AD-A238 026

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

...timated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, ...nd reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this ...his burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...he Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

...PORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL 01 May 88 to 31 Dec 90

4. TITLE AND SUBTITLE
Topographic Map Reading

5. FUNDING NUMBERS
GR - AFOSR-88-0187
PE - 61102F
PR - 2313
TA - A4

6. AUTHOR(S)
Herbert L. Pick, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Minnesota
Institute of Child Development
51 East River Road
Minneapolis, MN 55455-0345

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR-

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
Building 410
Bolling AFB DC 20332-6448

NL

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Work completed includes a field study and protocol analysis of experienced map readers solving a localization problem. The analysis focused on identifying both the features the map readers attended to and the information processes involved in solving the problem. The information processes identified were reconnaissance, map orientation, feature matching, relation or configuration matching, and hypothesis generation and evaluation. The laboratory simulation studies were conducted on the basis of the field study. They both involved matching a photographic scene to a station point and direction of view on a topographic map. Results of these studies tended to confirm the description of the features and processes from the field study. Two supplementary studies were carried out. One used an alternative laboratory map reading task and provided converging evidence for the effects of manipulation of the amount of information available in the map. The other supplementary study was carried out to determine the precision with which map readers could judge physical distance and slope from photographic scenes and maps. This ability would constrain the possibility of using metric information in map reading. Preliminary specification for a computational architecture for the problem solving aspects of the drop-off problem was completed.

14. SUBJECT TERMS

15. NUMBER OF PAGES
20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	UNLIMITED

GR - AFOSR-88-0187

The model includes both knowledge bases for features and hypothesized solutions as well as a control structure for guiding problem solving activity.

A-1

May 12 1991

TOPOGRAPHIC MAP READING

(AFOSR-88-0187)

Final Report

AFOSR-TR- 91 0590

Submitted by:

Herbert L. Pick Jr. & William B. Thompson
Principle Investigators

Introduction.

The work supported by this grant was directed at understanding how people use topographic maps to solve localization and navigation problems. On the basis of this analysis work was begun on developing a computational model of these processes. The major part of the research was carried out in the context of a localization problem. The localization problem is most generally conceived of as matching a scene of the terrain to a position on a topographic map, that is a map-terrain correspondence problem. Localization of the map-terrain correspondence sort may be characterized as lying along a dimension defined by amount of initial information available. At one extreme is an updating problem. With updating one starts off with specific knowledge of where one is on the map, that is of the map-terrain correspondence at an initial time. After movement for some interval time it is necessary to find one's position on the map again. At the other extreme is the "drop-off" problem. One has minimal information about where one is on the map at the initial time and the task is to establish that map-terrain correspondence. The research completed during the grant period was oriented towards understanding the drop-off problem.

The research involved three studies of map reading per se. The first study was a protocol analysis of experienced map readers' verbal reports while they were solving a drop-off problem in the field. Data from the first study suggested which map and terrain features are used in solving the drop-off problem and what kinds of information processing map readers engage in. The second study used a laboratory simulation of a localization problem to vary the amount and type of map information available to map readers to further assess the features of importance in solving localization problems. The second study involved matching positions and directions of view on a map to photographic scenes. In the third study verbal protocols of map readers solving the laboratory simulation problems were analyzed in order to investigate further both the features of importance and the information processes involved in the problem solving. In order to further investigate the terrain features map readers might attend to a scene memory task was conducted in which map readers and non-map readers recall memory for photographic scenes was compared.

One supplementary study was carried out using an alternative laboratory map reading task in which the location of the photographic scene was provided on the map and the map reader's task was to find where the scene was being viewed from. A second supplementary study was carried out to determine the precision with which map readers could judge physical distance and slope from photographic scenes and maps.

Preliminary specification of a computational architecture for the problem solving aspects of the drop-off problem was completed. The model includes a taxonomy knowledge base for aiding in recognition of topographic features and the assembly of configurations, and a hypothesis knowledge base for posting information on currently active hypotheses about viewpoint or map-terrain correspondences. A set of procedures forms a control structure for recognizing features and posting and evaluating hypotheses.

The first section of the body of this report will summarize the main results of the studies. This will be followed by a more detailed exposition of the experimental procedures and results.

Summary of Results

Map reading studies.

Three studies have been conducted specifically on reading topographic maps. The first involved experienced map readers solving a map-terrain correspondence problem. The task was a "drop-off" localization problem. This is a problem in which map readers were asked to find their current position on a map when they had a minimum of apriori information. That is, they had to rely primarily on the information available to them from their current view of the terrain and had to relate it to the information available to them from the topographic map. The map readers were asked to think aloud as they engaged in the problem solving process and these verbal protocols constituted a major part of the experimental data. The second study consisted of a laboratory simulation of the correspondence problem in which photographic scenes of terrain were to be related to specified positions and direction of view on a topographic map. The amount of map information available to the map readers in this study was varied by masking portions of the map around the specified positions. The third study employed a subset of the simulation problems used in the second study and again involved collection of verbal protocols of map readers solving these problems.

DROP-OFF LOCALIZATION PROBLEM. Results of success in solving the drop-off localization problem indicate that this type of problem, with a minimum of apriori information about initial position, is exceedingly difficult. Under one condition the map readers were not allowed to move away from their initial station point. Under these conditions no one was able to arrive at a successful solution. Under a second condition in which exploration was permitted fifty per cent of the map readers successfully solved the problem.

The verbal protocols provided information about the kinds of features and attributes the map readers attended to and the kinds of information processing strategies they engaged in. The features included: valleys, saddles, rivers, ridges, plateaus, lowlands, hills, and fields, etc., with attributes such as gradients, distances, contours, etc. Of particular interest was attention to relations among features or configurations. These relationships include purely topological descriptions (e.g. behind, in front of, next to) and occasionally quantitative properties such as actual elevation. The information processes identified were reconnaissance, map orientation, feature matching, relation or configuration matching, and hypothesis generation and evaluation.

The map readers typically begin their problem solving with a general reconnaissance, looking broadly across terrain and/or map ostensibly to get a general feel for what is there. The more successful problem solvers focus their reconnaissance on the terrain identifying features and relations among the features. Map orientation refers to aligning the direction of the map with the terrain which all map readers sooner or later do. While this isn't logically required by the nature of the problem it facilitates the search for correspondences between terrain and map. All explicit direction information has been removed from the maps used in these studies so that this map orientation is typically accomplished on the basis of correspondences between directions of drainage in terrain and map. Feature matching involves finding a correspondence between a feature in the terrain and on the map. It typically precedes hypotheses about where one's own position is on the map. Since simple identification of a feature such as hill or valley is not likely to be distinguishing, features are usually qualified with attributes such as relative size, elevation, slope or gradient, etc. These tend to be bipolar or qualitative descriptors: e.g., large or small, narrow or wide, steep or shallow. Relation or configuration matching serves the same purpose as feature matching and simply involves relations among features. However, configurations constrain the matching process more effectively than do individual features, even when they are qualified by specific attributes. Most configurations that are identified involve features that are actually physically contiguous rather than accidentally adjacent in the view of the terrain. In addition most configurations are composed of features which fall along or are parallel to a line-of-sight. Hypothesis generation is the positing of a distinct map location as corresponding to the viewing position. While viewpoint invariance is desirable in the spatial arrangements of features which define a configuration, viewpoint dependence is obviously necessary for hypothesis testing. The hypothesis must necessarily describe the relationship of topographic features to the viewpoint. This is done in rather simple, qualitative descriptions rather than using a more sophisticated trigonometric analysis. Evaluation of hypotheses is carried out in a variety of ways. In testing hypotheses the more successful map readers use a simultaneous comparative approach rather than one involving sequential generation and testing. A common source of error is to "explain away" disconfirming evidence. This involves discounting an expectation from the map based on the hypothesized position. In the present study the correct hypothesis often failed to be generated because of inadequate registration of the terrain in the immediate surround of the station point. The greatest gain for map readers allowed to explore was in more accurate registration of the area around the station point.

LABORATORY SIMULATION OF LOCALIZATION PROBLEM. In the second study, the laboratory simulation of the map-terrain correspondence problem, map readers attempted to identify a match between a station point and direction of view specified on a topographic map and a photographic scene. Different portions of map around the station point were masked. Correct identification of matches was best in the unmasked control condition (59%). However, accuracy was almost as high (56%) in the case in which the outer 1/3 of the map was masked. Accuracy was low both in a condition in which the inner 1/3 of the map was masked and when the outer 2/3 was masked. This pattern of results suggests it is not the sheer amount of area masked which reduces accuracy but the location of that area. Accuracy is low when the area right around the station point is occluded or when the largest amount of masking occurs. That condition eliminates the information both at the far and intermediate distances. Overall the near and intermediate distance information seems most crucial in this task.

Examination of performance for individual photographic scenes indicates that there are idiosyncrasies depending on particular locations. Thus reliance on near information in some cases led to performance well below chance level when the perceived layout around the station point from the photograph is in error.

PROTOCOL ANALYSIS OF LABORATORY PROBLEMS. To further explore these observations the third map reading study was carried out in which map readers provided verbal protocols by thinking aloud while they worked on a subset of the problems of the second study. The quantitative results confirmed the conclusions of the second study as far as accuracy in relation to degree of masking. The verbal protocols supported the inferences from the second study about the importance of particular features in leading to errors when nearby terrain in the photographic scene was misperceived. In addition, analysis of the protocols indicated the importance of configurations and relations among features for confirming hypotheses. As in the drop-off problem in the field configurations more tightly constrain the matching process than do individual features.

Scene memory study.

One goal of the map reading studies was to determine the features map readers rely on in solving map-terrain correspondence problems. To further specify these features a study of scene memory was conducted using photographic scenes from the same areas as those used in the map reading studies. Memory for photographic scenes was compared for map readers viewing photographs with the subsequent task in mind of solving a map-scene correspondence problem, for map readers simply asked to look at the photographic scenes, and for non-map readers.

Results indicate that all three groups recall approximately the same number of features. However, the type distribution of the features differ across the three groups. Thus, for example, the map readers with the map task in mind recall more terrain features and fewer vegetation features than do the non-readers. The map readers who simply view the scenes fall in between the other two groups.

The type of terrain features mentioned included hills, flat areas, valleys, and slopes. The proportion of these differed across groups. The map readers with task in mind recalled more valleys and slopes and fewer flat areas than did non-readers. Again the map readers just viewing the scenes fall in between the other two groups.

Experimental Methodology and Discussion of Results

Map reading studies.

DROP-OFF LOCALIZATION PROBLEM. Twenty-nine experienced map readers participated in the map reading drop-off field study. They were recruited from geology and geography departments, orienteering clubs and other outdoor and wilderness organizations. Their experience ranged from professionals who use topographic maps daily on the job to experienced recreational users including a national class orienteer and a person who hiked the length of the Brooks mountain range in Alaska.

Conditions for the drop-off problem were achieved by taking blindfolded map readers to a station point in a state park about an hour's ride away. Participants were led across a field and up a hill to the station point about 300 meters from a road access. No cultural features were visible in the terrain from the station point (and as noted none were included on the map). See Figure 1. At the station point the blindfold

Insert Figure 1. Here

was removed and the participants were given a map on a clipboard and were asked to find their position. They had been briefed during the trip to think aloud as they tried to solve the task. The verbal reports of a subset of 17 of the map readers were taped as they thought aloud while attempting to solve the localization problem. Their behavior was simultaneously videotaped to provide information about where they were looking and while they were thinking aloud. The verbal reports were transcribed and coordinated with the videotape to determine which features in the terrain and on the map were being referred to. The transcripts were parsed into statements at breathing pauses or at changes of focus of attention (e.g., from map to terrain). Each statement is a coherent utterance that makes a single point. A series of statements about features in a single viewing direction, e.g. northwest, form episodes of problem solving.

The coding scheme developed to analyze the protocols consists of an Information Trace and a Process Trace. The Information Trace documents the time course of attention to terrain and map features while the Process Trace documents the time course of cognitive processing about these features. The time course is indicated by sequential statements along a horizontal axis and the features and processes are categories along a vertical axis.

Overall the types of features and configurations of features used by the present subjects are probably constrained by the local topography. However, it is interesting to note that most feature description was qualitative rather than metric. Especially judgments of slope gradients were made in terms like steep, medium, shallow etc., rather than degrees. Quantitative judgments of distance were more frequent but still not heavily used. When they did occur they sometimes were in units of time, e.g. a 5-minute-walk, etc. It is also the case that there was little reference to the most distal features of the layout. This might have been partially due to the range of the map but that wouldn't account for the heavy neglect of such features.

Neither in the scene nor map descriptions and hypothesis testing were there any statements that could be characterized as reflecting global visualization. The descriptions specified features or at most configurations of features. This was somewhat surprising as informal reports in early pilot interviews of informant map readers included statements about looking at a map and visualizing the general overall topography.

In trying to summarize the information processing reflected by the protocol analysis it may be useful to think in terms of a focus on the map or the terrain and in either case attention to the station point and its immediate surrounds or attention to the more distal features and layout. In terms of such a two-by-two classification (map vs. terrain and station point vs. distal layout) the goal of the task is to arrive at a solution which specifies a station point on the map. Two general strategies are observed: a map driven strategy and a scene driven strategy. The more successful subjects appear to use the latter. The reconnaissance of the terrain informed the reconnaissance of the map; their feature matching was guided primarily by inspection of the terrain. This strategy is not surprising since the terrain imposes more constraints on hypotheses than the map. Everything visible in the terrain (subject to the criteria of feature and scale for representation on the map) is relevant. The information on the map is not constrained by what is visible from the station point and hence includes much more.

LABORATORY SIMULATION OF LOCALIZATION PROBLEM. In the laboratory simulation of the map-terrain correspondence problem map readers were asked to match a photographic scene with a direction line on a map. The first aim of the laboratory task was to investigate what amount of information people need to match a map with a scene. One specific question was whether performance is directly related to the proportion of visible map area. A second related aim was to determine whether, independently of size, certain areas of the map are in general more informative than others in solving correspondence problems. Specifically, where is the most useful map information for solving a localization problem, close to the station point or at intermediate or far distances.

To investigate such questions four groups of subjects were asked to match photographic scenes to a position and direction line on a topographic map. Map information was manipulated by masking portions of the map to varying degrees for different groups of subjects. One group of subjects was presented with full or unmasked maps, while the other three groups of subjects were given maps with various portions masked. In the "inner 1/3" masked condition, an area defined by the third of the radius of the map directly surrounding the central station point was occluded. In the "outer 1/3" masked condition, the more distal third of the map's radius was masked leaving a central area corresponding to two thirds of the radius unmasked. Finally, in the "outer 2/3" masking condition, the distal two thirds of the radius was masked leaving only a small central area directly surrounding the station point unmasked. As a consequence, the "inner 1/3" and "outer 1/3" conditions were equivalent in terms of the radius proportion masked, while the "outer 2/3" masking condition had the smallest amount of visible area.

Sixty-three map readers participated in this study. They included geology graduate students, back packers, orienteers or members of the military recruited on campus and in local outing clubs.

The map readers were presented with topographic maps of five locations. Three of those locations were in Minnesota, one was in New Mexico, and one in Arizona. The maps were copies and enlargements of USGS topographic map overlays which did not include any cultural information.

Color slides were taken from a position corresponding to the center of each of those maps. The pictures were taken with a tripod leveled with horizontal. The complete set of pictures covered the whole 360 degree panoramic view. From this set, three non-overlapping pictures at each location were selected for the experiment. The slides were presented to the subject on a rear projection screen in a darkened laboratory room.

The localization task consisted of two symmetrical types of problems. In one problem type, the map reader was given a map on which one arrow was drawn, pointing away from the center. He or she was shown three successive slides corresponding to non-adjacent views taken from this center location. The task was to select the slide corresponding to the view in the direction indicated by the arrow on the map. Since the required response was the selection of a scene, this task is referred to as the "scene task". The map reader could cycle back and forth between the scenes as much as necessary.

For the second problem type three arrows separated by 120 degrees pointed away from the center location on the maps given to the readers. They were shown a single slide for each location and told that one of the three arrows corresponded to the viewing direction of the middle of the picture. Since their task was to select one of the three directions on the map, this task was referred to as the "map task".

Overall results indicate that on the average, accuracy significantly exceeded chance performance in all masking conditions, though not always at each location. Average performance in the full map and the outer 1/3 masked conditions was equivalent and significantly better than performance in the inner 1/3 and outer 2/3 masked conditions, which were also equivalent. This pattern of accuracy suggests that masking areas of the map impeded the solution of the correspondence problems only when those areas were close to the viewers' locations on the map or when large areas of the map were masked. Detailed results are summarized in Table 1.

Insert Table 1. Here

PROTOCOL ANALYSIS OF LABORATORY PROBLEMS. The laboratory map-terrain correspondence task is most valuable for the hints it provides as to the specific information used to solve the problems. Verification of those hints can be obtained by presenting such problems to map readers who describe aloud how they are going about the solutions. Such verbal protocols were collected for a subset of the original problems of the laboratory correspondence task in the third study.

Five problems were selected from the original set: three map problems and two scene problems. Ten map readers attempted to solve the problems twice, first in a masked condition and then in the unmasked full map condition. As in the field correspondence study the participants were asked to think aloud while solving the problems. Overall results indicated 44% successful solutions in the masked map condition and 64% successful solutions in the full map condition. These values approximate the average performance of the corresponding conditions in the original laboratory task. The protocols were scored in a manner similar to the field protocols. The main scoring category which did not apply the same way with the laboratory and field protocols was hypothesis generation. In the laboratory task the hypotheses were, in a sense, given and the task was to test them.

The protocols help explain the particular patterns of results obtained for the different problems. In addition they also illustrate a number of features that frequently occur in the problem solving of map readers in the field as well as in the laboratory simulation. One aspect is a tendency to focus on particular salient features. Even where the focus is on a salient feature such as a large river valley or particular hill, a second aspect involves attempting to find more reliable configurations or combinations of features. As mentioned before, a common source of error is incorrect registration of the area close to the viewpoint. It is also the case that metric information is often ignored. However, ordinal information about the relative heights of features or magnitude of distances may be sufficient to decide between hypotheses. Finally in testing hypotheses, especially in the laboratory, map readers realize that detection of one clear difference between an hypothesized position and what is visible in the terrain is sufficient to eliminate that hypothesis. However, acceptance of an hypothesis usually requires more converging evidence and that is one result of attending to configurations or relations.

SUPPLEMENTARY STUDY: VIEWPOINT LOCALIZATION. As a way of providing converging data for the laboratory simulation masking study described above and alternative "viewpoint" paradigm was developed. Instead of giving a station point on a map and and asking which of three viewing directions matched a scene, a target location was marked on a map and a scene was presented of that target location. The task was to find the position on the map from which the scene was being viewed. The amount of map information was varied. Each trial started with with only a small area of the map around the target visible. A judgment was made as to the location of the viewpoint. Then the visible map area was enlarged, and another judgment made and so on for a total of four steps of increasing map area. Accuracy of finding location was measured in terms of angle and distance from the true viewpoint. As might be expected accuracy improved with increasing amount of map area, but improvement was much greater for azimuth than for distance. Figure 2.) shows the amount of azimuth error as a function of the radius or area visible on the map for four different sites.

Insert Figure 2 Here

Of particular interest is when there is a sharp decrease in angular error with a change in area as happens in at least once case for the Afton, New Mexico, and Arizona (C3) sites. It is sometimes possible to infer which features account for the improvement as in the masking study described above.

SUPPLEMENTARY STUDY: METRIC SLOPE AND DISTANCE JUDGMENTS. Map readers in the field drop-off problem appeared to have some difficulty both with metric judgments of slope and of distance. This was also true of some slope judgments in the laboratory simulation task. To explore this further, map readers in the laboratory masking study were asked, after they completed the matching tasks, to make judgments about the inclination of slopes and distances of various target features in the terrain pictures. The slope judgments were made by setting a pointer to correspond to a specified terrain inclination and the distance judgments were made by responding in yards or miles to distance specified on the terrain photographic slide or map. The relation between the judged and actual distances for scene and map are shown in Figure 3.) As is obvious observers were quite good in making these judgments; relative accuracy was very high with judgments approximating the same rate of increase for judged and actual distance. In the case of terrain studies the judged distance was slightly but consistently overestimated.

Insert Figure 3 Here

Similar data for the slopes are shown in Figure 4.). Again the relative accuracy for the slope judgments was quite good . For both scene and map the estimated slope was linearly increasing function of the actual slope. However, the large upward displacement of the function form a perfect match indicates that in both cases the judged slope was much steeper than the true slope. This overestimation of steepness of slope is often observed in everyday situations. (The incline of even the steepest highway hills is rarely more than 6 or 8 percent but we often feel we are going down a 45 degree hill.) Why the map slopes were similarly overestimated is not at all clear. The observers were asked not to calculate the slopes which they might have done by counting contour lines and relating the vertical change to the horizontal distance. The inaccuracy in absolute slope judgments could lead one to make errors in a map-terrain correspondence problem if one were relying on the inclination of a particular feature as was a case in the New Mexico photographic scene. This shouldn't be a problem if one were using a configuration of features even if it were based on changes in slope.

Insert Figure 4 Here

Scene memory task.

The aim of the scene memory task was to explore what features of terrain scenes are salient when persons are given the specific task of remembering the scene. Would this be different if the person knew that they would be performing a map reading task relevant to that scene. The scene memory of three groups of 16 participants each was compared: map readers with a map-terrain correspondence task in mind, map readers simply asked to remember the scenes, and a group of non-map readers. Each of five photographic scenes (a subset of those used in the laboratory map-terrain mapping task) was presented on a rear projection screen. Half the participants in each group viewed each slide for fifteen seconds and half for thirty seconds. After a 3-minute delay period they were asked to recall and then draw all they could remember.

The average number of features remembered was approximately equal for all groups ranging from 7.3 for the map readers with map task in mind to 8.0 for the map readers just viewing the scene. However, as noted above the proportions of types of features differed for the different groups. The map readers with correspondence task in mind recalled more terrain and fewer vegetation features than the non-map readers. The different groups also differed with respect to the types of terrain features recalled. For example, the map readers with correspondence task in mind recalled more slope features than the other two groups and all the map readers recalled more valley terrain features than the non-map readers. Not unexpectedly somewhat more features were remembered by the the participants viewing the scenes for 30 seconds than for 15 seconds (8.2 and 7.0 respectively).

The performance of the map readers with correspondence task in mind on the subsequent map reading task was examined in relation to their recall of features. These map readers were given the laboratory simulation task described above of choosing which of three arrows on a topographic map specified the photographic scene they had previously been shown. On the average they solved 2.88 of the five problems correctly, performance significantly above chance. Performance on the map task was significantly correlated with total number of spatial features recalled (.55) and with number of terrain features recalled (.66). Also performance on the map task was negatively correlated with the number of vegetation features recalled (-.64). It would appear then that scene memory with a map-terrain task in mind does reflect features that will be useful in topographic map reading tasks.

Computational Model

The taxonomy, scene, and map knowledge bases and control structure are all elements of the computational model for solving localization problems. Figure 5 shows an example of map data partially instantiated against partially interpreted scene data. The taxonomy knowledge base is used to create a hierarchy starting with topographic features and continuing on down through the solid (subclass) links to map and scene features, primitives and configurations, etc. In this example, the image knowledge base consists of the frames representing two peaks (P-1 and P-2) and a valley (IV-1) which have been recognized in the scene. These frames have been attached appropriately in to the taxonomy domain by membership (dashed) links. The map knowledge base consists of three hanging valleys (hanging valley-1, -2, -3), three canyons (moran canyon along with its south-fork and north-fork), and a col (col-1). These are attached to appropriate places in the taxonomy hierarchy via membership links.

Several of the control structure procedures are shown in Figure 5. These procedures are divided into two classes. General strategy rules

Insert Figure 5 Here

include reconnaissance (both initial and follow-up), map orientation, feature matching (both scene to map and map to scene), configuration matching, hypothesis generation and evaluation, and conclusions. Specific procedures perform tasks such as grouping configurations and attentional processes such as looking for unique or unusual dat like prominent high points or unusual configurations.

Conclusion

The research supported by the grant has provided a detailed description of the problem solving of experienced map readers as they attempt to solve a drop-off localization problem. The description includes both the kinds of features attended to and the information processing procedures in which the map readers engage. One important constraint on successful problem solving is that the map readers must be permitted some mobility. This is especially useful for providing information about the local terrain around the view point. The results from investigation of the field localization problem have been extended by means of laboratory simulation tasks. A computational model is being elaborated which captures both the features and processes identified in the analysis of expert map reading behavior.

Table 1

Masksing Study

Testing Location	Masking Condition			
	Full Map 	Inner 1/3 	Outer 1/3 	Outer 2/3 
O'Brien A Map Task*	31[a]	35	73	40
O'Brien B Map Task	56	60	33	40
Arizona Map Task**	61	47	47	20
New Mexico Map Task**	12	50	13	7[b]
Afton Map Task**	69	18	80	57
O'Brien A Scene Task	75	82	67	73
O'Brien B Scene Task**	75	12	73	60
Arizona Scene Task	62	47	73	47
New Mexico Scene Task	31	65	47	47
Afton Scene Task	50	41	53	53
Mean*	59	45	56	45

Note. *N* = 16 in the full map condition, *N* = 17 in the inner 1/3 masked condition, and *N* = 15 in the outer 1/3 and 2/3 masked conditions.

[a] Underlined percentages are significantly greater than chance (33%) at $p<.05$.

[b] Percentage correct is significantly less than chance at $p<.05$.

* $p<.05$ ** $p<.01$ --for comparisons between masking conditions.

Figure 1. Example of topographic map used.

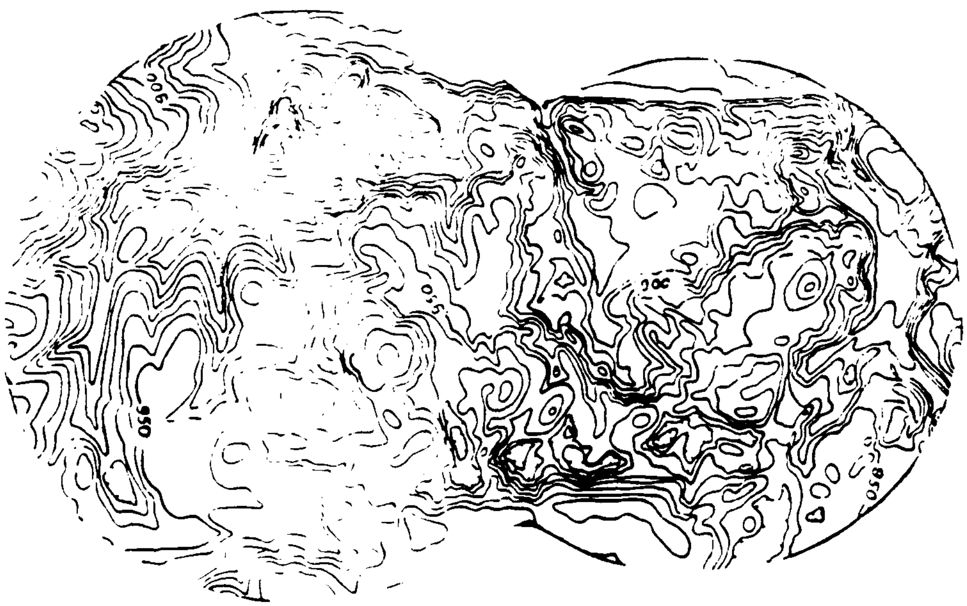


Figure 2. Graph of Azimuth error for viewpoint localization.

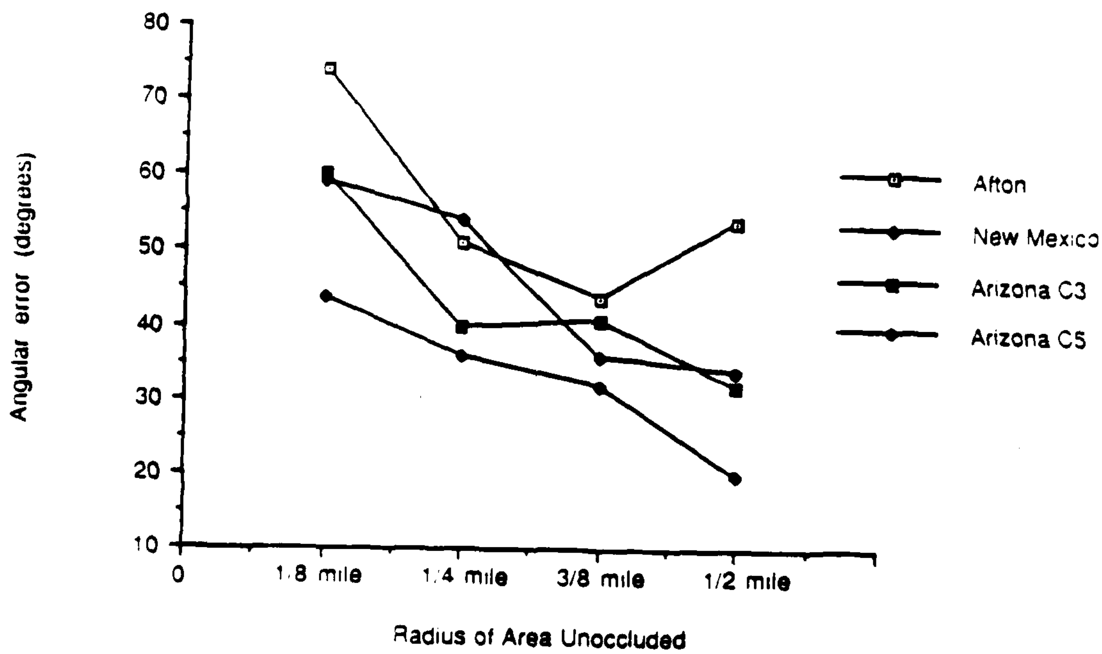


Figure 3. Relation between judged and actual distance.

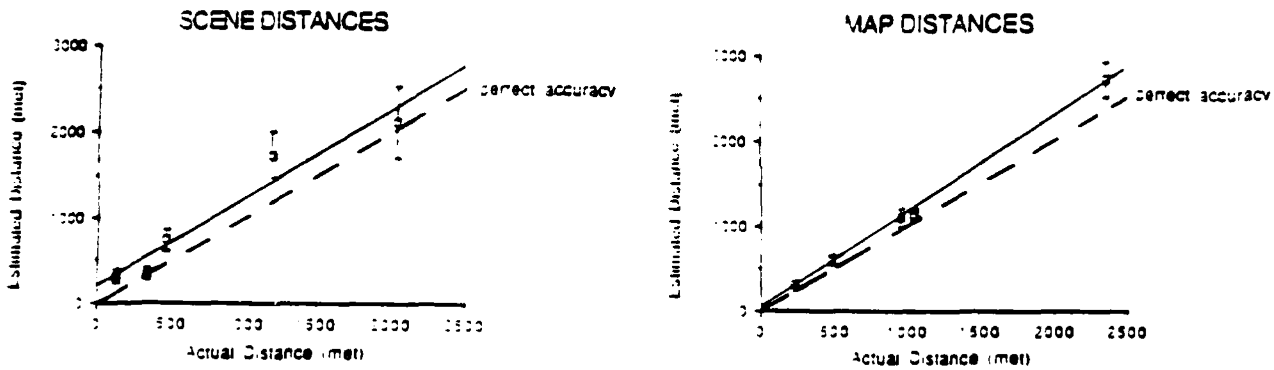


Figure 4. Relation between judged and actual slopes.

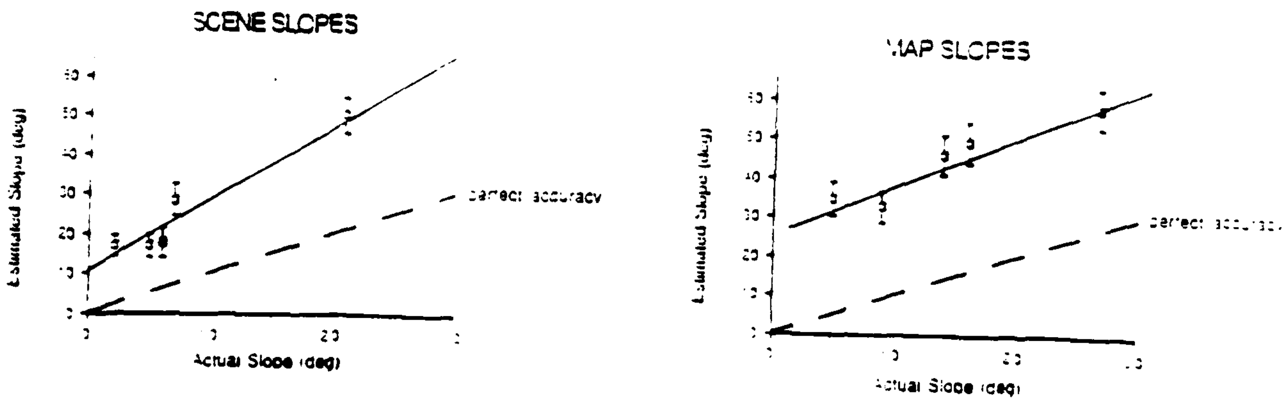
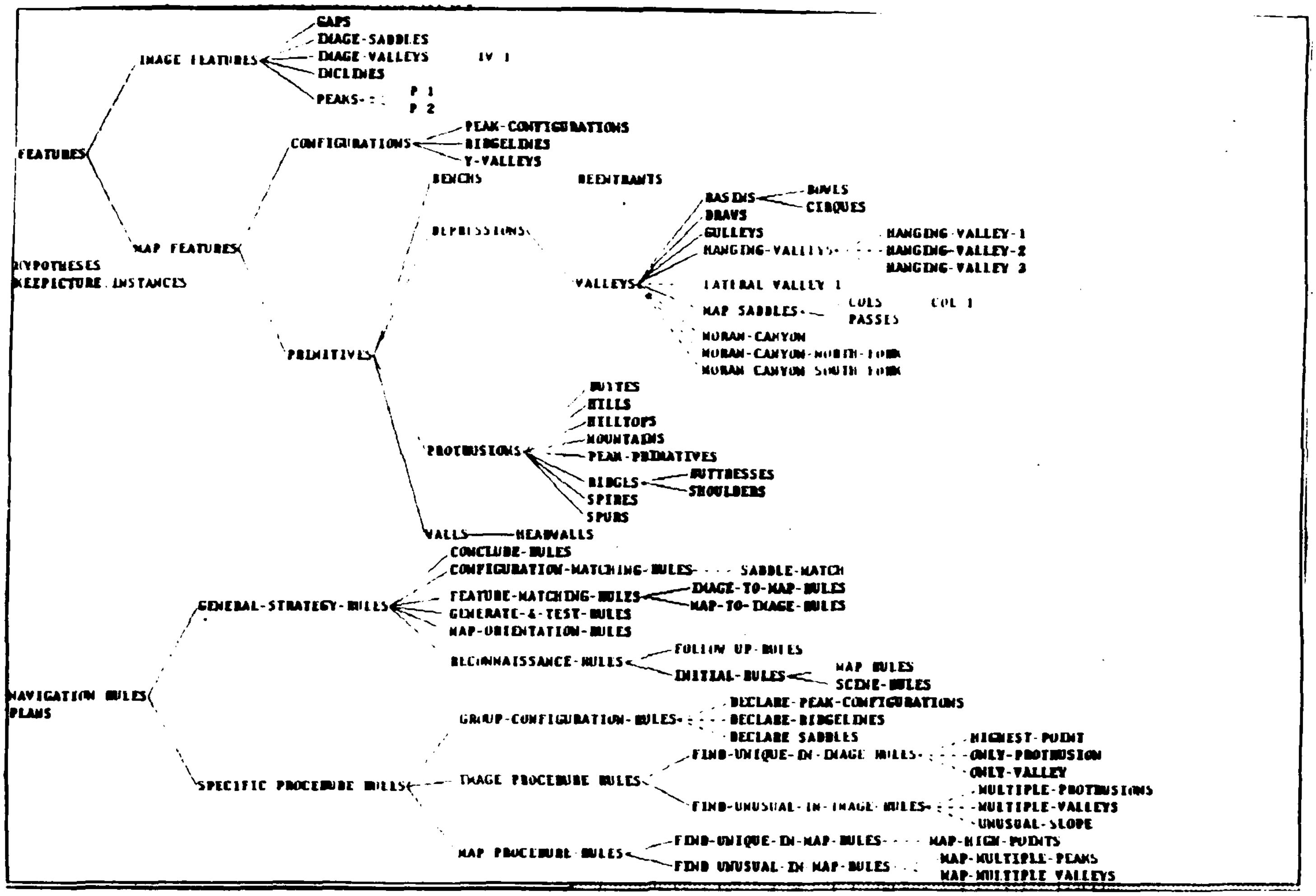


Figure 5. Example of relation between scene and map knowledge bases and control structure.



Appendix

1. Reading topographic maps. Draft of paper to be submitted to <u>Cognition</u>.

2. Smith, K., Heinrichs, M., & Pick, H. Where am I! (Summer, 1991.) Similariy judgment and expert localization. Paper in <u>Proceedings of Cognitive Science Conference</u>.

3. Thompson, W. B., Pick, H. L., Jr., Bennett, B., Heinrichs, M., Smith, S. L., & Smith, K. (September, 1990). Map-based localization: The "drop-off" problem. Paper in <u>Proceedings of Image Understanding Workshop</u>, DARPA.

DRAFT 16 April 1991

Reading Topographic Maps

Marian Heinrichs, Daniel Montello, Herbert Pick,
Christian Smith, Catherine Sullivan, William Thompson
(In some permutation)

University of Minnesota

Introduction

Topographic maps are interesting representations of the layout of the environment. The 2-dimensional layout of the map is formally similar to the 2-dimensional layout of the world in the sense that displacements in particular directions on the map systematically correspond to displacements in particular directions in the world. However, the third dimension or elevation in the world is encoded symbolically on the map by means of contour lines. The present study is concerned with the use of topographic maps to solve localization problems, that is, finding the position on a map which corresponds to where one is in the world. In order to do this a map reader, in addition to relating the contour line information to the elevation in the world, must be able to cope with the difference in perspective in viewing the map and the world. The perspective on the map is typically a bird's eye view; the map is roughly perpendicular to the line of sight, while the perspective on the world is typically from eye level viewing in a direction almost parallel to the ground. How do skilled map readers accomplish this task?

Although both geographers and psychologists have been interested in map reading researchers from neither discipline have studied the use of topographic maps to solve localization problems. In the first place with few exceptions psychologists have focused on the use of road or street maps and/or political maps while geographers have been more interested in thematic or political maps. In the second place research has tended to be focused on reading maps alone rather than relating maps to the environment. One major emphasis has been on how particular map characteristics affect the extraction of information from the map. Thus using psychophysical procedures investigators have examined the discrimination of specific features (Shortridge, 1979), the perceived size of point symbols (Crawford, 1971; Chang, 1977; Flannery, 1971) or perception of symbol size differences (Crawford, 1973; Meihofer, 1973; Shortridge, 1979). Other studies have involved simply asking subjects what type of representation they prefer. For example, Shurtleff & Gieselman (1986) asked novice map readers which of a number of symbols representing map features (e.g., lakes, rivers, trails) commonly found on topographic maps were the best representatives of their referents.

Another emphasis in map reading research is the investigation of processes underlying extraction of information from maps. A popular method in such research has been to examine the eye movements of subjects engaged in map reading tasks. Eye movement recordings have been considered useful for measuring focus of attention , depth of processing, and use of peripheral vision (Chang, Antes, & Lenzen, 1985). Analysis of individual differences in map reading performance has also been used as a way of investigating information processing in map reading. Sholl and Egeth (1982), for example, in a systematic psychometric approach, related performance on a number of topographic map tasks to several more general standard psychometric measures. The map tasks such as land form identification, slope identification, spot elevation, terrain visualization were factor analyzed yielding two major factors, one described as a spatial visualization factor and the other as an altitude estimation factor. Surprisingly standard tests of spatial ability are not highly correlated with the spatial visualization map reading factor whereas verbal analytic measures are. Standardized measure of mathematical ability is related to the altitude estimation factor. In general the results seem to suggest that our standardized tests don't capture very well the abilities used in a practical skill like topographic map reading. While these studies are concerned with the processes involved in map reading they concentrate on the map itself and not on the relation between the map and the environment. Even when the tasks call for matching a map feature to an environmental feature the environmental feature is usually a schematic diagram and is a very small fragment of what one would see in the world.

One study which does examine the detection of correspondence between a section of map and a more extended representation of the environment surface is that of Eley (1988). He investigated the effect of differences in orientation in view point on speed of matching a map position to the topography of a pictorial representation of a surface. Typical mental rotation results were obtained. The more the target viewing angle to the map deviated from the subject's own orientation toward the surface the longer the reaction time to press a button for the depicted surface topography. In a second experiment reaction time was measured for surface views from different elevations. Results indicated that an elevation providing a view point of 30 degrees above horizontal was more effective than either higher or lower elevations. The effects on map reading performance, of mismatch of orientation of map and environment, has also been found with street maps (Levine, Marchon, O'Hanley, 1984; Adeyemi, 1982). However Eley's results implicating elevation of view point also suggest that perspective taking in the third dimension is a factor when topographic features are involved. It should be kept in mind that even in Eley's study the topographic surface was a computer generated wire screen presented on a CRT.

In sum the research on map reading tends to use tasks that concentrate on the map alone or if they do involve relating map and environment use schematic or impoverished representations of the environment. The present study is an investigation of map reading starting with the task of relating a map to the real environment in the context of a localization problem.

Localization problems range along a continuum of amount of initial information available to a person as to where they are. Towards one end of the continuum would be an "updating" problem. A person knows where they are at some particular time. They move away from this initial position and at some later time try to update where they are on the map. Near the other end of this continuum is the so called "drop off" problem. A person has minimal information as to where they are to start with but need to find their location on a map. This might occur , for example, when a person doesn't keep track of their movement through a strange environment and suddenly realize that they are lost. In the present study the behavior of experienced map readers is examined as they try to solve drop off map reading problems. The drop off problem was used as apriori it would seem to place the greatest demands on the map readers. The study consists of three parts. The first is a field experiment with protocol analysis of subjects trying to solve a drop off problem. The second is a laboratory simulation of the drop off problem in which the amount of map information available to the subject was manipulated and the third is a protocol analysis of subjects trying to solve the simulated drop off problem.

Experiment 1 Field Experiment and Protocol Analysis

The goal of Experiment 1 was to collect and analyze protocols of expert map readers attempting to solve drop off map reading problems in the field. The goal was to obtain descriptions of the terrain and map features attended to and the strategies used by the map readers in attacking these problems. This information could then be used to design laboratory simulations of the drop off problem where terrain and map information could be more carefully controlled. Subjects were taken blind-folded to a station point in a state park approximately thirty miles distant. When on station the blindfold was removed and the subjects were asked to find their position on a topographic map. Initially subjects were not permitted to move more than a few steps from the initial station point. However, the task proved almost insoluble with that constraint. Therefore, a second condition was introduced in which later subjects were permitted to move freely while attempting to solve the problem.

Materials and setting.

The study was conducted in a local state park with generally rolling hills and valleys from a station point near the top of a grassy hill. The station point was selected to permit a view of two to three miles in several directions. (Figure 1 is a view from the station point toward the north.)

INSERT FIGURE 1 HERE

No cultural features such as roads, buildings, or power lines, etc. were visible from the station point although a park trail was visible (which did not appear on the map). The map provided the subjects was a portion of a geodetic survey topographic map overlay. (See Figure 2.) The particular

INSERT FIGURE 2 HERE

overlay included all the topographic information available on geodetic survey maps but did not include any foliage information such as swamp or wooded areas. Nor did the map include any grid lines or geographic orientation indicators. The map segment itself was an irregular shape cut out of an original geodetic survey map so as to eliminate some cultural features and a distinctive river. (The relatively impoverished map was used so that subjects would have to rely solely on topographic information for solving the localization problem.) The map did have a distance scale marked on it and there were elevation numbers on some of the contour lines.

Subjects.

Subjects were 29 experienced map readers. They were recruited from among geology and geography graduate students, orienteering clubs, and other outdoor and wilderness organizations. Their experience in using topographic maps ranged from...... to .... Among the subjects was one who placed in along the .... Range of .....

Procedure and Design.

Subjects were driven blindfolded approximately 30 miles to a road access point about a 400 meters from the station point. They were then led across a level field and up a hill to the station point. The blindfold was removed, they were given the topographic map segment affixed to a clipboard and they were asked to find where on the map their current position was. They had been briefed on the procedure while driving out and were instructed to think aloud, in as detailed a way as possible, as to how they were solving the task. A subset of 17 of these subjects were videotaped, trying in so far as possible to record where they were looking and what they were pointing at during their explanations. The records of this subset of subjects were subjected to protocol analysis.

Seventeen of the 29 subjects were also instructed not to move appreciably from their initial position. (They were permitted to move a few feet to see past a bush or in turning.) The remaining twelve subjects were permitted to move around and explore as desired in solving the task. (Three of these were permitted to explore only after coming to a an initial solution. They were told at the end of their stationary verbal protocol that they could move if they desired to confirm their solution.)

Results.

Solution. Of the seventeen stationary subjects only one arrived at a correct solution while six of the twelve exploring subjects identified the correct map position. Such a difference is statistically significant (X[2], df [1] = 5.6; p < .05)

Protocol analysis. The subjects' verbal protocol was taped as they thought aloud while attempting to solve the localization problem. Their behavior was simultaneously videotaped to provide information as to where they were looking and pointing while thinking aloud. The verbal protocols were transcribed and each one was coordinated with the subject's videotape to determine in ambiguous cases what features in scene and on map were being referred to. On the basis of these transcriptions a coding scheme was developed with which it was possible to analyze all the protocols.

Each protocol was described in terms of a Problem Trace which was a 2-dimensional display or graph. The horizontal axis of the Trace represents the temporal sequence of the problem solving in terms of the sequences of statements made by the subject. Each statement is a coherent utterance with a single focus of attention. (Statements are typically separated by breathing pauses or by changes in the focus of attention.) A series of statements comprise problem solving episodes which are directed toward a single higher level goal. The vertical axis of the Problem Trace captures three aspects of the localization task: 1) source of information (map or terrain), 2) type of feature attended to, and 3) the goal-directed activity or process being engaged in at the time. The Trace actually includes separate tracks for each of these aspects. An example of such a Trace is shown in Figure 3.

Insert Figure 3 Here

The numbers along the horizontal axis are sequential statements. The vertical lines separate different episodes. The source of information is indicated by the lower track, map or terrain. The type of feature is indicated by the middle track and the activity or process is indicated by the upper track. In general all three tracks are simultaneously relevant. For example, when a subject says: "Looking down hill, it looks like I'm looking into a very broad valley." The source of information here is the terrain, the feature attended to is a valley and the process is reconnaissance. The type of features attended to in the map and/or scene as evidenced by the protocols were: valley, saddle, river, ridge, plateau, lowlands, hill, gradient, field, distance, and contour. There would likely be other or additional features if the localization task were carried out in different terrain, e.g. mountainous or desert. There were six processes or activities identified: reconnaissance, map orientation, feature matching, configuration matching, hypothesis generation and evaluation, and conclusion. A detailed description of the coding process is available from the authors. Here a general description of the processes and how they function will be provided.

<u>Reconnaissance</u>. Localization problem solving is almost always initiated by an extended period of reconnaissance. The search is extended broadly without any particular focus. Perceptually distinct topographic properties of the map or scene that are potentially relevant to establishing map-image correspondences are identified. The more successful problem solvers seem to spend most of this initial time examining scene features, organizing the information into a cohesive representation of features and configurations. Initial reconnaissance focusing on the map seems less successful.

> An example of reconnaissance focused on the scene: "So, umm standing on a slope here, it's sloping down on pretty much all the way, like 180 degrees sloping down that direction, so. And it looks like there might be a hill behind us, although it's hard to say if it goes down on the other side. But it looks like a pretty high spot in the terrain area, so it's probably one of the higher areas on the map, especially and higher over there. That's about it." (OA 3-5). Typical of map reconnaissance would be: "Ahh, looking at the maps, the map, it uh, doesn't show trees so as far as the wooded area and... It doesn't show like the farms out in that direction, ah. I think the biggest thing is for me to use hopefully would be this long valley if it's a stream of river system. Umm, looking at the map there, there appears to be a couple of things that could be a stream valley, umm. This marks a depression with the slash marks." (DJ 5-9).

The purpose of reconnaissance is to gather information prior to the creation and/or evaluation of specific hypotheses about the viewing position. Follow-up episodes of reconnaissance generate additional information and can be prompted by three different situations. Acquisition of additional information is common during the evaluation of a hypothesis. The additional information is required whenever the current information is insufficient to establish the hypothesis. Follow-up reconnaissance is also useful during the refinement of an hypothesis that is being accepted. The additional information serves typically to fine-tune the hypothesis. The most common use of follow up reconnaissance is a "strategic regrouping" after the rejection of an hypothesis. This regrouping appears to serve the same purpose as the initial extended reconnaissance, the gathering of information required to support the targeting of a new hypothesis.

Map orientation. Map orientation involves relating the direction and scale of the map to the visible scene. In a typical way-finding situation the orientation of the map is given via direction lines and/or a compass rose on the map as well as the usual orientation of grid lines and print and orientation with respect to the scene is available if the way finder has a compass. In the present experimental situation all conventional directional information has been removed from the map and the subjects do not have a compass. Solving the localization problem does not logically require orienting the map with respect to the surrounding environment. However, doing so aids in constraining and systematizing the other processes and almost all subjects specifically engage in efforts to orient the map with respect to the environment. The general lay of the land in the area of the present field situation provides information for determining a corresponding map orientation. Direction of sun (when visible) in conjunction with time of day provides geographical direction information to a way finder but this by itself is not useful with present localization problem because the geographical information has been removed from the map. The following exemplars from the protocols are characteristic of the orientation process:

> "Since the land on the map falls off away from us this way, and most of the land appears to be falling off in that direction, I figure this is the way the map is on the land here." (SK, 21)
>
> "Since the general slope of the land does go from behind us, the high behind us, to lower this way, and this (contour on map) is 900 down here, the land is getting lower on this side (of the map). So this would be low over here and high here. I would have to orient it this way." (RB, 49).
>
> "Maybe I'm holding the map upside down. I don't think so because this has, the general slope of the terrain's that way. The general slope here is sort of going down this way, and if I hold it upside down there's no place on the whole map that would mention the slope going that direction, because it's all that low, going up, so it must be this, must be correct. This is the way the map should be oriented, perhaps a little like this." (OA, 41).

Feature matching. The major activity during the localization task is matching features in the image to features in the map or vice versa. Feature matching does not require the existence of a specific hypothesis about viewing location. Such matching can establish general correspondences between the environmental scene and map, facilitating the generation of specific hypotheses. After hypotheses are formed feature matching plays a key role in their evaluation.

Feature matching is based on a common identification and similar characterization of topographic structures in map and scene. Identification is done in terms of a set of labels and properties, often specific to a particular geographic landform. In the geological area of the present study the most common features attended to are hills and valleys. Attempting to find correspondences for the mere presence or absence of a hill or valley is not particularly diagnostic of location. Accordingly map users more commonly attend to properties of these features rather than just their existence. To differentiate they focus on relative size, elevation, and slope (gradient).

Most subjects tend to impose a bipolar or qualitative classification to differentiate properties of features. Features are large or small, narrow or wide, steep or shallow. Comparison is another common strategy to differentiate features. One feature is described as larger. broader, or steeper than another. Consider the following examples:

> "Then down there there's a big valley so I guess that could be this valley going down here, and if that's the case, the high area we're seeing , might be this ridge extending out here, and umm" (OA 15-16).
>
> "This area right here, ah, gently sloping while fairly flat on top, so maybe look for some kind of plateau on the map, and, that drops off relatively to my left to the water and to the front. There's a couple of areas on the map that look gently rolling like this area here or over in here, umm, both of them to have a water area off to the left" (DJ 22-26).
>
> "Hmmmm, I don't know. There should be a hill on the other side of that, on this wet land right in there. There is a hill I see over there, a grassy hill. Trees behind it. Could be, could be this hill here. It's kind of steep slope, indicated by the closeness of these topo lines right here" (RB 32-34---hypothesis evaluation).

Configuration matching. Configuration matching serves the same purposes as feature matching. The only difference between the two are that the pieces of information that are being attended to are assemblies of features. Configurations are specified in terms of the features of which they are composed and the relationships between those features. These relationships include purely topological descriptions (e.g., behind, in front of, next to), and quantitative properties (e.g., actual elevation). More expert map users tend to do more configuration matching and less feature matching than do less proficient individuals. The complexity of the configurations is usually relatively small, however, typically involving two to four individual features. Success in the localization task appears to depend on the accurate establishment of appropriate configurations for matching.

Configurations constrain the matching process more effectively than do individual features. There are fewer matches to "a hill with a dip and a ridge" than there are to individual hills, valleys, or ridges. By bundling features together search can be restricted to more unique configurations.

A pair of simple but highly effective heuristics characterize the assembly of features in the configurations of experienced map users. The first heuristic restricts configurations to features that are contiguous. Features that are combined to form configurations are actually physically adjacent (e.g., "the flat area that slopes down and then up again to a ridge"), rather than just adjacent in the scene due to occlusion. Field subjects have been observed to trace out in the air with a finger the connection between features as they refer to a configuration.

The second, less-rigorously applied heuristic, restricts configurations to features that align along a line-of-sight. The majority of configurations (perhaps 80%) used by the present subjects are composed of features that fall along or parallel to an azimuth extending away from the observer. Most of the remaining configurations (the other 20%) focus on the distribution of features along prominent ridge-lines that cut across the line-of-sight. These configurations share a property of linearity. Explicit reference is often made to non-linearity when a feature in a configuration does not line up (e.g., the crook in a ridge-line or the slight offset in a string of hills and valleys).

Most configurations conform to both heuristics. Both derive their power from the fact that they disallow configurations that could be products of accidental viewpoints. Both connectivity and linearity are viewpoint invariant properties of a scene that survive the transformations required for matching. Lowe (1987) emphasizes a similar importance for viewpoint invariant configurations of features in object recognition. Typical examples from the protocols include:

> "I see three, a low hill, a very gentle dip, and another kind of low hill and then a third one to the west of this, well this, I know the wind is from the northwest so, I assume this is north, out this way somewhere. Ahh, so I'd see those over there. These would be the three hills I could pick out" (RB 25-27 hypothesis evaluation).
>
> "There are some big hills on the other side of this gulley/ravine through here. There's a ridge with four little, kind of a rolling ridge which I think would be off to the south off that way and then it drops off fast down into a far away into a kind of a big into the valley, the main valley, so that would be back in here" (RB 89-91 hypothesis evaluation).

<u>Hypothesis generation and evaluation</u>. An hypothesis posits a distinct map location and direction as corresponding to the viewing position. The hypothesis is initially triggered by possible map-scene correspondence between a small number of features or configurations. Hypothesis evaluation proceeds by examining other scene and map features or configurations using expectations derived from the hypothesis. Often a brief reconnaissance of a local region in the map and/or scene will be required to identify additional features and configurations useful in the evaluation process. The strategies involved here have much in common with those used in other diagnostic tasks, e.g., Johnson, Moen, & Thompson (1988).

While viewpoint invariance is desirable in the spatial arrangements of features which define a configuration, <u>viewpoint dependence</u> is obviously necessary for hypothesis testing. An hypothesis must necessarily describe the relationship of topographic features to the viewpoint. The protocols of the present experts indicate the use of rather simple, qualitative descriptions rather than a more sophisticated trigonometric analysis.

The search through alternate hypotheses can proceed in a variety of ways.

A breadth-first strategy, typically not very effective, involves generation of a large number of hypotheses before attempting to evaluate any of them. The generation of each individual hypothesis is based on a small number of features, often only one. More focused search strategies are characterized by generation of successively more precise hypotheses based on increasingly a richer sets of configurations. These focused searches may involve alternate generation and evaluation or generation of a small set of possibilities with subsequent simultaneous evaluation. The following illustrates the generation and rejection of hypotheses:

> "And that other open area that we just barely see, between the, that seems that could be this area here. Umm. But yeah, it looks pretty good. The other side, if that is the case, that we're actually down here now, that, if we were down here, that should be like umm, a ridge going out. I guess there is sort of like a ridge right down there. A little ridge. I don't see it bending to the right though. There's definitely a valley going down there, but yeah. Oh I see. Maybe that valley is this valley. In that case, this makes us up, more... no that doesn't look right either, cause then it should be pretty flat, and it looks sloping more going down there. Hmm, maybe I'm in a completely different location on the map. Hmm, it goes..." (OA 32-40).
>
> "So what I'm looking at, is that we're kind of on a hump that kind of comes out quite a ways. Maybe I should, seems like we're coming around in direction up a hill like that. So maybe an oblong, more oblong shape hill. Probably something like this down here (referring to map). Umm, there's more of a ravine on that, but this, this hill here doesn't look like it's big enough. This is the big, seems like the biggest hill in the area, and so to me, that isn't a big enough contour or a big enough hill on this map to signify that's where I'm at" (JP 50-57).

Conclusion. Hypothesis evaluation leads to the tentative rejection or confirmation of hypotheses that have been generated. A final step in the localization process produces the best estimate of actual location and viewing direction. Depending on the search strategy used, this may be based on a comparison of the likelihoods of competing hypotheses or may simply be the identification of a single hypothesis which survived a sequential generate-and-test procedure. The subject may be satisfied (success) or unsatisfied (failure) with the final statement. An example of each:

> "Let's say these are about 10,20, this is 30 feet above this line so these actually should be the same height. OK, so I probably wouldn't notice it too much. I think, ah, we're here on this ridge. Umm, let's just look at this one more time here. This is, umm, OK, I think we're here" (RB 96-99).

> "This one doesn't match because it was too steep. What about this one? Maybe this one. I have to...Let's see. But then, umm, there should be a very sharp or steep valley here, but I don't see that at all, so it's probably not that place. And it's probably not on this...Unless it's down here, umm. Because it's very steep below that, and it's certainly not generally sloping here. So I think the best guess is that we're about here. That's my best guess. It doesn't match completely though" (OA 73-79).

Overall the types of features and configurations of features used by the present subjects are probably constrained by the local topography. However, it is interesting to note that most feature description was qualitative rather than metric. Especially judgments of slope gradients were made in terms like steep, medium, shallow etc., rather than degrees. Quantitative judgments of distance were more frequent but still not heavily used. When they did occur they sometimes were in units of time, e.g. a 5-minute-walk, etc. It is also the case that there was little reference to the most distal features of the layout. This might have been partially due to the range of the map but that wouldn't account for the heavy neglect of such features.

Neither in the scene nor map descriptions and hypothesis testing were there any statements that could be characterized as reflecting global visualization. The descriptions specified features or at most configurations of features. This was somewhat surprising as informal reports in early pilot interviews of informant map readers included statements about looking at a map and visualizing the general overall topography.

In trying to summarize the information processing reflected by the protocol analysis it may be useful to think in terms of a focus on the map or the terrain and in either case attention to the station point and its immediate surrounds or attention to the more distal features and layout. In terms of such a two-by-two classification (map vs terrain and station point vs distal layout) the goal of the task is to arrive at a solution which specifies a station point on the map. Two general strategies are observed: a map driven strategy and a scene driven strategy. The more successful subjects appear to use the latter. The reconnaissance of the terrain informed the reconnaissance of the map; their feature matching was guided primarily by inspection of the terrain. This strategy is not surprising since the terrain imposes more constraints on hypotheses than the map. Everything visible in the terrain (subject to the criteria of feature and scale for representation on the map) is relevant. The information on the map is not constrained by what is visible from the station point and hence includes much more.

In testing hypotheses the more effective subjects seemed to use a simultaneous comparative approach rather than one involving sequential generation and testing. A common source of error in evaluating hypotheses was to "explain away" potentially disconfirming evidence which did not fit expectations based on a station point hypothesis. The error would be to discount the expectation from the map. The most common problem of subjects in the present study was not generating or accepting the correct hypothesis because of inaccurate registration of the terrain in the immediate surround. Inadequate reconnaissance of the current position led to simplistic description of the station point without concern for disambiguating constraints. The greatest gain for the exploration condition was the more accurate registration of current terrain position.

Experiment 2. Laboratory Simulation of the Localization Problem

The second part of the present study was a laboratory simulation of the localization problem in which the amount of map information available to the subject was manipulated. In this experiment subjects were asked to match a scene with a direction line on a map. The purpose of the laboratory task was to examine the information used in solving a localization problem in a more controlled manner. By restricting the demands of the map-scene correspondence task to forced-choice answers, the laboratory task was designed to explore a subset of questions addressed in the field studies. It is clear, however, that only an evaluation of both the laboratory and field data can give us a truthful picture of both what people can do, and what they normally do when asked to solve localization problems.

The first aim of the laboratory task was to investigate what amount of information people need to match a map with a scene. One specific question was whether performance is directly related to the proportion of visible map area. A second related aim was to determine whether, independently of size, certain areas of the map are in general more informative than others in solving correspondence problems. Specifically, where is the most useful map information for solving a localization problem, close to the station point or at intermediate or far distances.

What are the particular features or group of features that are the most useful for solving the correspondence problem? A finding that, hills for example are preferred over valleys or more generally, that the use of features is favored over a more wholistic approach, would be of value for a more general understanding of the map reading process.

To investigate such questions four groups of subjects were asked to match photographic scenes to a position and direction line on a topographic map. Map information was manipulated by masking portions of the map to varying degrees for different groups of subjects. One group of subjects was presented with full or unmasked maps, while the other three groups of subjects were given maps with various portions masked. In the "inner 1/3" masked condition, an area defined by the third of the radius of the map directly surrounding the central station point was occluded. In the "outer 1/3" masked condition, the more distal third of the map's radius was masked leaving a central area corresponding to two thirds of the radius unmasked. Finally, in the "outer 2/3" masking condition, the distal two thirds of the radius was masked leaving only a small central area directly surrounding the station point unmasked. As a consequence, the "inner 1/3" and "outer 1/3" conditions were equivalent in terms of the radius proportion masked, while the "outer 2/3" masking condition had the smallest amount of visible area (Figure 4).

Insert Figure 4 Here

This experimental manipulation allows us to directly address the question of the amount of map information needed to solve the task, as well as whether particular areas are favored over others. If the amount of available map area is the only variable affecting performance on correspondence tasks, the full map control condition should produce the best performance, the performance under the "inner 1/3" masked condition would be the next best, followed by the "outer 1/3" masked condition. Finally, the most errors should occur in the "outer 2/3" masked condition since it has the most map area masked. Any deviation from these predictions will allow us to infer which areas are the richest in information. Examination of these areas would enable specification of the features most important for problem solution.

Subjects were asked to solve map-scene correspondence problems of two types. In one, the map task, they had to select the one of three direction arrows from a station point on the map which corresponded to a photograph of a scene projected on a large screen. In the other, the scene task, they had to select the one of three pictures which corresponded to the view that would be seen from a station point on a map in the direction of of an arrow emanating from that station point. The geography sampled across these problems included gently rolling hill areas in Minnesota and more rugged hilly and mountainous terrain in Arizona and New Mexico.

These tasks are obviously more constrained than the field localization problems where subjects are asked to find their location on the map. However, they do constitute a subset of such map-scene correspondence problems since the subjects have to match features on the map to scene characteristics to succeed.

In addition to evaluating the overall effect of masking across conditions we were also interested in looking at the differential effect of masking across the various locations that were used for the study. Any general research conclusions about how people use topographic maps must include both the strategies that are favored by most (for example a tendency to focus attention at a certain distance from the station point), and how those tendencies interact with the idiosyncrasies of the particular location. For example, it is possible that for some locations restricting the distal information may actually help the subject focus on the important proximal information, while for other areas the masking may be hiding the one single significant feature that would help the subject solve the problem. By collapsing across all three masking conditions for each given area one can determine which third of the radius was the most informative for solving the particular problem. If a prominent feature is included in the only visible portion of the map and performance is accurate, it can be concluded that it was important for that feature to be visible for the response to be correct. These possible differences across locations may point to interesting interactions between usual performance characteristics and the particularities of the map or scene studied.

Subjects

The subjects participating in the experiments ranged in age from 16 to 58 years old (Mean age = 28.4, SD = 8.08). Most of the subjects were geology graduate students, backpackers, orienteers, or members of the military recruited on campus and in local outing clubs. There were a total of 12 females and 51 males in the sample. Out of the 63 subjects, 16 were included in the full map control, 17 in the "inner 1/3" masking, 15 in the "outer 1/3" masking, and 15 in the "outer 2/3" masking condition. Subjects in the different groups did not differ significantly either in the amount of field experience with topographic maps or in the amount of formal training although the amount of variability across groups was considerable.

Materials

The subjects were presented with topographic maps of five locations. Three of those locations were in Minnesota, one was in New Mexico, and one in Arizona. The maps were copies and enlargements of USGS topographic map overlays which, as in Experiment 1, did not include any cultural information such as roads, trails, and houses or vegetation. The external boundary of each map was chosen to correspond approximately to the distance to horizon visible in the slides. Once this boundary was selected, the maps were enlarged or reduced to a similar diameter (about 16-18 cm).

Any information allowing the subject to align the map with the geographical coordinate axes was removed: the maps were circular, the grid lines were absent, and the numbers indicating the altitude of the contour lines were rewritten in random orientation. The maps were presented with no preferred orientation to the subjects who were told that the top of the circular map was "not necessarily north". In the "outer 1/3" and "outer 2/3" masking conditions, the distal one third or the distal two thirds of the map's radius was hidden by a black occluder, leaving a visible central area with a diameter of approximately 12 cm and 6 cm respectively. In the "inner 1/3" masking condition, a circular mask was placed over the center of the map, with the station point marked by a dot in its center, covering a diameter of about 6 cm.

A scale was marked on the maps indicating a distance corresponding to half a mile (880 yards). The length of the scale representations differed from map to map ranging from 2.2 cm to 6.2 cm in length. The contour interval was also indicated on the map and was 10 feet for the three Minnesota locations and 20 feet for the Arizona and New Mexico maps.

Color slides were taken from a position corresponding to the center of each of those maps. The pictures were taken with a tripod at leveled with horizontal. The complete set of pictures covered the whole 360 degree panoramic view. From this set, three non-overlapping pictures at each location were selected for the experiment.

The slides were presented to the subject on a rear projection screen in a darkened laboratory room. The size of the projected slide was 159 cm in width and 106 cm in height. The subject was sitting 120 cm away from the screen with a reading lamp illuminating the map from behind. A remote control allowing the subject to advance the slides was attached on the arm of the chair.

Procedure

In a short introductory period subjects were questioned about their experience and training with topographic maps. They were told that the maps contained no cultural information. They were instructed to perform the task as quickly and as possible without making a mistake.

The localization task consisted of two symmetrical types of problems. In one problem type, the subject was given a map on which one arrow was drawn, pointing away from the center. He or she was shown three successive slides corresponding to non-adjacent views taken from this center location. The subject's task was to select the slide corresponding to the view in the direction indicated by the arrow on the map. Since the required response was the selection of a scene, this task is referred to as the "scene task". The subject was instructed that the scenes presented in this task had no particular order, and that he or she could use the remote control to go back and forth between the scenes as much as necessary.

For the second problem type three arrows separated by 120° pointed away from the center location on the maps given to subjects. They were shown a single slide for each location and told that one of the three arrows corresponded to the viewing direction of the middle of the picture. Since their task was to select one of the three directions on the map, this task was referred to as the "map task".

Results

Table 1 presents response accuracy on the forced-choice tasks for each location tested, as a function of masking condition. On the average, accuracy significantly exceeded chance performance in all masking conditions, though not always at each location. Average performance in the full map and the outer 1/3 masking conditions was equivalent and significantly better than performance in the 1/3 inner and 2/3 outer masking conditions, $\underline{t}(14) = 2.80$, $\underline{p}<.01$, which were also equivalent. This pattern of response accuracy suggests that masking areas of the maps impeded the solution of the correspondence problems only when the areas were close to subjects' locations on the map or when large areas of the maps were masked.

INSERT TABLE 1 HERE

Masking did not uniformly disrupt performance at each location in this manner, however. As Table 1 shows, a variety of patterns of results occurred at different locations. Like the pattern of averaged results, accuracy on the O'Brien A map task was high in the full map and outer 1/3 conditions, but it was low in the inner 1/3 and outer 2/3 conditions. This suggests at least that the outer 1/3 radius area did not contain necessary information to solve the task. Accuracy on the New Mexico map task, however, was very poor in all but the inner 1/3 masked condition, suggesting that map information within the 1/3 radius area was possibly misleading to subjects. On the Afton map and O'Brien B scene tasks, accuracy was high in all conditions except the inner 1/3 masking condition, indicating the importance of information within the 1/3 radius area for success on these tasks. At still other locations, accuracy was either uniformly high across conditions (O'Brien A scene task) or uniformly mediocre (Afton scene task). These results suggest that the entire area of a map representing part of the visible landscape is not typically necessary for the solution of correspondence problems. And they also suggest that any particular place on the map (such as near the person) is not consistently necessary to solve the problems.

Discussion

Masking arbitrarily by area is the crudest kind of manipulation to get at the important information. The eventual goal is to predict specifically the features and configurations which are critical for map readers. The results for individual scenes have been examined to begin to get at this question. It is not the case that the result is task constrained that any distinguishing feature will be used if it is the only one available. Consider the O'Brien A map task, for example. The results for this problem fit quite closely the overall pattern of mean results with the full performance on the full map condition and on the outer 1/3 condition quite good and performance on the inner 1/3 masking and outer 2/3 masking quite poor (essentially at chance level). In this problem a distant large far away valley on the left could serve as a distinguishing feature in choosing the correct line. This was visible in the inner 1/3 masking condition but not in the outer 1/3 masking condition but apparently was not used. A similar result was obtained in the Afton map task. In the Afton scene the foreground contains a very distinctive pair of hills which were visible in all the conditions except the inner 1/3 masking. While the midground was not very informative there was one very distinctive wide valley far away on the right of the picture, which was clearly visible on the map in all the conditions. While the performance in the two outer masking conditions was as good or better than the full map condition (confirming the importance of the two hills in the foreground) the performance in the inner 1/3 condition was even below chance, suggesting that this far away valley was once again not used. (It may, of course, be that it is not the distance of this features that is important but the fact that they are valleys. There is nothing, however in the protocol data above which would suggest that valleys are not noticed and responded to in solving such problems. Valleys, ravines, depressions are, in fact, mentioned very frequently.)

Assuming that subjects do have a bias toward reliance on foreground features in solving the map tasks this tendency may lead them into trouble in some situations. The New Mexico map problem is a case in point. Subjects performed best when the inner 1/3 was masked, but when the outer 2/3 was masked, i.e. when only the inner 1/3 is visible the subjects performed at a level significantly below chance. Subjects' comments at the time of testing suggest that they misjudged the foreground slope perceiving it to be flat or inclining down even though it actually was slightly rising.

This present laboratory simulation localization task is most valuable for the hints it provides as to the specific information that subjects are using to solve the problems. Verification of these hints can be obtained by presenting such problems to subjects who are asked to describe aloud how they are going about the solutions. Such protocols were collected for a subset of the original problems of the laboratory simulation task in Experiment 3.

Experiment 3. Protocol analysis and laboratory simulation of localization.

Five problems were selected from the set of problems of Experiment 2. Three were map problems and 2 were scene problems. Ten subjects solved the problems twice, first in a masked condition and then in the unmasked full map condition. The particular problems and conditions posed were those circled in Table 1. As in the field study subjects were asked to think aloud while solving the problem. Except for this they were instructed in a manner similar to Experiment 2. The subjects were under no time pressure and their performance was not timed.

The overall results indicate 44% successful solutions in the masked map condition and 64% successful solutions in the full map condition. The performance on the full map condition is significantly above the chance level of 0.33 while performance under the masked condition does not differ significantly from chance. (**** Do statistics on this****) However, these values approximate the average performance values from Experiment 2 considering the full map, inner 1/3 masking, and outer 2/3 masking conditions from which this subset of problems was taken. Thus the overall results represent a reasonable replication of the results of Experiment 2.

The subjects were instructed as in the Experiment 1 field protocol task to think aloud as they solved the problem and their protocols were scored in a manner similar to the field protocols. The main scoring category which did not apply the same way with the lab protocols and field protocols was hypothesis generation and testing. With the laboratory task the hypotheses were in a sense given to the subject and their task was testing them.

Let us consider for detailed analysis two of the problems used in this laboratory simulation: the New Mexico map task and the Afton map task. As mentioned above and evident from Table 1, the New Mexico map task is one where performance is paradoxically better under the inner 1/3 masking condition than on the full map condition. The results of the present experiment replicate the earlier findings. Six of the ten subjects chose the correct of the three map arrows under the masking condition but only one of the ten under the full map condition. Conversely with the Afton map task performance in Experiment 2 under the full map condition is better than under the inner 1/3 masking condition. Again the results replicate here: all ten subjects gave the incorrect answer under the masked condition while nine responded correctly under the full map condition.

The protocols help account for these patterns. In the inner 1/3 masked condition of the New Mexico map task the mask covers most of the terrain presented in the slide. This pushes the subjects toward a disconfirmation strategy with which they are generally successful. One incorrect direction arrow has a prominent hill in the background which the subjects surmise would be in the background of the slide. This permits rejection of that arrow. Then they are able to guess between the other two. For example:

1) "Umm. The slide is ah, the slide is a fairly flat area, I can't ah.... The map doesn't look particularly flat. Ah, O.K. I guess I could look for, I guess I could look for things in the distance and see if... It probably isn't (arrow) 2 because if it was in the direction of 2 there's some sort of hill in that direction. And since I'm not seeing a hill in the distance, it probably isn't there, although the trees could be obscuring it. Umm... I guess ah, let's see now. It's hard to say. I'm just going to eliminate 1 for the same reason I guess. Well... 7400 ft. fairly close there, whereas, in that direction (arrow 3) there's also a 7400 ft. point but it's a little further off, so that would be more likely obscured on 3. So 1 is the best guess I can make." (Correct, JS New Mexico Map-masked condition.)

When they get to the full map condition they choose the arrow that has the gentlest slope close to the station point. The terrain in the slide appears to be almost flat although in fact it is rising thus accounting for their erroneous response. Here, as in the field study, errors are caused by incorrectly assessing the terrain of the station point.

2) "This one looks so flat it's hard to tell anything. I guess if anything those trees are maybe a little bit higher, ah... It's really hard to tell ah, I guess maybe I can try and eliminate things,

umm... O.K. if I was looking in the direction of (arrow) 1, I would expect to be looking up, right in front of me. Well I don't know how steep a slope that is , but I guess it's a couple of contour lines. Umm, let's see, smaller lines are 20 ft. intervals so that would be up about 40 ft. in the space of 100 yds. It doesn't look like it's going up that much. I'm inclined to think that (arrow) 1 would be going up a little bit more than this one is. Ah, (arrow) number 2 there's generally sort of a , from the left to the right, it's going down. This seems so flat. Doesn't even seem like there's a slight downhill. Number 3 is I guess the flattest looking one. Ah...Hmm... I guess since number 3 looks the flattest looking and this looks so flat, I'm going to guess number 3. Cause there just doesn't seem to be... Number 2 , it's too steep a hill going up. I'm sorry, number 1 (rejects arrow 1), and number 2 I'd expect to see a little more of a left to right, left to right downhill, some sort of angle. It seems so flat that I'll say number 3. That's a hard one." (Incorrect JS New Mexico Map-full map condition.)

The Afton Map task is one in which performance under the inner 1/3 masked condition is markedly deficient in comparison with the full map condition as is evident in Table 1. In fact the pattern of results for all the mask conditions would suggest that the crucial information for distinguishing amoung the arrows on the map is close to the station point. From examination of the map and scene a nearby prominent hill would appear to be the primary critical distinguishing information. Two strategies were identified from the protocols, one where more attention is paid to the map and the other where more attention is spent on the scene. When the center masked map is the focus of attention the salient feature is a large river valley and an attempt is made to see how this could fit into the scene. Then subjects choose between two plausible direction arrows.

3) "I am starting by looking at the map and am trying to determine the general shape of the terrain. I believe that this area here represents some high land and this is a river running in a quite deep gorge as indicated by the very close contour lines. This here represents I believe a valley.... probably a stream valley which comes up between this high land some other high land on the other side. I am a little surprised looking at the picture because I expected that the land form, for instance, that this describes would appear steeper land than I, appears on the picture. If I was looking in this direction (arrow 3) I think I would be looking downhill and across this...I assume this is a river but maybe I...no I think it must be... and then on to some banks on the other side. If I was looking in this direction (arrow 2), I am looking constantly downhill. And though I don't what is out here, it doesn't appear to be what I am looking there. This (arrow 1) shows that it is slightly downhill and then over perhaps a high point there . Which I think is probably that. I choose direction 1." (Incorrect JD Afton Map task---masked condition.)

When the scene is the focus of attention in this masked map condition subjects seize on the salient feature, the hill and ignore the distance scale and incorrectly select the direction arrow that shows a hill.

> 4) "I guess I'm looking up a hill. From the map I was gonna guess that I was gonna be looking down on pretty much everything. There's looks like two trails going though. I don't have those on the map. I don't, looking for a river but I don't see that in there which should make me eliminate choice 3. Looking to see if there is another hill on here... It appears that number, ah, choice number 1 goes across a low area and then back up a hill to a higher point. And that would be the direction I would choose of the three geographic areas. Number 1." (Incorrect TH Afton Map task---masked condition.)

In the full map condition of the Afton Map task all subjects focus on the hill feature and an attribute, orientation of the hill or the distance of the hill from the station point. This readily yields the correct answer.

> 5) "O.K. This is much easier. O.K. now for number 3 to be correct I want to see a big reentrant, two big valleys right ahead of me and leading into a lake. I don't see that at all. So, 3 doesn't make any sense at all. Now (arrow) 1. I would see a slight downhill and then a smaller hill in front of me before it drops off into a big valley. There is no indication of a big hill from what I am seeing on the slide to indicate that. So that doesn't make sense. Number 2 does have...ummm... this hill here, this big knoll could easily be that big hill on the map on the slide. And it also looks like you could see some of the things that we're seeing in the background--- the place where the road goes and comes in a lower spot and goes around the hill. That could definitely be around here. And you probably can't see anything off here because it is just too far. So now I would say that it is number 2." (Correct PD Afton Map task---full map condition.)

The protocols help explain the particular patterns of results obtained for the different problems. In addition they also illustrate a number of features that frequently occur in the problem solving of subjects in the field as well as in the laboratory simulation. One aspect is a tendency to focus on particular salient features. This occurs with the large river valley in 3) above and with the hill in 4) and 5) above. Even where the focus is on a salient feature a second aspect of the problem solving involves attempting to find more reliable configurations or combinations of features as happens with the attributes of the river valley in 3) and the observation of the low area and hill going to a higher point in 4). As mentioned before, a common source of error is incorrect registration of the area very close to the viewpoint which occurs in 2). It is also the case that metric information is often ignored which can lead to error as in 4). However, often ordinal information about the relative heights of features or magnitude of distances is sufficient to decide between hypotheses. Finally, in testing hypotheses, especially in the laboratory simulation, subjects realize that detection of one clear difference between a hypothesized position and what is visible in the terrain is sufficient to rule out a hypothesis. This is exemplified by the disconfirmation strategy in 1).

However, acceptance of an hypothesis usually requires more converging evidence (Smith, Heinrichs, & Pick, 1990) and that is one of the results of attending to configurations.

Summary and Conclusions

Solving localization problems with the help of a topographic map is a highly skilled task accomplished by orienteers, geologists, soldiers, etc. as part of their expert activity. The drop-off problem is a particularly difficult variation of the localization problem and even experts are unable to solve it under the constraint of not moving as in one condition of the first experiment above. When permitted to move around success rate jumps to fifty per cent. As noted, the protocols suggest that a major source of error in solving the drop-off problem in the present setting is inaccurate perception of the area around the observation point. Indeed when subjects are permitted to move they focus on acquiring information about the nature of the observation point and are less concerned with obtaining distal information. The same difficulties appeared in the laboratory simulation task in the second and third experiments.

The field protocols of the first experiment indicate similar problem solving activities in both the successful and unsuccessful subjects. These included general reconnaissance, map orientation, feature matching, configuration matching, and hypothesis generation and evaluation. The general reconnaissance activity occurs at the beginning of the task and sometimes later on when starting afresh to formulate new hypotheses. As noted, reconnaissance beginning with the scene is more likely to be efficient and successful, presumably because the scene constrains the search for relevant features on the map more than the converse.

Although orientation of the map is not necessary for subsequent feature matching, most subjects align the map with the environment. This is done on the basis of the general lay of the land as specified by direction of drainage. Such alignment activity is congruent with information processing research on mental rotation which indicates that search and matching would be facilitated (Eley, 1988; Cohen-Cliffer, 1991).

Feature matching involves establishing a correspondence between salient features in the environment and on the map. This does not require a specific hypothesis about one's own location. The salient features include high hills, ridges, depressions, valleys, etc. These are generally described with qualitative and ordinal rather than metric values. Evidence from a companion study of memory for photographic scenes (Montello & Sullivan, in preparation) suggests that experts engaged in map reading tasks find contour features and valleys more salient than non map readers or map readers no motivated by map reading tasks.

Configuration matching is also concerned with establishing correspondences between map and scene but with combinations or clusters of features, usually adjacent and often along the line of sight. The use of configurations reduces the likelihood of accidental correspondences.

In generating hypotheses there is no evidence for any quantitative triangulation processes. It is possible that a crude form of triangulation is being carried out semi-automatically and doesn't appear in the protocols or that a qualitative decision as to which side of a line between a pair of features one is on (cf Levitt, Lawton, Chelberg, Koitzsch, & Dye, 1988) is being used in conjunction with local features (e.g. I am on a hill of a particular shape on this side of this pair of features). Hypotheses are evaluated by making additional predictions about features to be seen on map or in scene. Logically a single bit of negative evidence should disconfirm a hypothesis while confirmation should demand more exhaustive correspondence. However, in the difficult drop-off problem, especially when the opportunity to gain further information by moving is precluded, subjects will often explain away negative evidence if the fit is otherwise reasonably good.

In the laboratory simulation task of the second experiment the full map condition (with least masking) produces the best results. However, it is not just sheer amount of unmasked map available that is crucial since the condition with the outer third of the map masked produces almost as good results. This pattern would suggest, that up to a point, proximal map information closer to the observation point is more valuable than distal information. Of course what information is crucial is dependent on the particular problem setting and the protocol analyses of a subset of the Experiment 2 problems in Experiment 3 help indicate the specific information which is being used for better or worse. The laboratory task protocols yield problem solving strategies similar to those of the field protocols with the exception of hypothesis generation. Since the laboratory task provides three specific hypotheses in each case the evaluation procedures engaged in by the subjects can be more systematically examined which is being done in a future paper.

A final observation is that in neither the field task nor the laboratory simulation is there any evidence for global matching of the scene to map. Subjects seem to do their searching and matching for features or configurations of features. Informal comments by some map users that they look at a map and visualize the overall layout of the terrain is not supported by the present data.

Where am I? Similarity Judgment and Expert Localization

Kip Smith, Marian Heinrichs, & Herbert Pick, Jr.

Center for Research in Learning, Perception, and Cognition
University of Minnesota
205 Elliott Hall
75 East River Road
Minneapolis, Minnesota 55455
kip@umnsom.bitnet

Abstract

How do skilled map-readers use topographic maps to figure out where in the world they are? Our research addresses this question by studying the problem solving of experienced map-readers as they solve localization - Where am I? - problems. Localization relies upon judgments of similarity and difference between the contour information of the map and the topographic information in the terrain. In this paper we discuss experiments that focus on how map-readers use attributes and structural relations to support judgments of similarity and difference. In our field and laboratory experiments, experienced map-readers implicitly define attributes to be detailed descriptors of individual topographic features. They use structural relations that link two or more topographic features as predicates. The time-course of their problem solving suggests that attributes and relations are psychologically distinct. Attributes like slope, e.g., "steep (hill)", support only initial judgments of difference. Relations like "(this hill) falls steeply down into (a valley)" are more powerful, supporting both judgments of difference and judgments of similarity. Judgments based on relations are used to test hypotheses about location. Experienced map readers exploit the distinction between attributes and relations as they solve localization problems efficiently.

Localization - the 'Where am I?' problem in navigation[1]

Localization is the familiar task of finding the point on a map that represents your viewpoint in the world. Anyone who has ever been lost knows that localization can pose a difficult problem. It is a fundamental component of all navigation in large-scale space. Diverse professions (e.g., geology and airborne infantry) require individuals to become skilled at localization. The work reported here elucidates the roles of topographic features, attributes of features, and relations among features in the judgments that establish correspondence between map and terrain.

Maps are representations that preserve with fidelity a selected subset of the information available in a section of the world. The information contained in a map provides a context for the map-reader: localization judgments based upon a map can only be made with reference to the type of information it makes available. We restrict our study to topographic maps because they provide a clear, familiar, and pragmatically useful context for constructing a theory of localization that will assist the design of intelligent systems to control vision-based robot navigation.

When using a topographic map to solve a localization problem, the map-reader must find the location among the contours that matches the *viewpoint* in the terrain. The viewpoint is the location in the world where one happens to be standing. It determines what can be seen and what is occluded. The viewpoint dependence of terrain information tightly constrains problem solving. It determines the topographic features and relations among those features that can be used to generate and test hypotheses about location. Map-readers necessarily relate terrain information to their viewpoint.

The constraint of viewpoint dependance and the context provided by topographic maps transform localization into the task of finding the contours on the map that characterize a layout similar to that seen from the viewpoint. Determining the correspondence between map and terrain relies on judgments of similarity and difference between the contour information of the map and the topographic information in the terrain.

There are frequently many locations on a topographic map that appear similar to the viewpoint. Each may be entertained as a hypothesis. Selecting the hypothesis that provides the best match to the terrain relies on judgments that discriminate among competing hypotheses. Thus, there are two basic steps to localization problem solving: (1) generating hypotheses that relate map and terrain information and (2) testing these hypotheses by identifying the best match. Similarity judgment is essential to both.

Localization and similarity judgment

Gentner (1983) and Medin, Goldstone, and Gentner (1990) emphasize the role of relations among objects in judgments of perceptual similarity. Their notion of structure-mapping holds that the relations among objects constrain judgments of similarity and are, in fact, more central to the process of judgment than are the individual objects themselves. This emphasis on the structure that relations impose on their constituent objects is intuitively consistent with the correspondences that map-readers must make to compare a map to the terrain they see.

Tversky (1977) introduces the notion that judgments of similarity depend on the context of the task in which they are embedded. He specifies a rule for calculating similarity. He implies that application of the rule is dependant on the task context but does not indicate specifically how. Medin et al. (1990) seize on this insight and suggest that the

[1]This work was supported by NSF Grant IRI-8901888 (funded in part by DARPA) and AFOSR Grant 88-0187.

relational structure among objects provides the context missing in Tversky's (1977) model. This too matches the demands of the localization task. Topography provides not only a context but also an intrinsic structure within which a map-reader views both the terrain and the map.

A key component of the structure mapping hypothesis is a fundamental distinction among objects, relations, and attributes. We embrace this distinction. In this paper, we call the topographic objects that capture a map-reader's attention *features*, e.g., "I see a valley". An *attribute* is a property, like gradient, that embellishes the description of a feature, e.g., "I see a steep valley". A *relation* is a connective property that cannot be hung on any one feature; relations span two or more features, e.g., "When I look southeast, I see the ground falls abruptly into a valley". In this example, the relation is a predicate that links the map-reader's viewpoint to a distant feature.

Since localization is a veridical task, individuals who have developed this skill are readily identifiable. They include professionals who make their living finding their way around the world using topographic maps (e.g., geologists and wilderness guides) and serious recreationists (e.g., orienteers and outfitters). By investigating the problem solving of experienced map-readers as they solve localization problems, we gain insight into the methods used in efficient localization problem solving.

These considerations lead us to believe that studies of localization problem solving can shed light on three current issues in similarity judgment: (1) the claim that the structure of relations among features is more vital to these judgments than features taken independently, (2) the utility of making a distinction between relations and attributes, and (3) the processes by which these judgments are made.

Experiment 1: Field studies

The goal of Experiment 1 was to address these issues using as data the thinking-aloud reports (protocols) of experienced map-readers solving a localization problem (Thompson, Pick, Bennett, Heinrichs, Savitt, & Smith 1990).

Method

Subjects. A total of 29 experienced map-readers including professional geologists, champion orienteers, and wilderness guides participated in Experiment 1.

Procedure. Individual subjects were blindfolded and driven approximately 30 miles to a road access about one-quarter mile from the station point: the point in the terrain to be found on the map. They were led across a level field and up a hill to the station point. Once there, the blindfold was removed and they were given a topographic map attached to a clipboard. The map is a cropped U.S.G.S. topographic map from which all non-topographic information (culture) has been deleted. The map contains only contour information about elevation.

The station point is a roughly circular hill that is the westward extension of a larger highland. A distinctive attribute is its steep slope to the southwest. A second round hill with a similar orientation is selected as an alternative hypothesis by all subjects. This hill forms a garden path hypothesis (Johnson, Moen, & Thompson 1988): its similarity to the correct solution and its position in the center of the map lead many subjects to consider it early in their problem solving. Both have a pond to the north. Other alternatives are also considered by most subjects. Some subjects consider as many as eight different alternatives. Selection of the correct solution does not appear to depend on the number of alternatives considered.

The subjects' task was to find their viewpoint on the map. During the drive to the site, they had been briefed on the procedure and instructed to think aloud and to point to what they were talking about as they addressed the task. Subjects spent an average of 45 minutes on the task.

Subjects' verbal reports were recorded as they thought aloud. Simultaneously, their behavior was videotaped to provide information about where they were looking (and pointing) while thinking aloud. The verbal reports were transcribed and the composite audiovisual protocols coordinated and scored. Scoring focused on two aspects of problem solving: on the type of information attended to and how that information was used.

The scoring procedure identifies the source of information, the map or the terrain, and three categories of information - features, relations, and attributes. We define *features* as individual topographic objects that our subjects identify with a familiar count noun, e.g., hill, valley, pond. Each subject's lexicon is small and consistent. The composite lexicon across subjects provides a taxonomy of useful topographic terms. *Attributes* are properties that modify individual features. Subjects tended to use bipolar, qualitative attributes to differentiate among similar features, e.g., narrow or wide, steep or shallow. *Relations* are connectives that conjoin two or more features into a single structural unit we call a *configuration* of features. Some relations are purely topologic connectives, e.g., behind, below. Most configurations are expressed by qualitative predicates, e.g., "and then it (feature 1) gets steep down into (feature 2)". Use of quantitative relations, e.g., higher than, a mile apart, is less common.

Configurations constrain problem solving more effectively than do individual features. For example, there are fewer matches to "a high spot going down steeply to some lakes" than to an individual hill or pond. Distinguishing among features, attributes, and relations is consistent with the arguments made by Gentner (1983) and Medin et al. (1990).

Analysis of the protocols also identifies components of the problem solving process. Three of these processes involve judgments of similarity and/or difference. Localization problem solving is initiated by an extended period of reconnaissance. *Reconnaissance* identifies features, attributes, and relations for subsequent processing. Subjects return to reconnaissance to gather additional information. *Matching* is a form of argument that marshalls evidence that the features or configurations seen in the terrain correspond to those seen in the map, or vice versa. *Hypothesis generation* is an explicit statement about a particular location on the map that may represent the viewpoint. Localization concludes with wholesale acceptance of a hypothesis.

Condition 1. The 17 subjects in the first condition were instructed to remain at the station point as they attempted to solve the problem. They were permitted to move a few feet in turning around. As this task proved extremely difficult, a second condition was introduced.

Condition 2. In the second condition, the 12 subjects were free to walk about and to explore the terrain.

Results

Solution. Of the 17 stationary subjects, only one arrived at the correct solution. Six of the 12 exploring subjects arrived at the correct solution. This difference in performance is significant, $\chi^2 = 5.60$, $p < 0.05$, $df = 1$.

Judgments of similarity and difference. All subjects begin by identifying salient features from the terrain and the map. They may begin with the terrain and move to the map, or begin with the map and move to the terrain. Subjects may identify a large number of features or key on a few salient features. The subject highlighted in Table 1 begins by describing his own position in the terrain as a relatively high area and identifying similarly high areas in the map (lines 4-5). Based on the few features he extracts in the first 25 seconds of reconnaissance, he generates a pair of hypotheses (lines 10-11). One of these hypotheses is the correct location on the map. The second is the garden path hypothesis.

Reconnaissance followed by hypothesis generation is typical of highly proficient subjects. Identification of features appears to be sufficient to generate informed hypotheses. Many subjects spend considerable time identifying features and assembling configurations of features before generating hypotheses. Subjects then proceed to focus on relations and judgments of similarity and difference to evaluate those hypotheses.

Single attributes often provide sufficient information to judge that a map feature cannot stand for a terrain feature. That is, difference judgments are often based on single features. An example of a judgment of difference based upon an attribute is shown in Table 1 (lines 87-90).

As shown in Table 1, the subject follows his generation of hypotheses with the assembly of several configurations, one of which is contained in lines 15 to 19. He conjoins his description of his viewpoint, "a knob", to the "stream valley below" with the predicate "gets pretty steep down into". The steep descent from his knob to the stream valley becomes a structural constraint on similarity judgment.

After assembling several other configurations both in the terrain and the map, he proceeds to attempt to match them. This matching necessarily entails judgments of similarity. One such match is shown in Table 1 (lines 38-43). He begins by reiterating a configuration extracted from the terrain (line 38). He turns his attention to the map to match features constrained by the same relation (lines 41-43). He then judges the two configurations to be sufficiently similar to support the hypothesis.

TABLE 1 SIMILARITY JUDGMENT IN LOCALIZATION

key: 4 - line number; M - map information; T - terrain information

1 Identify features including viewpoint, relations, and attributes. Match features to guide assembly of configurations.	4	T	All right, well I noticed I'm at one of the higher points within this area, so that's important.
	5	M	So I'm first looking on the map, for some higher points on the map.
2 Assemble configurations - descriptions of the topographic layout of relations among features including the viewpoint	15	T	but what I was actually looking at is how steeply the hill drops off.
	16	T	And it's kind of a knob right here we're standing on,
	17	T	and then generally not very steep
	18	T	and then it looks like it gets pretty steep down into a valley to the east.
	19	M	Ok, so I'm looking for the same types of things on the map.
3 Generate viewpoint hypotheses	10	M	Um, for example say, somewhere here (HYPO 1 - CORRECT),
	11	M	or on a hill here (HYPO 2 - GARDEN PATH).
4 Eliminating alternatives using attributes to make judgments of difference	87	M	And, see I'm kind of looking up here (HYPO 3)
	88	M	'cause this also has a hill
	89	T	but um, ... Now straight to the north it should be quite steep
	90	M	So, that doesn't seem likely. (REJECT HYPO 3)
5 Matching configurations using relations among features to make judgments of similarity	38	T	I'm at a high point. Directly north is a fairly flat area and north of that it get's steep and then there's the lake, ok.
	39	T	and I'm trying to match those features with what I see here on the map.
	40	M	Ok, ah, for instance, again. Let's go back to this place (HYPO 1) , ok, so here's a higher area.
	41	M	Here's a generally flat area.
	42	M	Then it goes down steeper here and it looks like there's valley coming through here.
	43	M	And then possibly some lakes or ponds here.
6 Comparing hypotheses using relations to make judgments of difference	47	M	Ok, I still kind of like this area (HYPO 1),
	48	M	But then I was looking up on the map. I also have a high here, (HYPO 2)
	49	M	and with (a pond?), .. that's not very far at all.
	50	M	It just doesn't seem to work well
	51	M	because there's a fairly steep and long gradient here before you get to a flat part
	52	T	and I don't see that where we're standing.

This is a consistent pattern in our protocols. Judgments of similarity are used to support hypotheses. They are also used to compare hypotheses. One such comparison is shown in lines 47-52. In this passage, a map configuration relating a high area and a pond is compared with the terrain configuration stated in line 38. This relation is a suitably strong constraint to reject this alternative.

Summary

The field protocols reveal the critical role of similarity and difference judgments in localization problem solving. Of the six components itemized in Table 1, the final three involve similarity and difference judgments. Topographic relations that link features (including the viewpoint) support both similarity and difference judgments. Attributes of features support difference judgments.

This difference in power between relations and attributes may explain the difference in performance between the stationary and exploring conditions. Subjects who were allowed to explore the terrain had better access to information in general and better information about their viewpoint in particular. They used this information to assemble richer configurations that included the viewpoint. The experimental manipulation cannot distinguish whether the information from the viewpoint or about the terrain at large is the more valuable for successful localization.

Experiment 2: Laboratory studies

Experiment 2 is a laboratory simulation of the localization task in which the amount of map information available to the subjects was manipulated. In an earlier study, maps were masked so as to obscure a portion of the map (Heinrichs, Montello, Nusslé, & Smith 1989). There were three masking conditions in that study: the mask covered either the inner one-third, the outer two-thirds, or the outer one-third of the map. The control condition used an unmasked map. Subjects were presented with one of the four conditions. Five pairs of masked and control conditions were selected for the present experiment.

The goal of experiment 2 was to determine whether map information around the viewpoint is generally more informative than other regions of the map. A second aim was to elucidate better the different roles of relations and attributes. The third aim was to examine a variety of locations in order to generalize beyond the single location used in the field experiment.

Method

Subjects. Ten subjects from the same pool of experienced map-readers participated in Experiment 2.

Apparatus. Subjects were presented with topographic maps of five locations, two in Minnesota, two in New Mexico, and one in Arizona. As in Experiment 1, the maps are enlarged and cropped copies of U.S.G.S. topographic maps with all culture removed. The maps were marked with a single point at the center of the map that identified the location from which color slides were taken. The slides were taken with a camera mounted on a tripod at eye level. The line of sight was horizontal. A complete set of twelve pictures covered the whole 360° panoramic view. For the experiment, either one view or three non-overlapping views were selected for presentation to the subject. Slides were presented on a rear projection screen in a darkened laboratory room. The subject sat 120 cm away from the screen with a reading lamp illuminating the map from behind. A remote control allowed the subject to advance or reverse the slides at will.

Procedure. Subjects solved two types of localization problems. In the first, one arrow was drawn on the map leading from the center point. Subjects used a remote control to view the three slides. Their task was to select the slide that corresponded to the terrain that would be seen looking in the direction of the arrow on the map. In the second type of problem, three arrows separated by 120° were drawn on the map leading from the center point. Subjects were shown only one slide. Their task was to select which of the arrows on the map corresponded to the view of the terrain in the slide. These procedures presented options that subjects could entertain as hypotheses.

As in the field study, subjects were asked to think aloud while solving the problems. Collection of concurrent verbal reports, videotaping, and scoring of the resulting protocols followed the procedures of Experiment 1.

Condition 1. To investigate whether information about the viewpoint is favored over information from more distant areas, map information was selectively masked in the first condition. In four of the five trials (one trial for each set of maps and slides) a black circle masked the inner 1/3 of the map. In the fifth (Arizona) a black annulus masked the outer 2/3 of the map.

Condition 2. As a control condition, subjects also solved the same set of five tasks in a 'full map' condition in which the masks were removed from the maps. Each subject solved the problems twice, first in the masked condition and in the full map condition. This manipulation allowed within-subjects and within-location comparisons.

Results

Solution. Accuracy was significantly better in the full map condition (66%) than in the masked condition(44%), $t(9) = 3.16$, $p < 0.05$. In the full map condition, performance is significantly different from chance, $t(9) = 6.27$, $p < 0.001$, but not in the masked condition.

Judgments of similarity and difference. In this section we compare judgments within subjects and across conditions for three of the five tasks.

In the first task, subjects viewed one slide of rolling terrain typical of southeastern Minnesota. The major discriminating feature in the slide is a prominent hill. Many subjects find the hill so salient that they base their judgments on a match to this feature. They were given a map with three arrows to choose among. In the masked condition, the mask covers the inner 1/3 of the map and totally obscures the prominent hill. One of the three arrows crosses a hill in the unmasked region of the map. Subjects who spend a disproportionate amount of time on the slide select this (incorrect) arrow. The salient hill leads them down a garden path. They ignore information about the relation of distance between the viewpoint and the feature and are led to an incorrect solution.

In the full map condition all subjects select the correct answer. The availability of information near the viewpoint pulls their attention to the distances between the viewpoint and the various hills along the arrows. As only one of these distances is similar to what is seen in the slide, the correct arrow is selected.

In the second task, subjects viewed one slide of mountainous Sonoran Desert terrain and were given three arrows on the map to choose among. In the masked condition, the mask covers the outer two-thirds of the map. This task is unique in that most subjects correctly answer both the masked and full-map conditions.

Subjects focus their attention on the orientation of a series of small ridges and valleys. It is clear that the relation of parallelism among these features (not including the viewpoint) is sufficiently diagnostic to raise only one of the offered choices to the level of a hypothesis.

The full map condition produces a second finding. It reveals a high hill in the distance to the left of one of the arrows. Subjects find that the hills in the slide are not as high and immediately eliminate that arrow from further consideration. This result supports the inference that attributes are sufficient to support judgments of difference.

In the third task, the inner one-third of the map is occluded and one arrow is shown on the map. Subjects viewed three slides of an area adjacent to a large river valley in eastern Minnesota. Their task is to select the slide that contains the terrain they would see looking in the direction indicated by the arrow on the map. The mask covering the viewpoint makes it appear as though the viewpoint is within a valley and the viewing direction is up at a hill. The viewpoint is actually on the crest of a small ridge that is completely obscured by the mask.

One of the slides contains a long gentle slope up to a distant hill. In the masked condition, many subjects make a reasonable assumption and incorrectly select this slide. By occluding the viewpoint, the masked condition eliminates vital information about the distribution and relations among features in the terrain and prevents correct solution. Correct solution of this localization problem clearly requires matching on the basis of relations of features that include the viewpoint.

Summary

The first task shows the superiority of a judgment of similarity based on a configuration over a judgment based solely on a salient feature. The second task also shows the superiority of a judgment of similarity based on a configuration over a judgment based solely on a feature. In addition, it reveals reliance on the attributes of a feature to justify a judgment of difference. The third suggests that successful localization often requires full knowledge of the relations that tie the viewpoint to nearby features.

Discussion

Experienced map-readers adopt a basic generate and test strategy to solve localization problems. They move in either direction, from map to terrain and from terrain to map, as they attempt to figure out where in the world they are. Judgments of similarity and difference inform both the generation and testing of hypotheses.

The two experiments reveal that structural relations of features play a key role in both the generation and testing of localization hypotheses. They also show a fundamental difference in the roles played by relations and attributes. The protocols reveal that experienced map-readers make this distinction. Attributes are used to make preliminary judgments about potential hypotheses (Table 1, Section 4) whereas relations are used to scrutinize hypotheses (Table 1, Sections 5 & 6). Features and relations guide the assembly of configurations. Attempts to match map and terrain configurations inform hypothesis testing. These tests rely on judgments of the similarity of relations. Relations are also used in judgments of difference to discriminate among competing hypotheses. In contrast, attributes are used only for judgments that either eliminate an alternative or raise it to the status of hypothesis.

Three questions remain: In judgments of topographic similarity and difference, are some *relations* more important than others? In the judgments of topographic difference, are some *attributes* more important than others? How do these vary with the nature of the terrain?

The roles played by relations and attributes in judgments of similarity and difference are part of a larger theory of localization problem solving. This theory is to be embodied in a system designed to control the navigation of vision-based mobile robots in dynamic environments.

References

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7:155-170.

Heinrichs, M.R, Montello, D.R., Nusslé, C.N., and Smith, K. 1989. Localization with topographic maps. In Proceedings of the AAAI Symposium on Robot Navigation, 29-32. Stanford, California.

Johnson, P.J., Moen, J. B., and Thompson, W.B. 1988. Garden path errors in diagnostic reasoning. In Bolc, L., and Coombs, M.J. eds. *Expert Systems Applications*. New York: Springer-Verlag.

Medin, D.L., Goldstone, R.L., and Gentner, D. 1990. Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*. 1(1):64-69.

Thompson, W.B., Pick, H.L., Jr., Bennett, B.H., Heinrichs, M.R., Savitt, S.L., and Smith, K. 1990. Map-based localization: The "drop-off" problem. In Proceedings of the DARPA 1990 Image Understanding Workshop, 706-719. Washington D.C.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84:327-352.

Map-Based Localization: The "Drop-Off" Problem

William B. Thompson, Herbert L. Pick, Jr.,
Bonnie H. Bennett, Marian R. Heinrichs,
Steven L. Savitt, and Kip Smith
University of Minnesota
Minneapolis, MN 55455

Abstract

Navigation based on maps requires frequent solutions to the *localization* problem. Localization is the process of establishing a match between particular locations in the environment and the corresponding locations on a map. Most often, localization involves determining the viewpoint and thus the location and heading of the navigating agent on the map. The solution requires both low-level extraction of image and map features and high-level problem solving to establish likely correspondences while avoiding prohibitively expensive search. We present a formalism within which the localization problem can be studied, information about how expert human map users deal with localization, and aspects of a preliminary computational model of the process.

1 Introduction

Localization is the process of establishing a match between particular locations in the environment and the corresponding locations on a map. Commonly, the environment location of interest is the viewpoint and viewing direction (i.e., the "where am I?" problem). Figures 1 and 2 illustrate a typical localization problem. Figure 1 shows a view of Moran Canyon in Grand Teton National Park. (Though we are primarily interested in ground level imagery, this particular example was taken from a helicopter flying approximately 1,300 meters above Jackson Lake.) Figure 2 shows a section of a topographic map which includes both the viewpoint for the picture and much of the terrain visible in the picture. The localization task involves determining the viewing position and direction on the map which corresponds to what is seen in Figure 1. In Figure 2, the true viewpoint location and direction is marked by a ←●.

At an abstract level, localization can be modeled as three interacting processes (Figure 3). Two perceptual processes identify appropriate map and image structures, a third process actually establishes correspondence. Perception needs to operate in both a top-down and bottom-up manner. Operating bottom-up, perceptual components of the process return the location and type of prominent features. Operating top-down, they search the data for features of a particular type at a particular location. In the third process, features which are candidates for matching are found in one set of features and then are searched for in the other set. The matching is bi-directional; that is, map properties can be searched for among image features or image features can be searched for among map features. The search is guided by a priori

Figure 1: View of Moran Canyon.

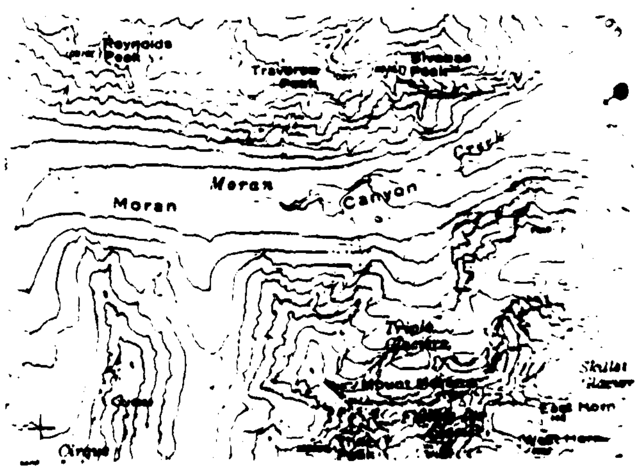


Figure 2: Topographic map of Moran Canyon.

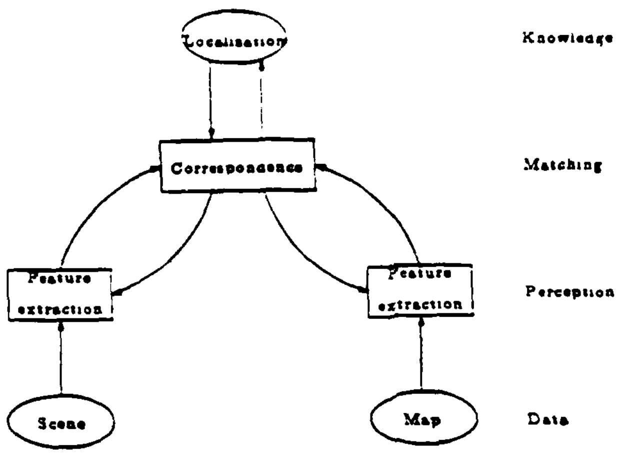


Figure 3: Top-level model of localization process.

knowledge of the likely viewpoint, together with heuristics that reduce the potential complexity. The localization problem itself is solved when correspondence is established between the observation point and a map location. Much of our research is aimed at understanding what features and feature properties are relevant to the perception level and what strategies are used at the matching level to guide the search.

Automated solutions to the localization problem are of obvious utility for mobile robotics in large-scale, outdoor terrain. In addition, a more precise understanding of the processes involved in localization can aid human map users through better training procedures. Finally, outdoor localization provides a challenging research environment within which to advance image understanding technology. Most of the "shape-from-X" techniques that have been developed are ineffective in large-scale, outdoor terrain due to the long distances involved and the complex reflectance models that prevail. New low-level analysis techniques based on occlusion cues and properties such as aerial perspective will be required.

In this paper, we describe a preliminary computational model for solving one type of localization problem. In addition we outline relevant information learned from studies we have done involving expert map users solving a variety of realistic outdoor and laboratory tasks. Subsequent reports will elaborate this model for a broader class of localization problems and show how lower-level vision modules and high-level spatial reasoning need to interact in order to perform localization while navigating outdoors.

2 Approach to the Problem

Localization problems can be characterized in terms of how much a priori information is available about likely observations points (Figure 4). At one end of this continuum, *drop-off* problems involve substantial initial uncertainty in viewing location and/or direction. (The name comes from the extreme case in which an observer is "dropped off" into a totally unfamiliar environment.) In *updating* problems, the task is to maintain a sense of the current position with respect to a map as the current position changes incrementally due to locomotion. We have initially focused our research on drop-off problems, since many of the techniques for solving drop-off problems are likely to be part of the solution to updating problems. In addition, the drop-off problem gives us a sense of base-level performance for map-based localization under a high degree of uncertainty.

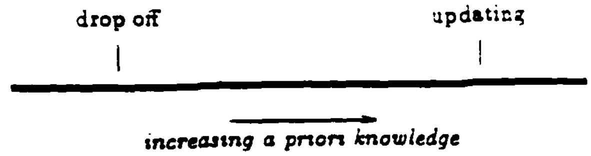


Figure 4: Variations in a priori knowledge affect the nature of the localization process.

Almost all of the previous work on localization using vision has been directed at updating problems. Knowledge of expected position (typically from dead reckoning) is used to predict visual features which are then searched for in the image. Deviations between expected and observed images are used to update the estimate of current location. While updating plays a necessary role in outdoor navigation, it is not sufficient in and of itself to solve the localization problem. Over the long distances and time intervals involved in large-scale outdoor navigation, dead reckoning errors accumulate to prohibitively large values, the maintenance of a visual fix on features necessary for updating becomes increasingly difficult, and dealing with the occlusion and disocclusion of tracked features introduces special problems.

Over shorter time intervals, it may be possible to start with an initial solution to the localization problem, use this to visually identify significant image features and note the corresponding map features, use low-level visual correspondence methods to track these features when moving, then use triangulation techniques to solve for the new current location. Even if this is possible, a continuous 360° view of the scene must be available and substantial computational resources are required.[1] Outdoor environments often have areas in which no distinctive features are visible. In such situations, it is essential that a method be available for reacquiring a sense of location on the map after moving into more varied terrain. Furthermore, low-level visual tracking of topographic features is not as simple as it might at first appear. The irregular shape of most topography together with the frequency of curving slopes presents significant problems. Relatively small movements of the observation point can produce significant changes in appearance of a single feature. Even worse, visually prominent aspects of one topographic feature may smoothly move to another feature as the viewpoint is changed. (E.g., a visual high point may correspond to a particular hill in

[1]Consider a real application: During tank battles, inertial position sensors drift at least one nautical mile per hour. global positioning system (GPS) information may be unavailable or unreliable, and it is clearly not possible for the tank crew to continuously and precisely keep track of all visual changes in the local topography.

one view and a different hill in a subsequent nearby view, without any obvious event in the imagery signaling that a different hill has come into view.) Finally, the frequent occlusion and disocclusion of structures needed for triangulation requires the visual acquisition of new features, presenting an additional possibility for significant error.

Real world topography involves complex shapes at many different scales. Even with an accurate map, the number of characteristic views (nodes in the aspect graph) grows rapidly with increasing uncertainty in viewing location. As a result, the combinatorics of the drop-off problem are such that it is usually not possible to use a verification strategy in which an expected view is matched against actual imagery. Instead, localization becomes more like a recognition problem in which the task is to decide what region of the map can act as a "model" to adequately explain visible portions of the scene.

While people can do object recognition rapidly and with little apparent effort, they have considerably more difficulty with localization problems. Effective utilization of a topographic map appears to combine use of visual skills with substantial problem solving. Localization is a high-level perceptual activity quite different from the recognition tasks that are more commonly studied. This suggests that localization may be an application in which lower-level image understanding techniques and methods from artificial intelligence may be naturally combined. It also suggests that the development of computational solutions for the localization problem can benefit significantly from research on how expert map users solve similar problems.[2]

3 Relationship Between Localization and Recognition

Vision is a process that extracts information about *what* and *where* from an image. Most of the research on higher-level vision has concentrated on recognition tasks. In recognition, the fundamental problem is to identify *what* is in the image. Aspects of the problem involving shape and position (*where*) may be both necessary and difficult, but they are typically subsidiary to the identification process. In contrast, issues of *where* are central to localization.

Many of the computational tools that have proven useful for recognition turn out to be also relevant to localization. Use of such formalisms allows a more formal specification of the localization problem while at the same time highlighting similarities and differences with existing recognition algorithms.

Grimson separates the problem of recognition into three conceptual components: selection of appropriate subsets of image features to match against object models, selection of appropriate object models, and establishment of correspondences between model and image features [Grimson, 1990]. Much research has focused solving the correspondence problem using *pose estimation* or *alignment* techniques in which the correspondence between model and image features is coupled with the estimation of the transformation between model and image coordinate systems (e.g., [Huttenlocher and Ullman, 1987]). Localization involves these same conceptual components, though there are distinct and significant differences.

In outdoor navigation, the relevant "model" is a representation of the topographic features visible from a particular vantage point. Because the number of vantage points is effectively unbounded, we no longer have a set of discrete models. Rather, the needed model of the topography must be assembled adaptively from the map. Severe combinatorial problems will result if this assembly of map features is not carefully constrained.

The selection of appropriate image features for matching is rather more straightforward, since all topographically distinctive visible features are potentially relevant. (In recognition, the image is typically cluttered with a large number of features unrelated to the object to be identified. For localization, the "clutter" is in the models, not in the image.) Proficient map users exploit this fact by driving the generation of hypothesized viewpoints based more on features visible in their view of the scene than on a search through possibly relevant map features. Still, the number of visible scene features usually presents combinatorial difficulties. Success in localization seems to involve organizing these features into easily matchable *configurations*.

Correspondence requires a one-to-one matching between particular subsets of map (model) and image features. Grimson describes this as a constraint satisfaction problem, distinguishing between unary constraints which apply to single pairings of an image feature with a model feature and n-ary constraints which apply to larger sets of pairings. (Grimson actually considers nothing more complex than binary constraints.) In localization, unary constraints consist of equivalent identifications of map and image features (e.g., "hill"), possibly combined with descriptive information about the feature (e.g., "high"). N-ary constraints relate configurations of basic features (e.g., "two hills separated by a saddle").

In object recognition, pose estimation involves the determination of the transformation that will best match a particular model to a given set of image features. For three-dimensional models and two-dimensional image features, this transformation typically involves up to six degrees of freedom: two of translation, one of scale (or equivalently, depth), and three of rotation. Finally, the projection of the transformed model onto the image plane must be determined.

The situation is rather more complex for localization in outdoor environments. Recognition is not based on generic, three-dimensional models. Instead, topography leads to $2\frac{1}{2}$-D models, since the environment can be thought of as a 2-D, horizontal surface that has been

[2] The fact that localization seems to be harder for people than object recognition does not necessarily argue against studying human performance in order to build computational models. Experience with expert systems suggests that it is *easier* to build these programs based on how people solve difficult problems than based on seemingly effortless "common sense", since the processes used to solve the more difficult problems are easier to access experimentally.

distorted out of the plane. A map is in effect a 2-D, downward-looking view of this $2\frac{1}{2}$-D surface. The images on which localization must be based are horizontal-looking views of the same surface. Thus, in matching model (map) to image, we always have a 90° rotation to deal with. This *perspective shift* between downward-looking and horizontal-looking views is quite distinct from the other translations and rotations of the map necessary to establish the viewpoint.

The 90° perspective shift between map and image has important implications for the sorts of lower-level image understanding techniques necessary to support localization. Knowing that the shift occurs constrains, to some extent, the problem of finding the complete transformation which specifies the solution to the localization problem. Unfortunately, the "on end" view of the topographic model together with the difficulty of accurately determining range over long distances using passive vision means that it is not possible to extract a precise quantitative geometric description of the scene from the available images and then match this against the map.

4 Strategies for Localization

Expert map users use six distinct processes in solving localization problems. Competence in all six seems required for effective performance. It is likely that these same procedures will be required in automated systems which solve localization problems without precise a priori information on viewing position.

4.1 Reconnaissance

The purpose of reconnaissance is to gather information prior to the creation and/or evaluation of specific hypotheses about the viewing position. Perceptually distinctive topographic properties of the map or scene that are potentially relevant to establishing map-image correspondences are identified.[3] Reconnaissance involves an examination of either map or image features in isolation. The most successful map users seem to spend most of their reconnaissance time examining image features, organizing the information from the environment into a cohesive representation of features and configurations. Initial reconnaissance focusing on the map seems less successful.

Localization problem solving is almost always initiated by an extended period of reconnaissance. The search is conducted broadly, without any particular focus except as to distinctiveness and relevance of features. Follow-up episodes of reconnaissance generate additional information and can be prompted by three different situations. Acquisition of additional information is common during the evaluation of a hypothesis. The additional information is required whenever the current information is insufficient to establish the hypothesis. Follow-up reconnaissance is also useful during the refinement of a hypothesis that is being accepted. The additional information typically serves to fine-tune the hypothesis. The most common use of follow-up reconnaissance is as a "strategic regrouping" after the rejection of a hypothesis. This regrouping appears to serve the same purpose as the initial extended reconnaissance, the gathering of information required to support the targeting of a new hypothesis.

4.2 Map Orientation

Map orientation involves relating the direction and scale of the map to the visible scene. If an accurate compass is not available, the map is aligned with the general lay of the land. An approximate calibration is established between the scene and the map contour interval and distance scale. Map orientation can occur at a variety of points in the problem solving process. It typically is required only once, unless hypotheses based on a previously determined value are proving hard to verify.

4.3 Feature Matching

The major activity during the localization task is matching features in the image to features in the map or vice versa. Feature matching does *not* require the existence of a hypothesis about viewing location. Such matching can establish possible general correspondences between the image and the map, facilitating the generation of specific hypotheses. Once hypotheses have been established, feature matching plays a key role in evaluation.

Feature matching is based on a common identification and a similar characterization of topographic structures in the map and in the image. Identification is done in terms of a set of labels and properties that is often specific to a particular geologic landform. In the rolling terrain of southeastern Minnesota, the most common features attended to are hills and valleys. Matching for the presence or absence of an individual hill or valley is not particularly diagnostic of location. Accordingly, map users more commonly attend to properties of these features rather that just the existence of the feature. To differentiate among similar hills and valleys, they focus on relative size, elevation, and gradient (steepness).

Most map users tend to impose a bipolar classification system to differentiate properties of features. Features are either large or small, narrow or broad, steep or shallow. Comparison is another common strategy to differentiate features. One feature is said to be larger, broader, or steeper than another.

4.4 Configuration Matching

Configuration matching serves the same purposes as feature matching. The only difference between the two are that the pieces of information that are being attended to are assemblies of features. Configurations are specified in terms of the features of which they are composed and the relationships between those features. These relationships include purely topological descriptions (e.g., behind, in front of, next to), ordinal relations (e.g., taller than), and quantitative properties (e.g., actual eleva-

[3] Criteria for distinctiveness and potential relevance can vary significantly over different landforms. The kinds of features relevant to localization in Minnesota are very different than those relevant to the glaciated topography of the mountains in the western United States. Anecdotal evidence suggests that even expert map users may require adaptation before effectively dealing with novel sorts of terrain.

tion). Expert map users tend to do more configuration matching and less feature matching than do less proficient individuals. The complexity of the configurations is usually relatively small, however, typically involving two to four individual features. Competence in map reading appears to depend on the accurate establishment of appropriate configurations for matching.

Configurations constrain the matching process more effectively than do individual features. There are fewer matches to "a hill with a dip and a ridge" than there are to individual hills, to individual small valleys, and to individual ridges. By bundling features together into configurations, the map user effectively restricts search to models with more unique descriptors.

Experienced map users appear to follow a pair of simple but highly effective heuristics as they assemble configurations of features in the image. The first heuristic restricts configurations to features that are contiguous. Features that are joined together to form configurations are invariably physically adjacent (e.g., "the flat area that slopes down and then up again to a ridge"), rather than just adjacent in the imagery due to occlusion. Map users in the field have often been observed to trace out in the air with a finger the connection between features as they construct a configuration.

The second, less-rigorously applied heuristic, restricts configurations to features that align along a line-of-sight. The majority of configurations (perhaps 80%) used by map users are composed of contiguous features that fall along or parallel to a line-of-sight, along an azimuth that extends away from the viewer. Most of the remaining configurations (the other 20%) focus on the distribution of features along prominent ridge-lines that cut across the viewing angle. The common characteristic of these assemblages is their linearity. Whenever a feature in a configuration does not line up, explicit reference is made to its non-linearity (e.g., the crook in a ridge-line or the slight offset in a string of hills and valleys).

Most configurations are assembled in accord with both heuristics. Both derive their power from the fact that they disallow configurations that could be products of accidental viewpoints. Both connectivity and linearity are viewpoint invariant properties of the image that survive the transformations required for matching. ([Lowe, 1987] emphasizes a similar importance for viewpoint invariant configurations of features in object recognition.)

4.5 Hypothesis Generation and Evaluation

A hypothesis posits a distinct map location and direction as corresponding to the viewing position. The hypothesis is initially triggered by the possible map-image correspondence between a small number of features or correspondences. Hypothesis evaluation proceeds by examining other image and map features or configurations using expectations about correspondences derived from the hypothesis. Often, a brief reconnaissance of a local region in the map and/or image will be required to identify additional features and configurations useful in the evaluation process. The strategies involved have much in common with those used in other diagnostic tasks (e.g., see [Johnson *et al.*, 1988]).

While viewpoint invariance is desirable in the spatial arrangements of features that define a configuration, *viewpoint dependence* is obviously necessary for hypotheses. A hypothesis must necessarily describe the relationship of topographic features to the viewpoint. Our experience with expert map users suggests that they use rather simple, qualitative descriptions for these relationships rather than a more sophisticated trigonometric analysis. Whether this is the best approach to the problem or only a consequence of the difficulty people have in making complex quantitative judgements is not yet clear.

The search through alternate hypotheses can proceed in a variety of ways. A breadth-first strategy, typically not very effective, generates a large number of hypotheses before attempting to evaluate any of them. The generation of each individual hypothesis is based on a small number of features – often only one. More focused search strategies generate successively more precise hypotheses based on increasingly richer sets of configurations. These focused searches may alternate generation and evaluation or may generate a small set of possibilities and then simultaneously examine all at once.

The most common error made by map users is the failure to generate the correct answer as one of the candidates in a set of initial hypotheses. This type of error seems to have as its source an inadequate reconnaissance of the scene in the map user's immediate vicinity. An overly simple description of the location (e.g., "I'm on a big hill" or "This ridge is steep") ends up matching the most prominent "big hill" or "steep ridge" in the map, without concern for the greater constraints that would be provided by a richer set of configurations.

A second type of error is made during the evaluation of a hypothesis. A common evaluation strategy is to examine the map for features or configurations that can be expected in the image if the hypothesized location were correct. If the model of the environment generated from the map is poorly constructed, it is all too easy to "explain away" expectations that are not realized. The source of this type of error is the failure to use the model to identify disconfirmatory evidence in the image. This is an instance of confirmatory bias, a common source of failure in human problem solving [Wason, 1960, Mynatt *et al.*, 1977]. The chance for error is enhanced in the localization task by the inherent imprecision of the model upon which the evaluation is made. This is one situation in which we might expect automated perceptual systems to perform better than their human counterparts.

4.6 Conclude

Hypothesis evaluation leads to the tentative rejection or confirmation of hypotheses that have been generated. A final step in the localization process produces the best estimate of actual location and viewing direction. Depending on the search strategy used, this may be based on a comparison of the likelihoods of competing hypotheses or may simply be the identification of a single hypothesis which survived a sequential generate-and-test procedure.

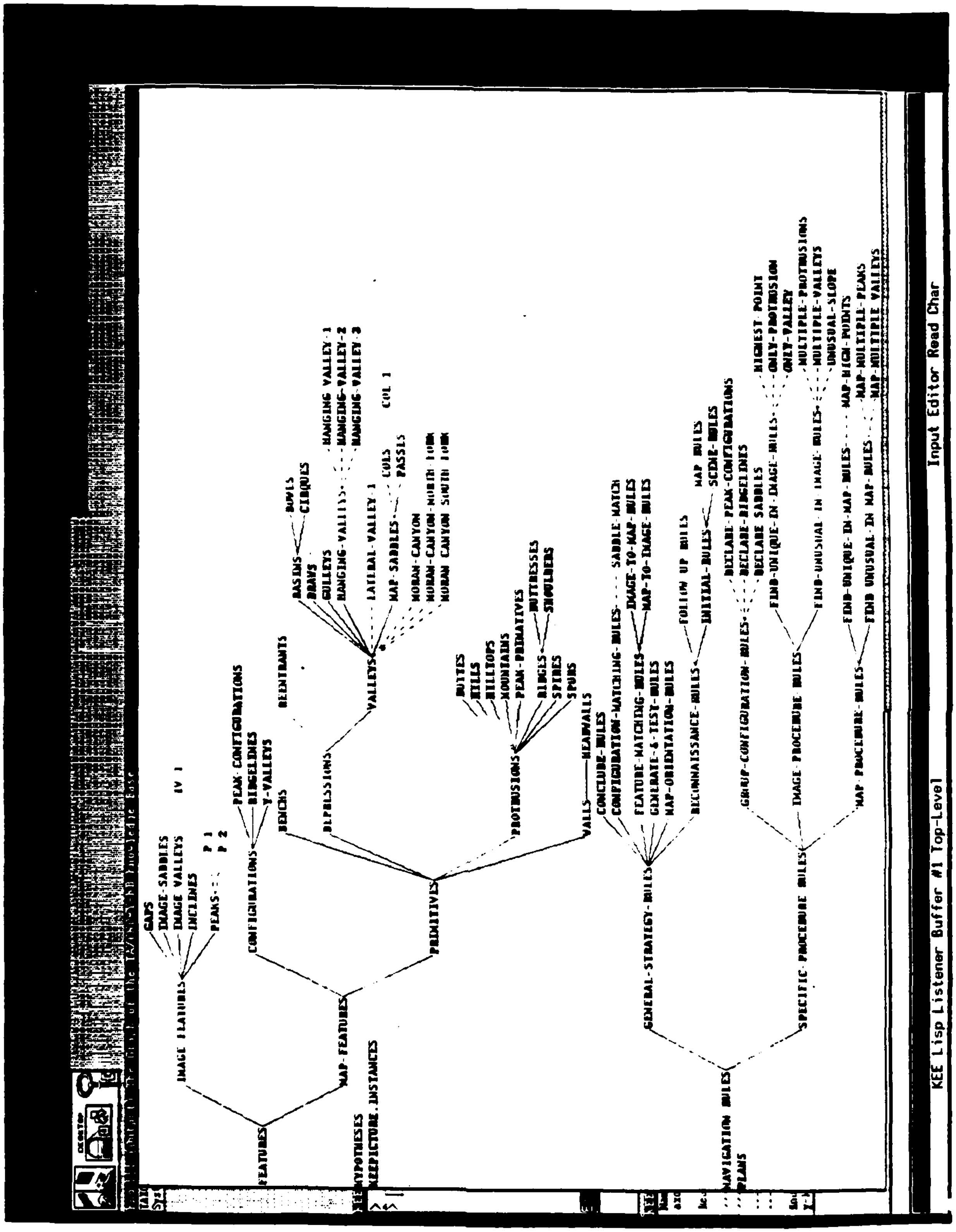


Figure 5: Matching of map and image features.

5 A Computational Architecture for Localization

We have completed the preliminary specification of a computational architecture for the problem solving aspects of drop-off problems. The model includes a *taxonomy knowledge base* for aiding in the recognition of topographic features and the assembly of configurations, *image* and *map knowledge bases* for representing information specific to the problem at hand, and a *hypothesis kn wledge base* for posting information on currently active hypotheses about viewpoint or scene-image correspondences. A set of procedures forms a *control structure* for recognizing features, assembling configurations, and posting, evaluating, refining, and accepting or rejecting hypotheses. In addition, the control structures have access to lower-level components responsible for extracting primitive features from map and imagery.

Figure 5 shows an example of map data partially instantiated against partially interpreted image data. The taxonomy knowledge base is used to create a hierarchy starting with topographic features and continuing on down through the solid (subclass) links to map and image features, primitives and configurations, etc. In this example, the image knowledge base consists of the frames representing two peaks (P–1 and P–2) and a valley (IV–1) which have been recognized in the image. These frames have been attached appropriately into the taxonomy domain by membership (dashed) links. The map knowledge base consists of three hanging valleys (**hanging-valley–1, –2, and –3**), three canyons (**moran-canyon** along with its **–south-fork** and **–north-fork**), and a col (**col–1**). These are attached to appropriate places in the taxonomy hierarchy via more membership links.

Several of the control structure procedures are shown in Figure 5. These procedures are divided into two classes. General strategy rules include reconnaissance (both initial and follow-up), map orientation, feature matching (both image to map, and map to image), configuration matching, hypotheses generation and evaluation, and conclusions. Specific procedures perform tasks such as grouping configurations and attentional processes such as looking for unique or unusual data like prominent high points or unusual configurations.

6 Lower-level Image and Map Understanding

Extraction of map features is aided greatly by the availability of accurate DTM (digital terrain model) data, since the interpretation of contour data involves a number of subtle interpolation problems. If DTM data is available, extraction of features such as peaks, ridges, and valleys can be done using relatively straightforward mathematical operators [Shapiro *et al.*, 1988]. A significant recognition problem remains, however, since important distinctions exist between features in the same class (e.g., a cirque is a very different feature than a canyon, though both are instances of valley features). This is different from the classic object recognition task, since terrain classification depends on sometimes subtle shape properties, not on a geometrically precise object model.

The problem of assembling configurations is difficult not only because the criteria for choosing members of the configuration is seldom clear, but also because there is no obvious way in which to determine the spatial relationships within a configuration. This problem arises because the individual features have spatial extent, thus limiting the degree to which relationships such as "adjacent to" can be effectively utilized. The fact that expert map users organize configurations in a linear structure may be caused, in part, by the need for finding a compact representation of spatial organization within the configuration.

The extraction of image features also suffers from the lack of precise object models. In addition, the primitive structures needed for feature identification are not well defined. As with other image understanding situations, a large number of effects can generate the same or similar patterns on an image. Simple edge detection is clearly not enough as a basis for finding topographically relevant image features. Many open questions remain in this aspect of our research.

The extraction of image features based on edges requires that only edges likely to be due to topographic effects be identified. Two approaches seem promising. One, similar to methods used in other recognition applications, involves organizing local edge elements into larger segments likely to correspond to some meaningful scene structure. (See [Sha'ashua and Ullman, 1988] and [Mohan and Nevatia, 1988] for examples in the domain of object recognition.) The second involves understanding the specific constraints that exist on image edges generated by topographic structures.

Figure 6 provides one example of using information about topography to generate constraints on edges. The figure shows a sketch of a ridge viewed from slightly different directions. In the right view, we see the ridge in profile. In the left view, the ridge is seen more end-on and the faces on both sides of the ridge have become visible. If the topography consists of approximately planar faces, then a ridge can be characterized in terms of its *rise angle* (the angle of the ridge itself relative to horizontal) and the *break angles* of each of the faces (the slope of the face measured along its fall line). For a horizontal viewing direction, the projection process is such that the angle of the ridge as projected into the image is never less than the rise angle. Furthermore, the projected ridge angle in the image for ridges seen in

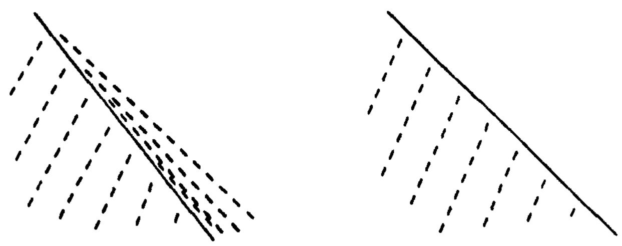


Figure 6: Ridge line seen from different vantage points.

profile is never more than the break angle of the hidden face. Thus, knowledge of the minimum rise angles and maximum break angles that are common in the scene constrain to an interval the projected ridge lines in the image. In most realistic situations, the viewing angle is sufficiently close to horizontal for this effect to be useful. Even in extremely rugged terrain, break angles are seldom more than 45°, thus providing a useful way to evaluate edges in the image.

Figures 7–11 illustrate a number of the lower-level image and map understanding problems that arise in localization. Figure 7 shows topographic features extracted from the DTM data using methods described in [Shapiro *et al.*, 1988]. Black lines indicate ridges, white lines indicate valleys. The features are overlaid onto an elevation image in which lighter values indicate higher altitudes. (Only a portion of the map shown in Figure 2 is shown.) Figure 8 shows the output of an appropriately thresholded Canny edge detector applied to the image in Figure 1. Figure 9 shows the Canny edges which have passed a multi-stage filtering operation involving spatial coincidence across scales, minimum edge length, and expectations about edge orientations. Figures 10 and 11 illustrate how simple textural patterns can aid in the identification of topographic structures. Figure 10 shows image edges filtered to preserve only those oriented down and to the right. Figure 11 shows edges oriented down and to the left. Concentrations of edges in Figure 10 indicate rightward facing slopes. Concentrations of edges in Figure 11 indicate leftward facing slopes. (The smaller clusters of edges in the upper left of Figure 11 are faces associated with the far walls of side valleys branching off from the main canyon.)

7 Related Work

Our research draws on a diverse range of past work. Localization is a fundamental problem in mobile robot navigation, and several different types of solutions have been developed to address it. Approaches for automated interpretation of reconnaissance imagery relate directly to the problems of determining locations in large-scale space. Finally, an extensive literature exists on human competence in map reading.

7.1 Computational approaches

Solutions to the localization problem in mobile robot navigation take two forms. Both approaches match the actual image with the scene that is expected given an estimated location, but differ in the level at which the matching takes place. In the first approach, a 3-D model of the scene and an estimate of the viewing location is used to predict what the 2-D image should look like. Edges from the predicted image are compared with edges found in the actual image. One example of this approach uses map data to project a potential image given a downward-looking perspective from an estimated position [Ernst and Flinchbaugh, 1989]. This potential image is then run through a low-level matcher which compares it to the actual incoming image. The resulting correspondences are then used to refine the estimated position. Another example, the PSEIKI system, also uses an estimated position which is derived from motion information to generate a two-dimensional projection of the structure in the expected scene [Andress and Kak, 1988]. Correspondences are found by using the Dempster-Shafer formalism. The HILARE system, too, uses motion information to estimate position, and explicitly represents positional uncertainty numerically [Chatila and Laumond, 1985]. The part of the world model near the estimated position which best corresponds to what is currently being perceived is then found using a global matching approach.

The second approach to localization in mobile robot navigation matches the expected scene with the actual image using landmarks at the level of objects and places. Distinguishable objects in the environment are identified using perceptual systems. The bearing and range to each landmark is then used to orient the system with respect to a "world model" (i.e., map) of the environment. One example of this approach is the NX robot, which, during an exploration phase, determines locally distinctive places by finding sensory features which are maximized at that place [Kuipers and Byun, 1987]. This *signature* is then used during later navigation to recognize the place. Levitt et al. developed a model of landmark-based localization in which landmarks are used in a highly error tolerant manner to partition the environment into places which are recognized by the landmarks configurations seen there [Levitt *et al.*, 1987, Levitt *et al.*, 1988]. Another method addresses the localization problem by combining low-level tracking or visual "servoing" with high-level perceptual verification using *milestones*. These milestones are defined in terms of landmarks such as buildings, for example, and their bearing [Arkin *et al.*, 1987, Fennema *et al.*, 1988]. Another system generates a 2-D and partial 3-D scene model from the observed scene. The matching problem is then solved by using object groupings and spatial reasoning [Nasr *et al.*, 1987].

Conventional approaches to landmark-based localization require that the identification and global position of landmarks be known a priori with a high degree of precision, and that perceptual systems exist which can accurately identify these landmarks and precisely determine their relative position with respect to the robot vehicle. Object recognition that is at the same time both general and robust is difficult to achieve. As a result, errors in landmark recognition will be common. In many environments, precisely localized landmarks may be scarce. Finally, the ambiguity associated with landmark-based navigation can lead to a combinatorial explosion of cases that must be analyzed. If there are many landmarks of the same type, then the complexity of the task matching landmarks to map features grows quickly.

The integration of sensed data with maps is central to many navigation tasks. Map-to-image matching has been extensively studied within the context of reconnaissance imagery (e.g., [Nevatia and Price, 1982, Clark, 1983, McKeown and Denlinger, 1984, Hwang, 1984, McKeown *et al.*, 1985]). Typically, meaningful features are found in the image and then matched to corresponding map features. Common matching items in-

Figure 7: Extracted topographic features.

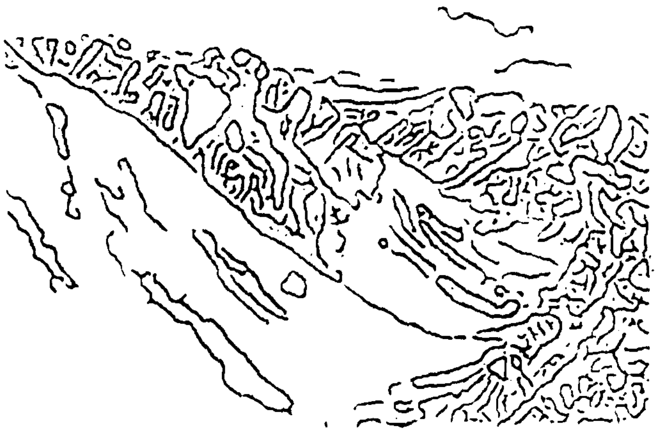


Figure 8: Canny edges from Fig. 1.

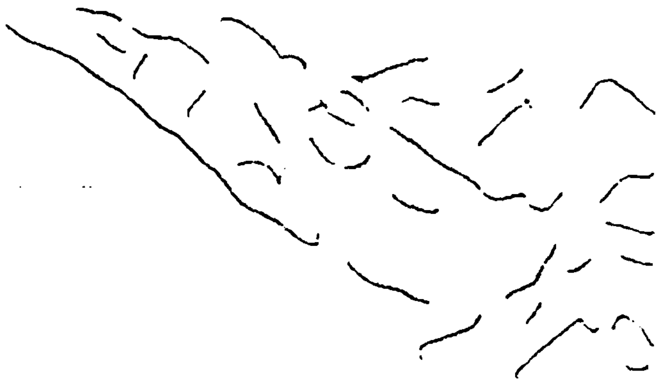


Figure 9: Filtered edges.

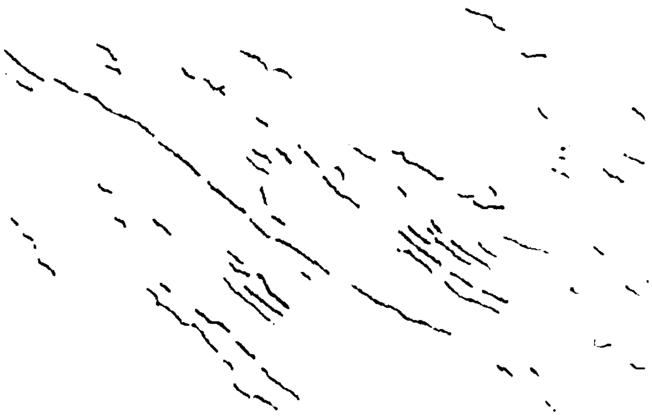


Figure 10: Diagonal edges from Fig. 1.

Figure 11: Edges oriented in opposite diagonal.

clude cultural features such as roads, cities, and airports, along with terrain features such as rivers, coastlines, and so on. [Little, 1982] describes one of the few map-to-image systems that makes heavy use of *topographic* features. Ridge lines are found in a digital terrain model and placed in correspondence with brightness discontinuities in an image. The matching is aided by information about illumination angle which is used to predict which ridges in the elevation model will generate distinctive changes in brightness. In all of these cases, imagery and maps have had a common, "downward-looking" perspective where both imagery and maps have a similar, two-dimensional coordinate system. The correspondence problem is essentially one of 2-D registration.

In the problems we are considering here, imagery has a near "horizontal-looking" perspective which is qualitatively different from the downward view common to nearly all maps. There has been relatively little work relating horizontal-looking imagery with maps. The work closest to our own is that of Lavin who was interested in a problem complementary to that of map matching [Lavin, 1979]. He investigated the creation of topographic maps from sketches of occlusion boundaries. Only a very simple model of topography involving uniform Gaussian shaped hills was used. Thus, many of the complexities encountered in more realistic situations were avoided. Related to both Lavin's work and the methods for matching reconnaissance imagery and maps are techniques for automatically rendering terrain views based on both aerial photography and elevation data [Quam, 1985]. Appropriate coordinate transformations and resampling are done to produce a horizontal-looking view from the original downward-looking photograph.

The perspective shift associated with combining visual data with other representations such as maps is related to several other three-dimensional reconstruction problems. Koenderink developed a relationship between the 3-D structure of solid objects and the topology of projected contours [Koenderink, 1984]. Giblin and Weiss describe how surface descriptions can be recovered from projected contours [Giblin and Weiss, 1987]. Neither of these approaches, however, is directly applicable to our problem. Complex terrain cannot be modeled as a simple, solid object. Furthermore, the inaccuracies of lower-level image analysis algorithms is likely to defeat any method based on the topology of projected contours. Finally, Shepard's work on mental rotations may provide some insight into human performance in perspective shift tasks [Shepard and Metzler, 1971].

7.2 The psychology of using maps

An extensive literature in psychology and cartography deals with problems associated with reading and using maps and the associated problem of recognizing aspects of scene geometry relevant to localization. While little of this literature deals with the actual processes involved in localization, it does provide useful insight into the sorts of computational models likely to be effective. Knowledge about the performance of expert map users can aid in understanding the heuristic strategies necessary to establishing correspondences between a map and an image. Lower-level image understanding methods which utilize passive vision are unlikely to work much better than humans. As a result, information about the limitations of human vision in large-scale, outdoor environments is potentially of great relevance in developing computational solutions for vision-based localization.

Preliminary research of ours using both protocol analysis of observers thinking aloud as they solved a drop-off localization problem and a memory paradigm of observers recalling photographic images suggested that much attention during observation of natural scenes may be devoted to qualitative topographic features [Heinrichs *et al.*, 1989]. Certainly there was more mention of such features than precise metric characteristics. Among the kinds of features noted were a variety of convex features (hills, ridges, rises), concave features (valleys, sinks, holes, etc.), inclinations (level plateaus and slopes). Although the organization of spatial knowledge has mainly been studied in urban or restricted laboratory environments the indications are that features or landmarks exert a strong influence on one's use of spatial information. For example, [Sadalla *et al.*, 1980] demonstrated that certain salient features serve as reference points for organizing spatial information. Once established, these reference points have a privileged role in spatial orientation, with one result being that the subjective distance between reference points and non-reference points is not symmetrical.

Other research has shown that spatial information is hierarchically organized. This is evidenced by the fact that making judgments (or thinking about) particular locations will facilitate subsequent independent judgments about locations that are physically nearby [Hirtle and Jonides, 1985, McNamara, 1986]. Another factor which contributes to such hierarchical organization is the extent to which various physical factors compartmentalize a space [Kosslyn *et al.*, 1974]. Distances between locations within the same subspace will often be judged as smaller than equivalent distances between locations in different subspaces. These subspaces might be defined by physical barriers such as rivers or fences, by optical barriers such as the edge of a field, or by political boundaries such as state or city lines.

Analysis of individual differences in map reading performance has also been used as a way of investigating the processes of extracting information from maps. [Chang *et al.*, 1985] studied how eye movements during reading topographic maps were related to individual differences in map reading experience. They found that the eye fixations of experienced map readers were shorter and more often focused on task relevant areas than those of inexperienced readers. Sholl and Egeth [Sholl and Egeth, 1982], in a systematic psychometric approach, related performance on a number of topographic map performance tasks to several more general standard psychometric measures. The map tasks such as land form identification, slope identification, spot elevation, and terrain visualization were factor analyzed, yielding two major factors, one described as a spatial visualization factor and the other an altitude estimation factor. Surprisingly, standard tests of spatial ability are not highly related to the spatial visualization map reading factor whereas verbal-analytic measures are. A standardized measure of mathematical ability is related to the altitude estimation factor, yet finding the altitude of points on a topographic map or finding the highest and lowest elevations wouldn't seem to involve very sophisticated mathematics. The authors suggest that the relationship is due to the arithmetic aspect of mathematical ability. In general, the results seem to suggest that our standardized tests don't reflect very well the abilities used in a practical skill like topographic map reading.

An obvious approach to understanding the processes underlying extraction of information from topographic maps is the use of information processing paradigms. There are few such studies, but one example [Eley, 1988] has examined the effect of differences in orientation in view point on speed of matching a map position to the topography of a surface. Subjects were shown a segment of a topographic map for inspection. After they had a chance to study the map, a point and direction of view was indicated on the map perimeter. Their task was then to imagine what the land surface would look from that perspective. When they were satisfied that they knew how the surface would look they pressed a button which presented a representational drawing of a surface. They then had to indicate whether the surface drawing corresponded or not to the specified view. Of particular interest was how the time required to imagine the view from the specified orientation was related to the viewing direction. Typical mental rotation results were obtained. The greater the required viewing direction deviated from the subject's own orientation the longer the reaction time to press the button for the drawing. In a second experiment reaction time was measured for land surface views at different elevations. Results indicated that an elevation providing a viewpoint of 30 degrees above horizontal was more effective than either higher or lower elevations. The effects on map reading performance of the mismatch in orientation between map and environment has also been found with street maps [Levine *et al.*, 1984].

Although space perception has been a topic of study for over one hundred and fifty years only so-called depth perception, the perception of the radial distance of objects from the observer, has received systematic intense investigation [Haber, 1985]. Psychophysical research has been concerned with how observers are able to obtain information about a 3-D world from 2-D sensory input. The few studies conducted in rich outdoor environments have suggested that a linear relationship exists between perceived and physical distance for spaces relevant to navigation. Unfortunately, all of these studies were done in flat open fields. No such studies have been carried out on even sloping or irregular (not to mention cluttered) landscapes.

Laboratory studies of the perception of the slant of surfaces indicate reasonable sensitivity to relative inclination as specified by optical texture and linear perspective (e.g., [Flock, 1965]). However, there is only one report of observation of the slope of a natural incline and that suggested that frontally viewed slopes were seen as steeper than they really were [Smith and Smith, 1965]. This result is consistent with anecdotal reports of hills often appearing steeper than they actually are when one is traversing them by foot or in a vehicle. In work preliminary to the present project, slopes were estimated from photographs at points for which the actual slopes varied from about 3 to 25 degrees. Results indicated a linear relationship between actual and perceived slope. Consistent with the observation by Smith and Smith slopes were perceived as steeper than they actually were.

The limited research that exists on reading of topographic maps is interesting and tantalizing. The results suggest a rather sophisticated skill, but neither an analysis of individual differences nor of tasks processes provides and adequate understanding of the nature of that skill. One reason is simply that there is relatively little research. Another is that the tasks used are artificial in two respects. The materials used are not realistic. The samples of maps themselves are real but often only very small segments are used. When the experimental tasks involve relating maps to the environment, the environment is typically represented by relatively impoverished sketches which may, on the one hand, emphasize features that wouldn't be as clear with natural terrain or, on the other hand, omit the incredible richness of natural terrain. The tasks are also artificial in the problems posed. Subjects may be asked only to find a high or low spot, to judge the qualitative nature of a land form, etc., and they are usually not even asked to solve a localization problem.

8 Implications For Training

A better understanding of the formal nature of the localization problem and the processes likely to be successful in solving localization problems has the potential for improving the training of map users. Knowledge about the perceptual limitations leading to localization errors can be used to warn map users of potential difficulties. Search and evaluation strategies which reduce the combinatorics and minimize ambiguity can be taught, while strategies known to be less effective can be avoided.

Map reading problems take a variety of forms. Localization tasks such as updating and drop-off problems involve map-image correspondence. Some other tasks focus solely on maps. These would include route planning, determination of intervisibility ("when looking from point A to C, would intermediate point B be visible?"), finding highest and lowest station points in an area, determining the direction of water flow, etc. Accuracy and efficiency in reading maps is important for both kinds of problems and accuracy and efficiency in perception of the scene is a necessary prerequisite for the correspondence problem. In addition, solving the map-image correspondence problem requires use of a variety of information processing and problem solving strategies. Establishing such a correspondence involves relating a two-dimensional plan perspective with an encoded third dimension to an eye-level view of a three-dimensional environment.

How accurate is our perception? As noted in section 7.2, the perception and memory of scene and map information is subject to a variety of distortions. Recall the evidence that slope of inclines is over-estimated and that distances between locations in different subspaces are over-estimated. Heights of hills and mountains can also be misperceived. Erroneous judgments of the relative heights of distant and nearer peaks may be caused by not properly taking into account one's own altitude and misperceiving whether one's own direction of gaze is above or below eye level. Such an error may have been a factor in a military plane crash [Haber, 1987].

Similar distortions occur in processing of map information (e.g., [Tversky, 1981, Tversky and Schiano, 1989]). For example, people tend to remember map features as more aligned than is in fact the case. In one case Tversky demonstrated that people will remember continents such as North and South America as more aligned with the cardinal axes of maps than they actually are. Thus, South America is considered to be almost directly south of North America. Such distortions can account for further erroneous judgments such as New York being typically judged as east of Santiago while in fact it is west. Similar distortions occur with more local features, such as city streets. In addition, features that are diagonal tend to be rotated toward cardinal frame axes and are remembered more nearly parallel or perpendicular to major features.

Where do problems arise in the process of solving map-image correspondence problems? On the basis of background literature and our prior work done related to this project, it has been possible to identify some problematic aspects of the solution process. Recall the studies mentioned above that indicate misalignment between map and scene increase the difficulty of the map reading problem. Orienteers are trained in always aligning their map to the scene as they traverse a course. They have found that this increases the efficiency of their map following when time is a premium and helps to reduce errors. It would be easy to demonstrate to the trainees the effects of misalignment between map and scene.

In our initial empirical work on map reading, protocols were collected from persons solving drop-off localization problems. Analysis of these protocols suggests that for drop-off problems a successful strategy is to work from the visible scene to the map. Apparently, specifying the scene features and configurations of features constrains the areas on the map that need to be examined. When this strategy is not successful, one reason is that the local features around the station point are misperceived. Trainees should be alerted to this danger. We observed a number of problematic strategies. One of the most frequent was a "garden path" kind of error in which attention was focused on one or on a very few possible solutions. Incorrect hypotheses were pursued over a long chain and disconfirming evidence was discarded or explained away.

In general trainees can be apprised of both successful strategies and procedures that are likely to lead to trouble. Trainees can be drilled on such problems and their errors pointed out. Unfortunately, field problems are very time consuming. Simulated problems in the classroom are a possibility [Barsam and Simutis, 1984]. However, the simulations need to be developed carefully. In one attempt to develop a laboratory analog to the actual drop-off problem using photographic images we found that the simulation distorted the process by eliminating some of the early stages of problem solving.

9 Discussion

Localization, particularly localization involving drop-off problems, fits well into the conceptual formalism that has been used for several successful approaches to object recognition. The most significant difference is that for localization, predefined object models are not available. Instead, drop-off problems require that models of the scene be created from information supplied on maps. This is possible only after preliminary hypotheses about viewing position and direction have been generated. (Updating problems are easier, in part, because the task of assembling models is much more straightforward.) The lack of predefined object models introduces significant added complexity over that involved in object recognition. This complexity can be overcome by the use of heuristic search strategies which combine sophisticated problem solving with more traditional perceptual processing.

Our formalism predicts the desirability of focusing the search based on an initial reconnaissance of the image before any exploration of the map occurs. This strategy is in fact often observed in expert map users. An interesting contrast occurs with localization problems involving a rapidly moving observer. Before the availability of more sophisticated navigation aids, fighter pilots were trained to do localization by first checking a stopwatch to determine the time spent on the current leg of the flight plan, then estimating their current location on the map and looking for distinctive map features, and finally attempting to visually locate those features in the environment [Ullman, 1990]. In our terminology, this corresponds to an initial reconnaissance focusing on the map – a sensible strategy when elapse time provides an initial guess as to position and the imagery is changing at a substantial rate.

As with alignment methods for object recognition, localization involves the recognition of viewpoint invariant configurations of features. Tentative correspondences between such configurations in map and image data can be established prior to the generation of hypotheses about the viewpoint defined transformation between map and image.

Future work will concentrate on strategies for additional types of localization problems and low-level computer vision requirements for localization. Segmentation algorithms tuned to outdoor scenes are required, as are techniques for recognizing topographic features such as peaks, ridges, and valleys in an image. The ability to actively move the view point will be explored, since an active observer can better determine scene properties such as slopes, while at the same time moving to distinctive positions that aid in the generation of viewpoint hypotheses.

Acknowledgments

Elizabeth Stuck provided important criticism on earlier drafts of this paper. Ting-Chuen Pong contributed to the detection of topographic features from DTM data. Some of the results are based on work by Daniel Montello and Catherine Sullivan.

References

[Andress and Kak, 1988] K. Andress and A. Kak. Evidence accumulation and flow of control. *AI Magazine*, 9(2):75–94, 1988.

[Arkin *et al.*, 1987] R. C. Arkin, E. M. Riseman, and A. R. Hanson. AuRA: An architecture for vision-based robot navigation. In *Proc. DARPA Image Understanding Workshop*, pages 417–431, Los Altos, CA, February 1987. Morgan Kaufmann.

[Barsam and Simutis, 1984] H. F. Barsam and Z. M. Simutis. Computer-based graphics for terrain visualization training. *Human Factors*, 26:659–665, 1984.

[Chang *et al.*, 1985] K. T. Chang, J. Antes, and T. Lenzen. The effects of experience on reading topographic relief information: Analyses of performance and eye movements. *The Cartographic Journal*, 22:88–94, 1985.

[Chatila and Laumond, 1985] R. Chatila and J.-P. Laumond. Position referencing and consistent world modeling for mobile robots. *1985 IEEE International Conference on Robotics and Automation*, pages 138–145, March 1985.

[Clark, 1983] J. Clark. Integration of imagery and cartographic data through a common map base. *Proc. SPIE*, 359:163–168, 1983.

[Eley, 1988] M. G. Eley. Determining the shapes of land surfaces from topographical maps. *Ergonomics*, 31:355–376, 1988.

[Ernst and Flinchbaugh, 1989] M. D. Ernst and B. E. Flinchbaugh. Image/map correspondence using curve matching. In *AAAI Symposium on Robot Navigation*, pages 15–18, March 1989.

[Fennema *et al.*, 1988] C. L. Fennema, Jr., E. M. Riseman, and A. R. Hanson. Planning with perceptual milestones to control uncertainty in robot navigation. In *Proceedings of the SPIE*, 1988.

[Flock, 1965] H. R. Flock. Optical texture and linear perspective as stimuli for slant perception. *Psychological Review*, 72:505–514, 1965.

[Giblin and Weiss, 1987] P. Giblin and R. Weiss. Reconstruction of surfaces from profiles. In *First Int. Conf. on Computer Vision*, pages 136–144, Washington, DC, June 1987. Computer Society Press of the IEEE.

[Grimson, 1990] W. E. L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints.* The MIT Press, 1990.

[Haber, 1985] R. N. Haber. Toward a theory of the perceived spatial layout of scenes. *Computer Vision, Graphics, and Image Processing*, 31:282–321, 1985.

[Haber, 1987] R. N. Haber. Why low-flying fighter planes crash: Perceptual and attentional factors in collision with the ground. *Human Factors*, 29:519–532, 1987.

[Heinrichs *et al.*, 1989] M. R. Heinrichs, D. R. Montello, C. M. Nusslé, and K. Smith. Localization with topographic maps. In *Proceedgins of the AAAI Symposium on Robot Navigation*, pages 29–32, March 1989.

[Hirtle and Jonides, 1985] S. C. Hirtle and J. Jonides. Evidence of hierarchies in cognitive maps. *Cognitive Psychology*, 13:208–217, 1985.

[Huttlenlocher and Ullman, 1987] D. P. Huttlenlocher and S. Ullman. Object recognition using alignment. In *First Int. Conf. on Computer Vision*, pages 102–111, Washington, DC, June 1987. Computer Society Press of the IEEE.

[Hwang, 1984] S. S. V. Hwang. *Evidence accumulation for spatial reasoning in aerial image understanding.* PhD thesis, University of Maryland, December 1984.

[Johnson *et al.*, 1988] P. E. Johnson, J. B. Moen, and W. B. Thompson. Garden path errors in diagnostic reasoning. In L. Bloc and M. J. Coombs, editors, *Computer Expert Systems*, pages 395–427. Springer-Verlag, 1988.

[Koenderink, 1984] J. J. Koenderink. What does the occluding contour tell us about solid shape? *Perception*, 13:321–330, 1984.

[Kosslyn *et al.*, 1974] S. M. Kosslyn, H. L. Pick, Jr., and G. R. Fariello. Cognitive maps in children and men. *Child Development*, 45:707–716, 1974.

[Kuipers and Byun, 1987] B. Kuipers and Y. T. Byun. A qualitative approach to robot exploration and map learning. In *Proc. of the IEEE Workshop on Spatial Reasoning and Multi-Sensor Fusion*, pages 390–404. Los Altos, CA, 1987. Morgan Kaufmann.

[Lavin, 1979] M. A. Lavin. Analysis of scenes from a moving viewpoint. In P. H. Winston and R. H. Brown, editors, *Artificial Intelligence: An MIT Perspective*. pages 17–82. MIT Press, 1979.

[Levine *et al.*, 1984] M. Levine, I. Marchon, and G. Hanley. The placement and misplacement of you-are-here maps. *Environment and Behavior*, 16(2):139–157, 1984.

[Levitt *et al.*, 1987] T. S. Levitt, D. T. Lawton, D. M. Chelberg, and P. C. Nelson. Qualitative navigation. In *Proc. DARPA Image Understanding Workshop*, pages 447–465, Los Altos, CA, February 1987. Morgan Kaufmann.

[Levitt *et al.*, 1988] T. S. Levitt, D. T. Lawton, D. M. Chelberg, K. V. Koitzsch, and J. W. Dye. Qualitative navigation II. In *Proc. DARPA Image Understanding Workshop*, pages 319–326, Los Altos, CA, April 1988. Morgan Kaufmann.

[Little, 1982] J. J. Little. Automatic registration of Landsat MSS images to digital elevation models. *Proc. of IEEE Workshop on Computer Vision: Representation and Control*, pages 178–184, 1982.

[Lowe, 1987] D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3):355–395, 1987.

[McKeown and Denlinger, 1984] D. M. McKeown, Jr. and J. L. Denlinger. Map-guided feature extraction from aerial imagery. *IEEE Workshop on Computer Vision: Representation and Control*, pages 205–213, 1984.

[McKeown *et al.*, 1985] D. M. McKeown, Jr., W. A. Harvey, Jr., and J. McDermott. Rule-based interpretation of aerial imagery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-7:570–585, 1985.

[McNamara, 1986] T. P. McNamara. Mental representations of spatial relations. *Cognitive Psychology*, 18:87–121, 1986.

[Mohan and Nevatia, 1988] R. Mohan and R. Nevatia. Using perceptual organization to extract 3-d structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:1121–1139, 1988.

[Mynatt *et al.*, 1977] C. R. Mynatt, M. E. Doherty, and R. D. Tweeny. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29:85–95, 1977.

[Nasr *et al.*, 1987] H. Nasr, B. Bhanu, and S. Schaffer. Guiding an autonomous land vehicle using knowledge-based landmark recognition. In *Proc. DARPA Image Understanding Workshop*, pages 432–439, Los Altos, CA, February 1987. Morgan Kaufmann.

[Nevatia and Price, 1982] R. Nevatia and K. E. Price. Locating structures in aerial images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-4:476–484, 1982.

[Quam, 1985] L. H. Quam. The Terrain-Calc system. In *Proc. DARPA Image Understanding Workshop*, pages 327–330, Los Altos, CA, 1985. Morgan Kaufmann.

[Sadalla *et al.*, 1980] E. K. Sadalla, W. J. Burroughs, and L. J. Staplin. Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5):516–528, 1980.

[Sha'ashua and Ullman, 1988] A. Sha'ashua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *Second Int. Conf. on Computer Vision*, pages 321–327, Washington, DC, December 1988. Computer Society Press of the IEEE.

[Shapiro *et al.*, 1988] L. G. Shapiro, R. M. Haralick, and T. C. Pong. The use of the facet model and the topographic primal sketch in image analysis. In C. Brown, editor, *Advances in Computer Vision*, volume II. Lawrence Erlbaum Associates, 1988.

[Shepard and Metzler, 1971] R. N. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 171:701–703, 1971.

[Sholl and Egeth, 1982] M. J. Sholl and H. E. Egeth. Cognitive correlates of map reading ability. *Intelligence*, 6:215–230, 1982.

[Smith and Smith, 1965] O. W. Smith and P. C. Smith. An illusion of slant in nature. *Perceptual and Motor Skills*, 20:1108, 1965.

[Tversky and Schiano, 1989] B. Tversky and D. J. Schiano. Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118:387–398, 1989.

[Tversky, 1981] B. Tversky. Distortions in memory for maps. *Cognitive Psychology*, 13:407–433, 1981.

[Ullman, 1990] S. Ullman, 1990. Personal communication.

[Wason, 1960] P. C. Wason. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12:129–140, 1960.