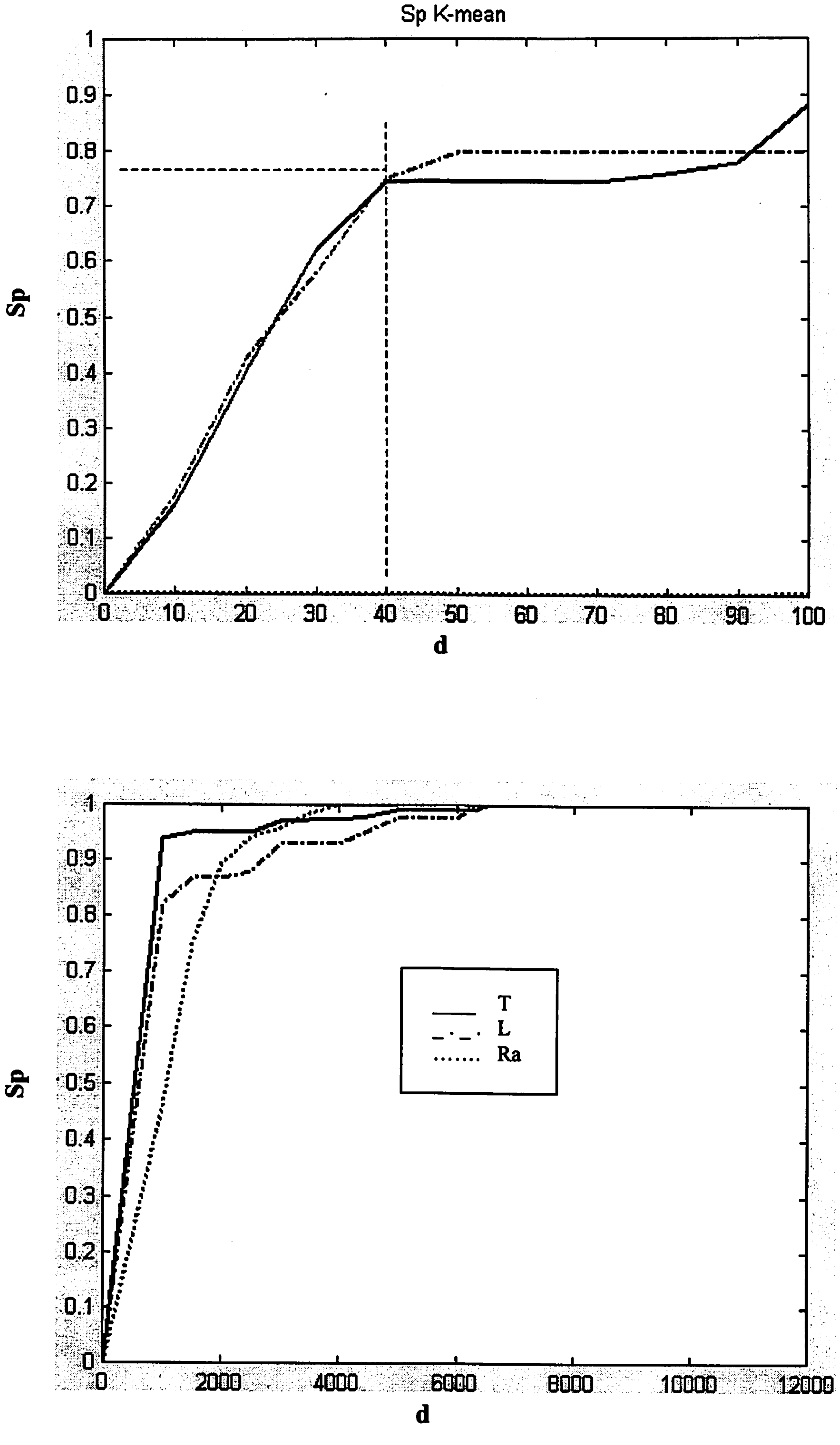
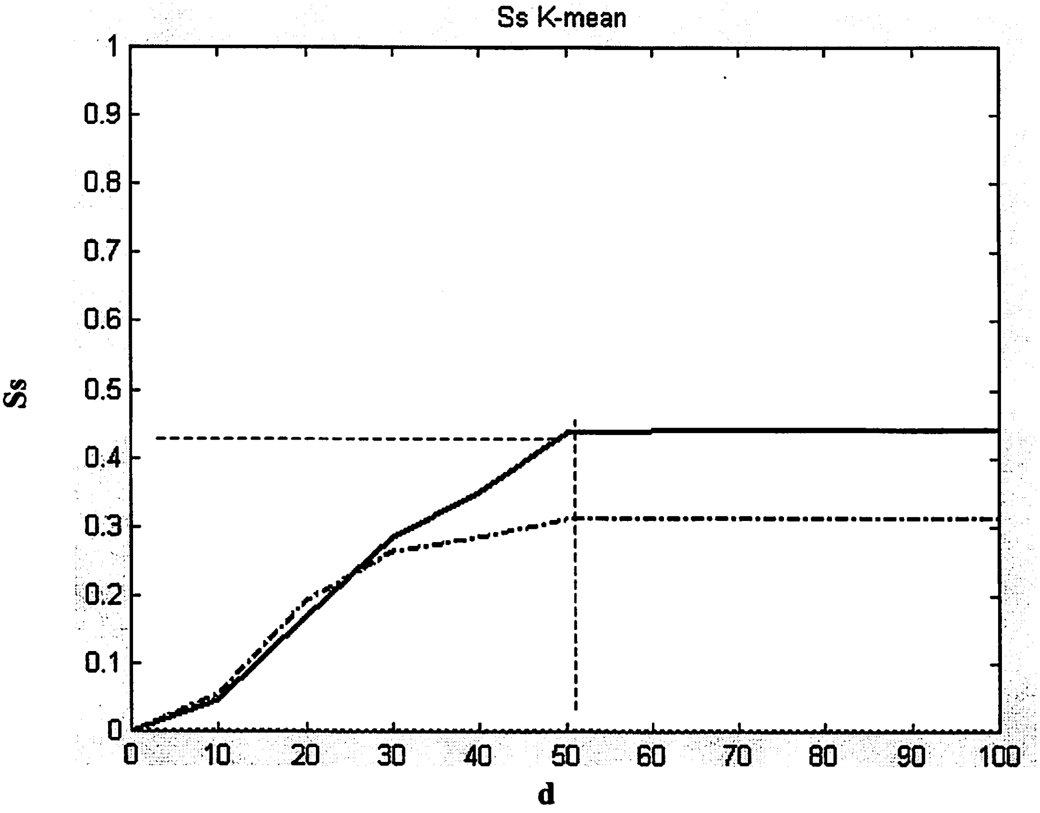


Figure 17: K-means approach to "scale" using Sp

Functions for targets, landmarks, and random all look similar on minified diagram (lower), but important results appear in structural elements of magnified diagram (upper).



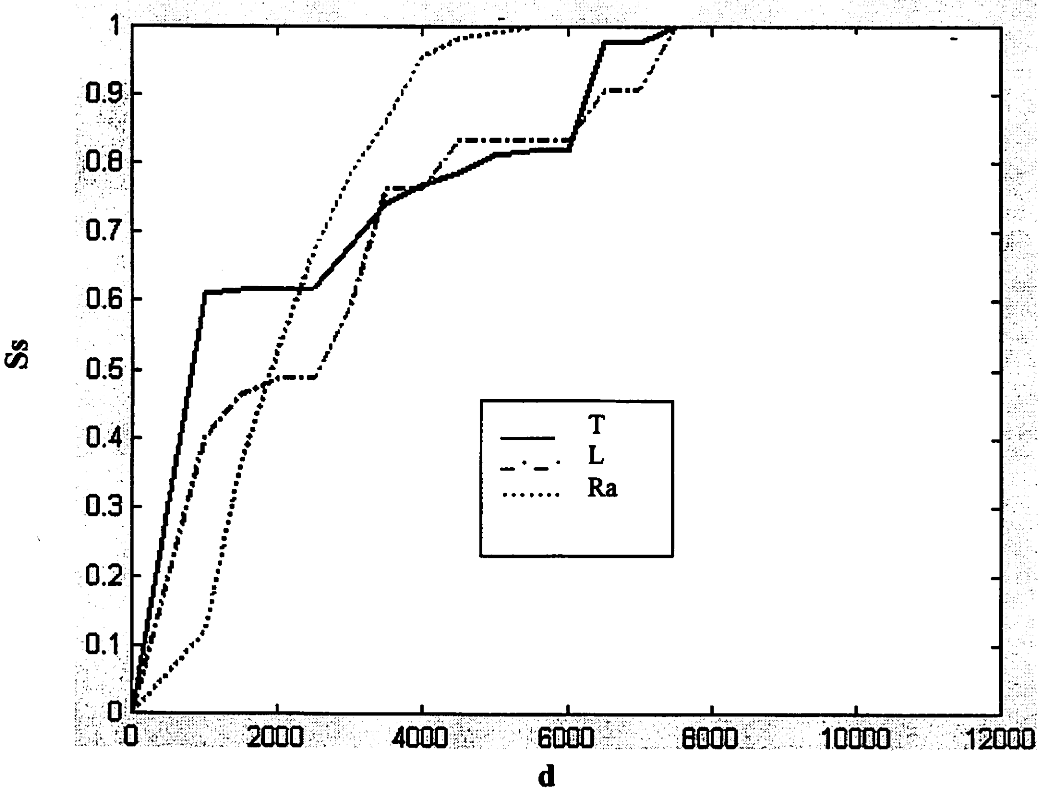
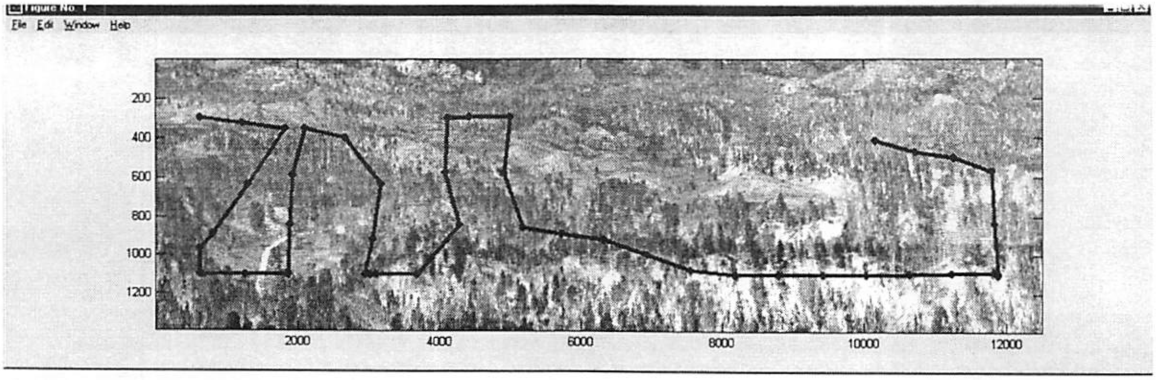
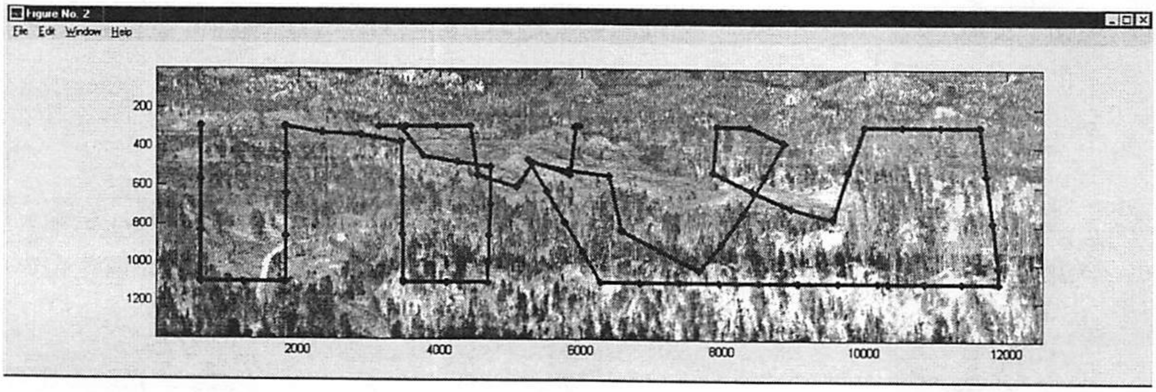
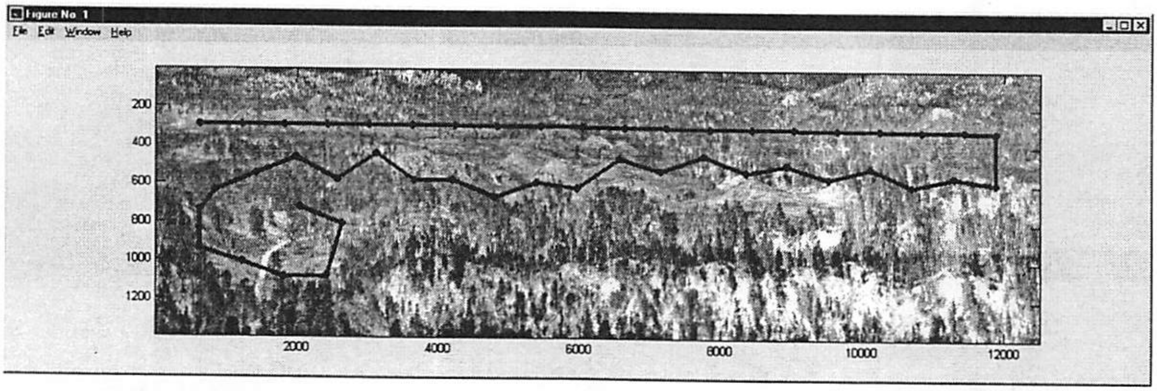


__Figure 18: K-means approach to "scale" (using Ss)__
As in previous figure, but note lower Ss indices and slightly larger "d" for landmarks.







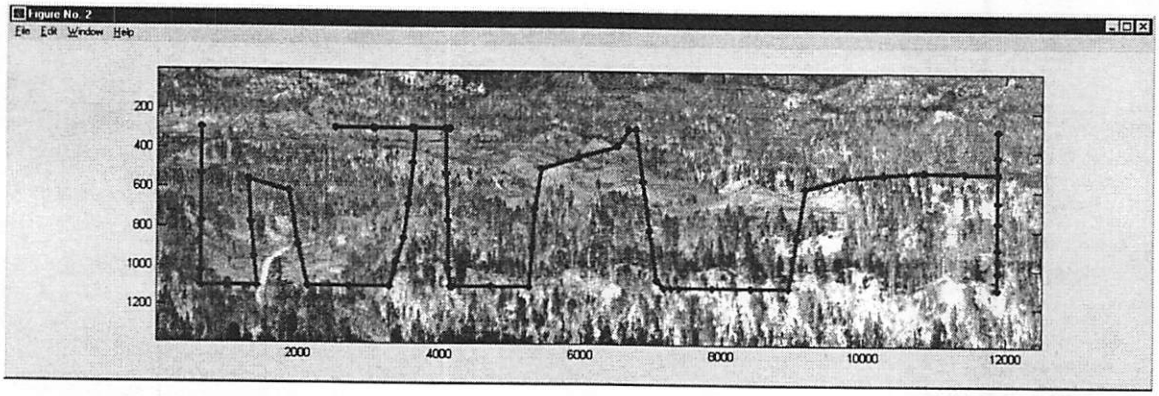
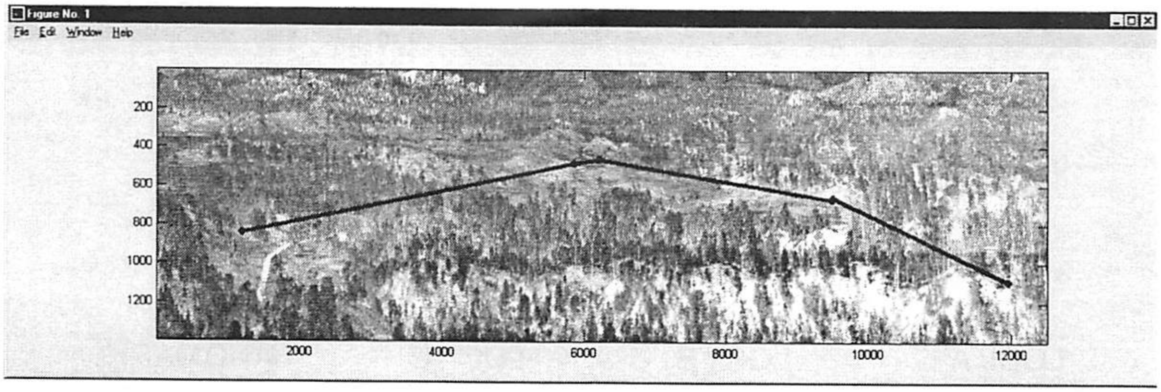
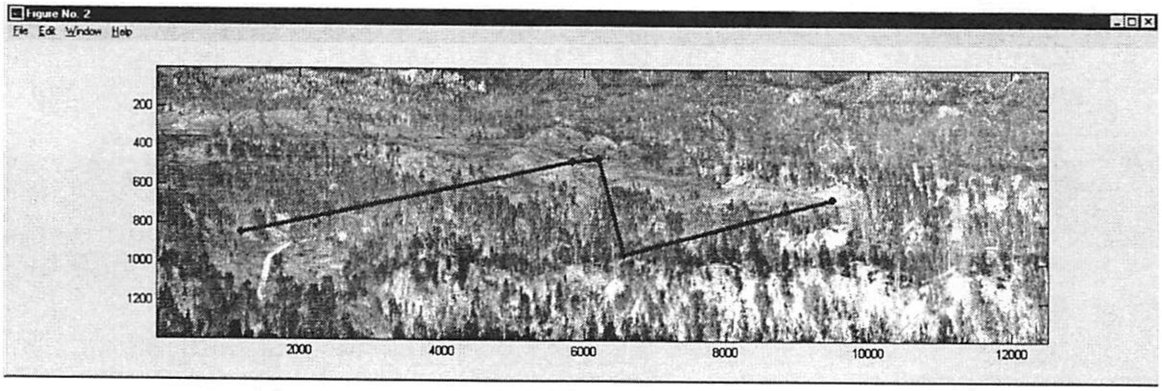


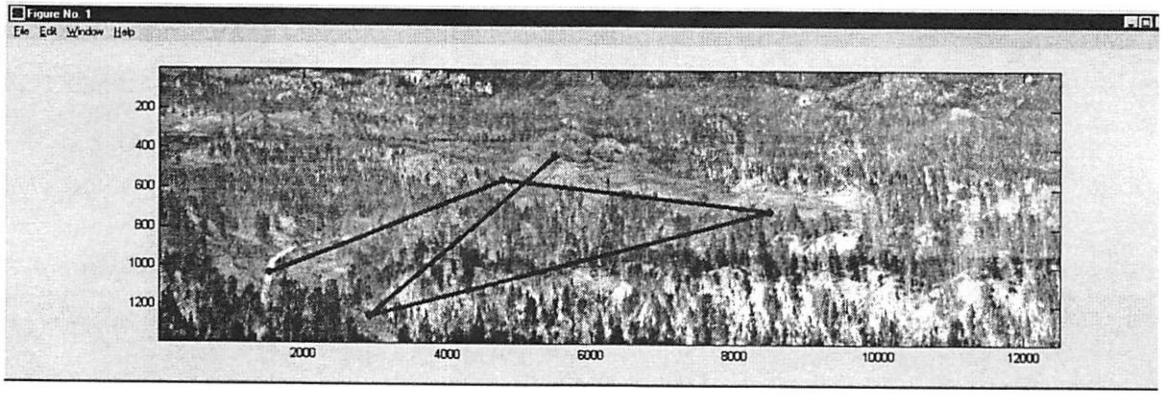
Figure 19: Similar and different subjects' instrumental searchpaths



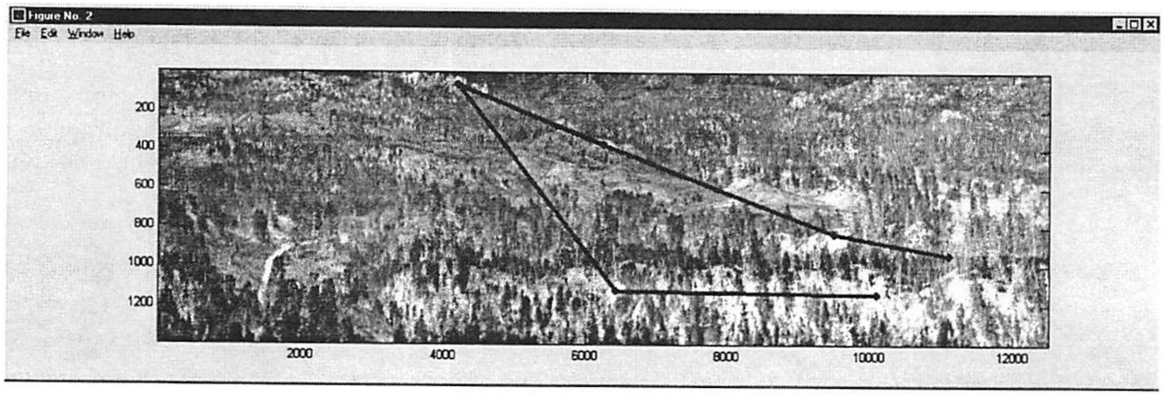
target



target



landmark



landmark

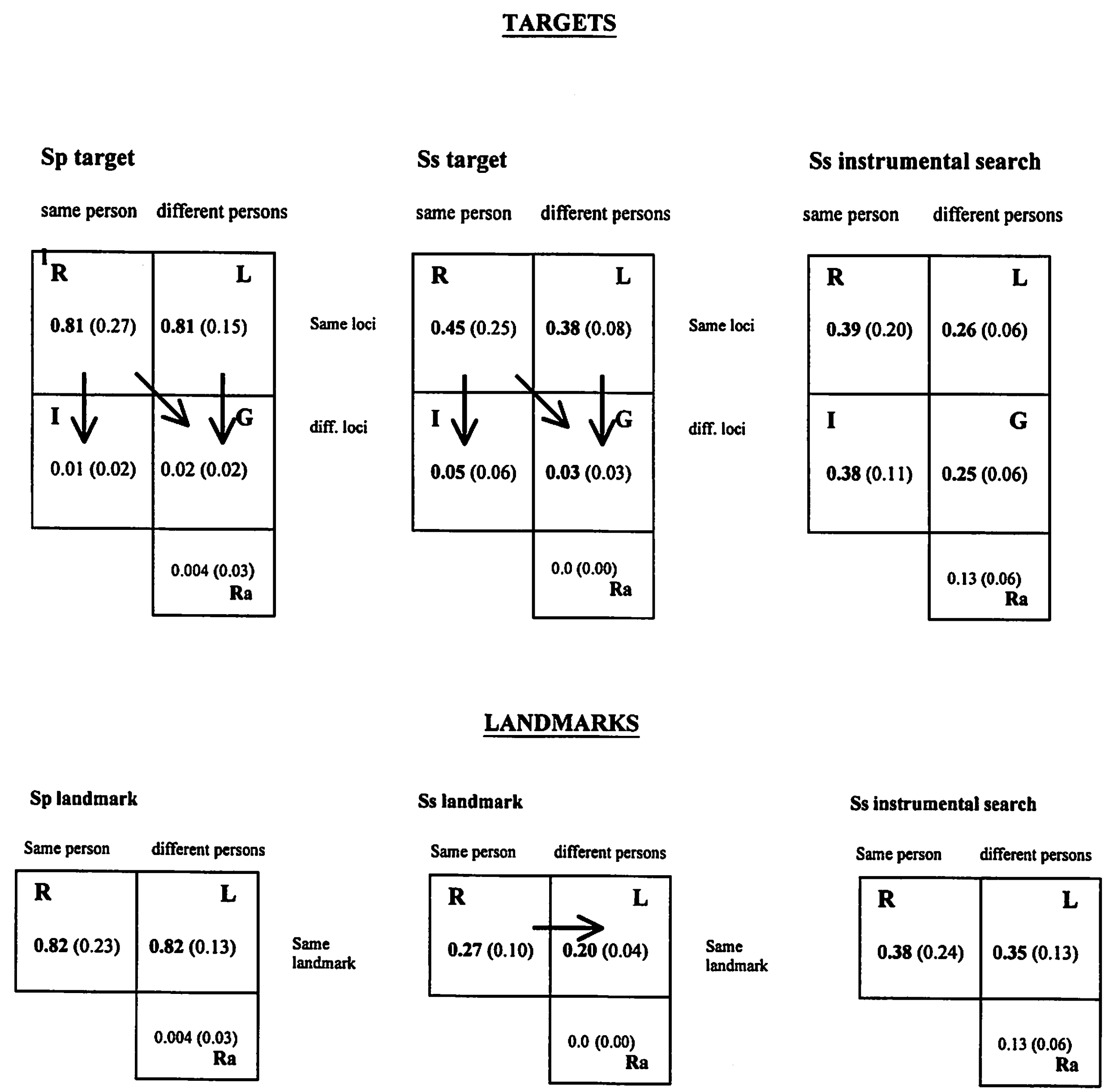
Figure 20: Similar target and different landmark searchpaths

Ss Y-matrix for Instrumental Searchpaths

		LOCI: A				LOCI: B				LOCI: C				LANDMARK		
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
	1	----	.20	.18	.24	.40	.28	.25	.27	.31	.17	.34	.43	.50	.41	.23
	2		----	.43	.42	.29	.22	.25	.42	.45	.34	.43	.35	.36	.42	.43
LOCI: A	3			----	.52	.29	.37	.46	.60	.42	.39	.50	.48	.44	.40	.45
	4				----	.45	.44	.43	.50	.42	.39	.41	.41	.46	.43	.39
	1					----	.41	.35	.31	.29	.31	.23	.27	.43	.51	.41
	2						----	.39	.41	.37	.38	.31	.30	.49	.43	.38
LOCI: B	3							----	.49	.30	.33	.31	.50	.43	.44	.45
	4								----	.42	.38	.57	.64	.43	.37	.42
	1									----	.43	.53	.43	.34	.31	.36
	2										----	.36	.47	.35	.34	.36
LOCI: C	3											----	.49	.39	.45	.49
	4												----	.37	.35	.38
	1													----	.42	.41
LANDMARK	2														----	.54
	3															----

Figure 21 Y-matrix

Actual example of experimental data for Ss for target searchpath similarities. These extensive collections of data are collected and averaged to supply the coefficients of the parsing diagrams (Figure 22). The coefficients are assembled in the Y-matrix in patterns.



TARGETS

Sp target

same person	different persons	
R **0.81** (0.27)	**L** **0.81** (0.15)	Same loci
I 0.01 (0.02)	**G** 0.02 (0.02)	diff. loci
	0.004 (0.03) **Ra**	

Ss target

same person	different persons	
R **0.45** (0.25)	**L** **0.38** (0.08)	Same loci
I **0.05** (0.06)	**G** **0.03** (0.03)	diff. loci
	0.0 (0.00) **Ra**	

Ss instrumental search

same person	different persons
R **0.39** (0.20)	**L** **0.26** (0.06)
I **0.38** (0.11)	**G** **0.25** (0.06)
	0.13 (0.06) **Ra**

LANDMARKS

Sp landmark

Same person	different persons	
R **0.82** (0.23)	**L** **0.82** (0.13)	Same landmark
	0.004 (0.03) **Ra**	

Ss landmark

Same person	different persons	
R **0.27** (0.10)	**L** **0.20** (0.04)	Same landmark
	0.0 (0.00) **Ra**	

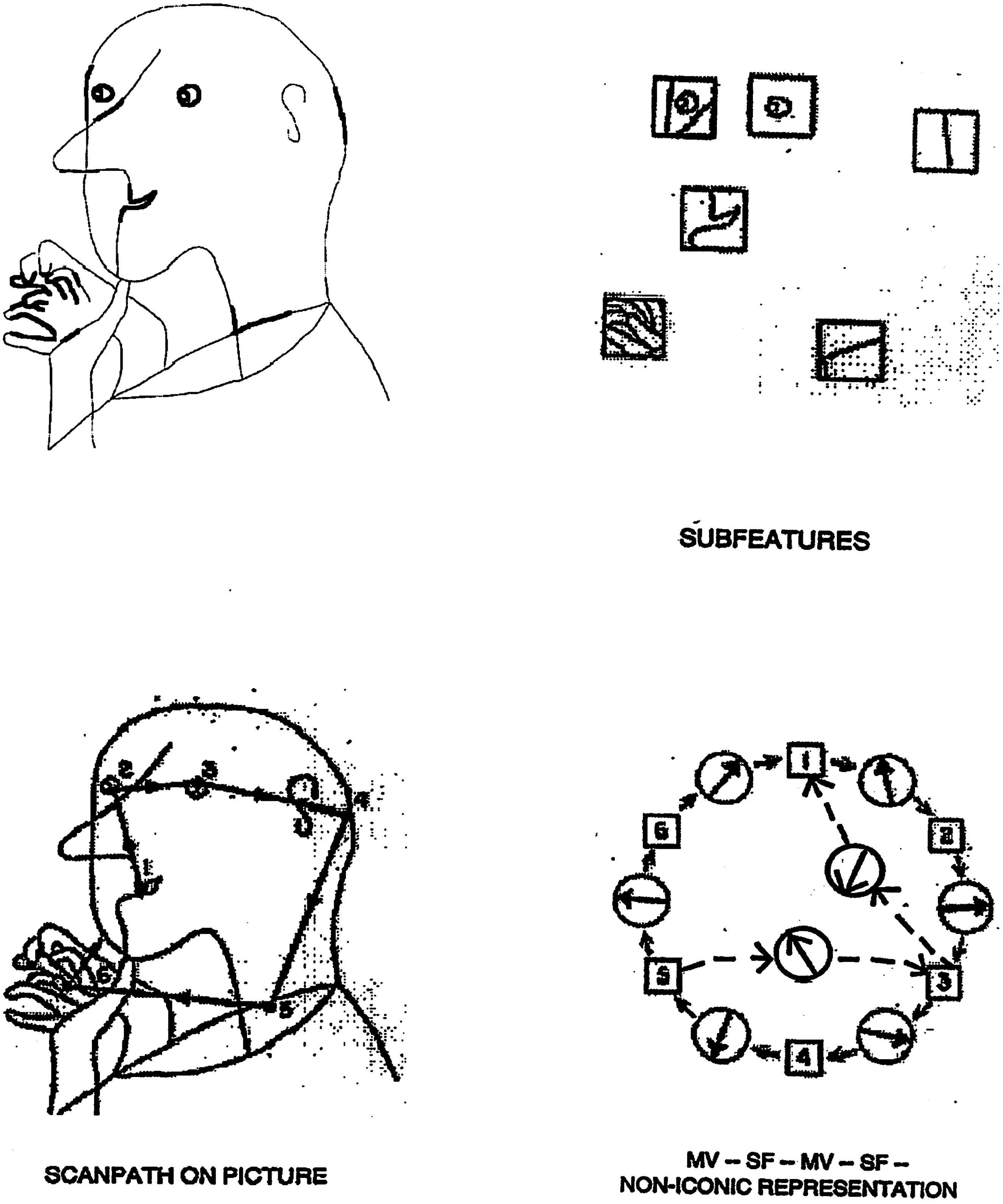
Ss instrumental search

Same person	different persons
R **0.38** (0.24)	**L** **0.35** (0.13)
	0.13 (0.06) **Ra**

Figure 22: Parsing Diagrams

		Sp	Ss
Before familiarization	Instrumental (1 – 4)	0.87	0.33
	Target (1 – 4)	0.79	0.34
	Landmark (1 – 3)	0.82	0.30
After familiarization	Instrumental (3 – 4)	0.90	0.50
	Target (3 – 4)	0.82	0.47
	Landmark (2 – 3)	0.81	0.26
	random	0.06	0.00

Figure 23 : Familiarization or consolidation effect



SUBFEATURES

SCANPATH ON PICTURE

MV -- SF -- MV -- SF --
NON-ICONIC REPRESENTATION

Figure 24: Scanpath Theory

INITIAL VISUAL SEARCH	**FAMILIARIZATION for EFFICIENT VISUAL SEARCH**	**NORMAL SCANPATH VISION**
1.- Pre-scan random search and viewing to ascertain information re boundary conditions for the search task	**1.- 'A priori' knowledge** geometry and orientation figure-background noise and clutter target-decoy MFs S/N; T/Clutter ratios	**1.- Internal spatial-cognitive model** already formed and used for checking
2.- Cover Tactics random search	**2.- Cover Tactics** systematic row search high probability areas	**2.- Cover Tactics** move fixations-foveations directly to predicted loci (parietal lobe)
3.- Detection wide peripheral vision is essential sensitive to effects of noise and clutter	**3.- Detection** special anti-filter developed for noise and clutter	**3.- Detection** not used except to readjust planned eye movement saccades
4.- Recognition (and Identification) careful foveal viewing	**4.- Recognition** (and Identification) matched filters for targets and decoys and for distinguishing features have been formed	**4.- Recognition** (and Identification) sub-features quickly checked against predictions with iconic maps in visual cortex

Figure 25: Bottom – Up and Top – Down Vision

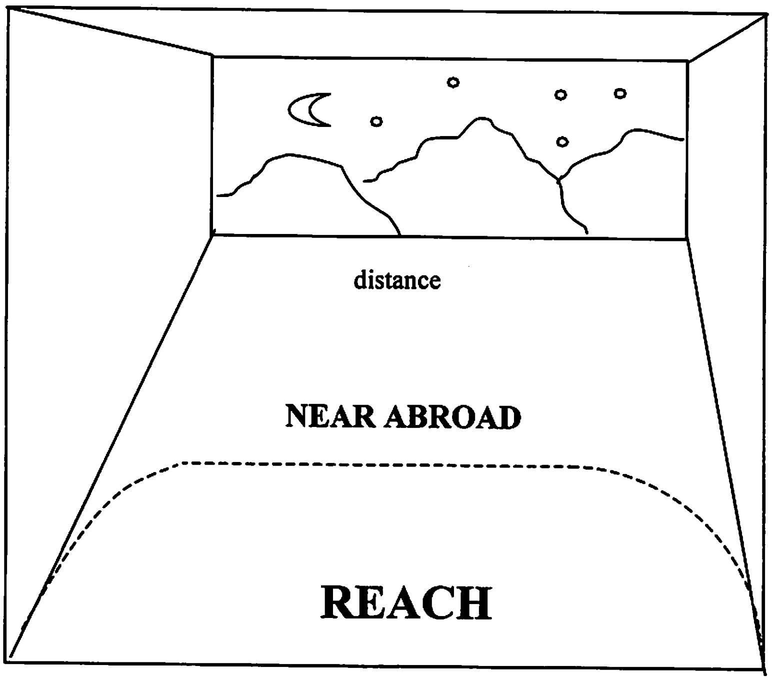


Figure 26: Over – all scales of distance for an abstract scene.