Vision-Based Navigation*

William B. Thompson
Department of Computer Science
University of Utah
Salt Lake City, UT 84112

Herbert L. Pick, Jr.
Center for Research in Learning
Perception, and Cognition
University of Minnesota
Minneapolis, MN 55455

Abstract

Navigation in outdoor terrain is difficult due to a lack of easily and uniquely identifiable landmarks. This paper outlines current research on extraction of navigationally salient features from images and maps, feature matching and viewpoint determination, landmark selection, detection and diagnosis of route following errors, perceptual issues related to vision-based navigation, and database and software availability.

1 Introduction.

Navigation involves two closely related tasks: *localization* and *route planning and following*. Most often, a navigating agent has available a map or some other model of the environment within which it is operating, together with sensor data about relevant aspects of that environment at the current instant in time. Localization finds the agent's position within the map or model frame of reference. Route planning involves the determination of a sequence of actions aimed at accomplishing some goal. This may be based in part on sensor data or completely on the map or model if they are sufficiently rich. Route following includes those processes which execute the plan and monitor for errors. These activities must be closely integrated. For example, accurate localization estimates are needed for route planning since an initial position is usually required and for route following to provide closed loop control of position.

Image understanding approaches to localization must necessarily contain three parts: *feature extraction*, *matching*, and *viewpoint inference*. Feature extraction involves the detection of salient patterns in both sensed data and the map or model. Extracted features are then matched, establishing a correspondence between the two frames of reference. Finally, this correspondence is used to place the viewpoint in the map/model frame of reference. At least in principle, these steps are relatively straightforward when downward looking aerial imagery is matched against a standard "plan view" map. (TERCOM is a classic example [Andreas *et al.*, 1978]). Features can be either raw data or simple, derived point or contour properties. Matching is essentially 2–D correlation. Viewpoint determination involves standard methods from photogrammetry.

Localization is much more difficult when performed at or near ground level due to the 90° change in perspective from sensed data to map. Passive image understanding techniques are likely to have serious problems estimating range to environmental features and thus the relative position of those features to each other and to the viewpoint in the map frame of reference. More sophisticated feature extraction and matching is required and viewpoint determination methods must be able to to function in the absence of accurate 3–D information from sensors. We have made progress in the following areas:

- *Feature extraction:* Domain specific feature extraction routines have been demonstrated which exploit constraints imposed by the geometry of terrain.
- *Matching and viewpoint determination:* Higher-level symbolic problem solving has been integrated with lower-level computer vision methods to produce an image understanding system capable of dealing with inference and ambiguity in localization.
- *Landmark selection:* Path following is significantly aided by selecting landmarks which minimize localization errors.
- *Diagnosis and recovery:* AI-like problem solving methods can complement lower-level computer vision in detecting failures in route following and diagnosis where the original error occurred.
- *Perceptual issues:* An understanding of the abilities and limitations of human perception of terrain features can give insights into the construction of automated navigation and also lead to better training methods.
- *Database:* Examples of panorama images registered to digital elevation data together with a variety of useful software tools are being made available to the research community.

Results of this work are of potential relevance to autonomous and semi-autonomous mobile vehicles, navigation aids, mission planning, simulation, and training.

*This work was supported by National Science Foundation grant IRI-9196146, with partial funding from the Defense Advanced Research Projects Agency.

2 Localization.

Our work on navigation has focused primarily on problems involving outdoor, unstructured terrain. Figure 1 shows typical feature correspondences that must be established. Since distinctive cultural landmarks are not available in such environments, considerable difficulties can be expected in reliably associating map and view features. One way to approach this problem is to use symbolic matching by first independently extracting from the view and map patterns likely to represent the same topographic features and then establishing correspondences using a hypothesize and test strategy.

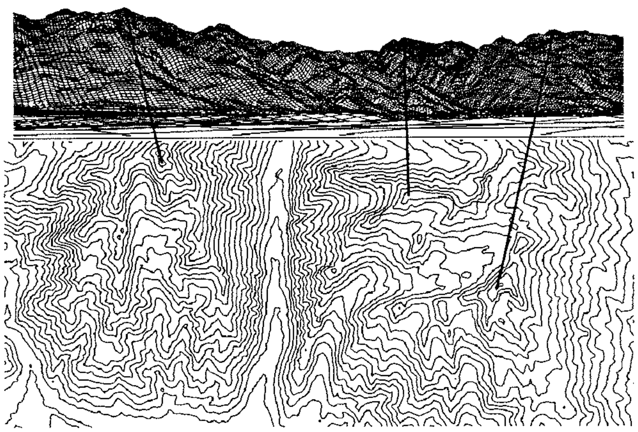

Figure 1: Correspondence between map and view.

Feature extraction from images of outdoor terrain is based on finding ridge contours with shapes indicative of peaks, saddles, and valleys. Peaks and saddles are simply vertical extrema in ridge line contours. Valleys are more difficult to find, since the actual valley terrain is usually not visible and must be inferred from other features such as T–junctions in ridge line contours.

Simple edge detection alone is not sufficient to find ridge contours in an image. Images of large-scale, outdoor terrain contain many important but indistinct features and many extraneous features which convey no useful information about the topography. The contrast across ridge contours is often low and of limited spatial extent. Often, local sections of a ridge contour are lacking in contrast variation altogether, while many non-ridge, high-contrast features are present.

Figure 2 shows a 40° portion of the panorama image shown in Figure 11. Figure 3 shows the results of applying a zero-crossing edge detector to this image. Hysteresis thresholding was used and parameters were carefully matched to the nature and scale of the image. As a result, this represents about the best that can be expected from edge detection alone. Figure 4 shows a new edge image in which a variety of filtering and gap filling steps have been applied. These steps are based on exploiting constraints about how ridge lines appear in horizontally-looking views of rugged terrain. Finally, Figure 5 shows extracted features and line segments.

Figures 6 and 7 show similar results for map features. Unlike the problem of extracting topographic structure

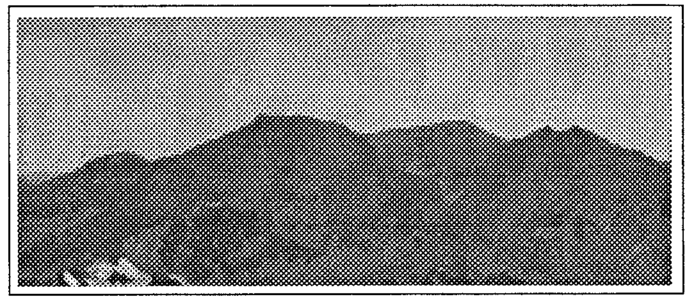

Figure 2: Original image.

Figure 3: Output from zero-crossing edge detector.

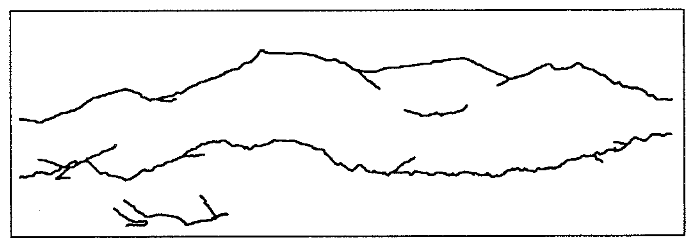

Figure 4: Processed edges.

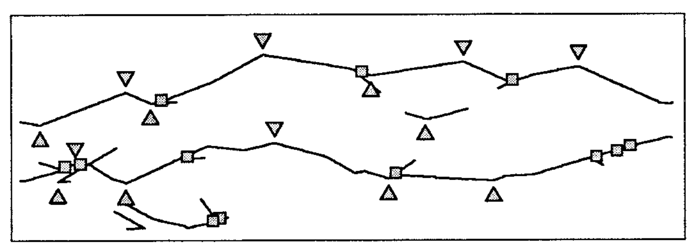

Figure 5: Extracted features.

from images, the "map-understanding" problem does not have to deal with the multitude of effects that can lead to contrast variation in images. Difficulties associated with scale are still very real, however. For example, ridge lines have a large spatial extent along their length. Across the length of a single ridge line, extent can vary from small (a sharp section of ridge) to quite large (a section where the ridge top is essentially a plateau). Peaks are likewise more difficult to accurately detect. Simply finding local maxima in elevation is not sufficient. Figure 6 shows the results of applying a local ridge detector similar to [Haralick *et al.*, 1983] to a portion of our elevation database (see section 6). Figure 7 shows the final results of feature extraction after thinning the raw results and filling gaps where ridge sharpness was low. In addition, peaks

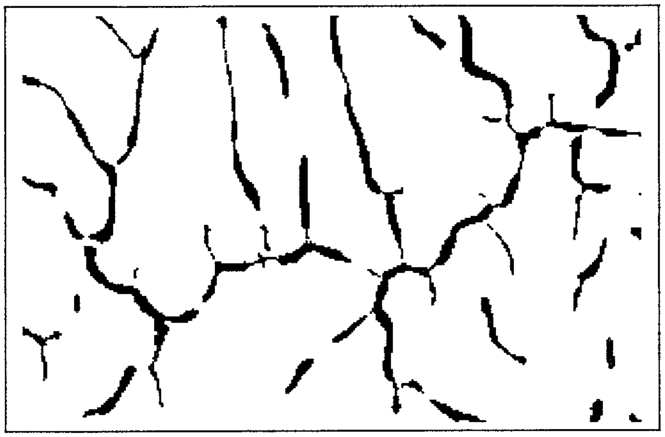

Figure 6: Unprocessed ridge features.

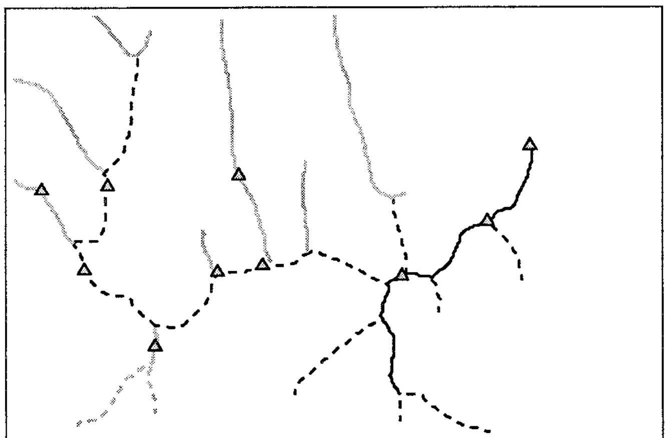

Figure 7: Peaks and ridge line hierarchy.

are found using a large area search that is more reliable than simple local maxima detection and ridge lines are organized into a hierarchy of importance that allows significant ridges to be used for initial matching while making available subsidiary ridges for subsequent verification operations. (The ridge lines to the northwest are not rendered in this view of the hierarchy, since they are actually part of the parent of the ridges shown.)

Features extracted using these processes still have a great deal of ambiguity associated with them. For example, lacking a priori information about viewing position and/or direction, it is hard to extract features such as peaks and ridges known to correspond in the view and map. This difficulty is similar to that faced by many symbolic problem solving systems dealing with tasks such as classification and diagnosis. In [Thompson *et al.*, 1993], we show that high-level hypothesize and test strategies can be integrated with lower-level feature extraction to solve difficult localization problems.

Additional details about feature extraction from views and maps can be found in [Savitt *et al.*, 1992, Savitt, 1992, Thompson *et al.*, 1993]. High-level strategies for feature matching are described in [Heinrichs *et al.*, 1989, Thompson *et al.*, 1990, Smith *et al.*, 1991, Heinrichs *et al.*, 1992] and computational implementations using these strategies can be found in [Bennett, 1992, Bennett, 1993, Thompson *et al.*, 1993, Thompson, 1993].

3 Landmark Selection.

We have previously demonstrated that the accuracy of landmark-based viewpoint determination is quite sensitive to geometric properties of the particular configuration of landmarks used [Sutherland, 1992]. Recently, the image understanding community has been paying increased attention to error estimation. Of equal importance are approaches which minimize the amount of error which can occur rather than only providing a posteriori characterizations of the error distribution.

The extraction of navigationally salient landmarks typically involves costs in time, computation, and sensing resources. As a result, there is benefit to be gained if simple strategies can be used to select a small set of landmarks which are likely to lead to accurate localization. Effective landmark selection methods are also relevant to mission planning, where one of the criteria entering into route selection should be the availability of landmarks sufficient to provide whatever degree of accuracy is required.

Error analysis is complicated by the lack of general sensor models which effectively describe position variability in properties used for viewpoint determination. This is particularly true when localization is based on bearings to features over a wide field of view, since sensing might involve mechanical scanning of cameras, fish eye optics, or more exotic technologies. We take a conservative approach in which we assume that the angular error in detected bearings to features is bounded, but the distribution of values within this range is not known. We then find the the region within which the viewpoint must lie to be consistent with these assumptions and are thus able to determine if conflicts with obstacles or untraversable terrain are possible. Figure 8 shows an example in which the relative bearing between two landmarks and the absolute bearing to a third landmark [Thompson *et al.*, 1993] separately generate possible viewpoint regions shown in light gray, the intersections of which are marked in dark gray.

Starting from the analysis of uncertainty regions, it is possible to develop simple heuristics for selecting landmarks likely to minimize the size of such regions [Sutherland and Thompson, 1993, Sutherland, 1993]. Note that

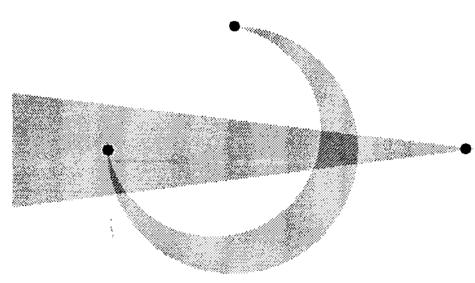

Figure 8: Intersection of viewpoint uncertainty regions.

this is not as easy as it might seem, since the problem must be solved with very minimal knowledge about the true viewing location. Figures 9 and 10 demonstrate the effectiveness of this method. Simulated navigators have identified landmarks on a map. Their task is to move along the segmented path shown by the dashed line. Current position is estimated at the beginning of each straight path segment, using relative bearing to three landmarks. In Figure 9, the landmark selection heuristic is used at each step to choose the three landmarks on which localization is based. In Figure 10, landmark selection is random. Both navigators start at the square at the left end of the dashed line. Direction and distance of move are based on estimated position. Uniform multiplicative error is assumed in both relative bearing measurement and in movement. The squares mark actual navigator positions at the end of each path segment for fifty trials. The scattering of location in Figure 10 is much increased over that in Figure 9.

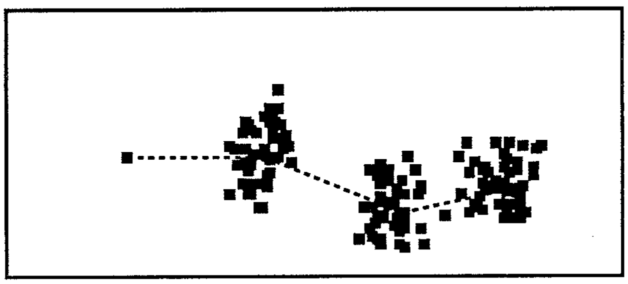

Figure 9: Fifty trials – "intelligent" landmark selection.

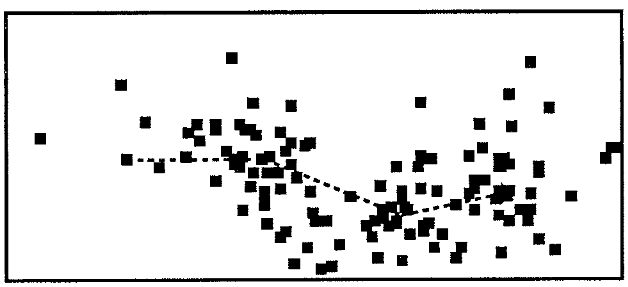

Figure 10: Fifty trials – random landmark selection.

4 Error Detection and Diagnosis.

Mobile robots capable of independent operation all employ some form of "perceptual servoing" to implement a sense-plan-act-verify cycle in which expectation about sensor data are compared with actual observations, and then differences are quantified and used to update estimates of current position and desired path (e.g., [Fennema *et al.*, 1990]). If a match between expectations and observations cannot be established, then some sort of replanning activity is initiated, a part of which requires a solution to the localization problem. This approach is only effective when a rich model of the environment is available, allowing for complete and specific predictions about the appearance of the world from any predicted viewpoint. Often, such models do not exist, particularly in tasks involving outdoor maneuvering. (Consider the effort that went into producing 5m resolution DEM data for the ALV site.) When this is the case, it is not possible to determine with certainty that an expectation does or does not match actual sensor values. At best some sort of confidence estimate can be produced. One consequence of this is that it is possible to travel substantial distances on what is in fact an incorrect path before determining with reasonable certainty that an error has occurred.

Sparse world models and the potential for substantial delays between when an error occurs and when it is detected mean that lower-level image understanding techniques are not sufficient in and of themselves to support effective plan monitoring in mobile robotics. We are addressing this problem by creating a qualitative model of error in vision-based navigation and using this model to characterize the sorts of errors that can occur, how they can be detected, and what sort of diagnosis is possible to determine the original source of difficulties [Stuck, 1992]. The research suggests a number of techniques that may usefully complement lower-level perceptual servoing.

5 Perceptual Issues.

Our approach to the development of novel methods for vision-based navigation is interdisciplinary, involving computational analysis, computer simulations, and studies of expert map users. Many of the strategies we use to automatically solve localization problems [Heinrichs *et al.*, 1992, Thompson *et al.*, 1993] arose out of experiments done with experts solving actual and artificial navigation problems [Pick *et al.*, in press]. In retrospect, these strategies make excellent computational sense since the experts are highly adapted to dealing with the ambiguity and complexity inherent in these problems. Nevertheless, the strategies were not obvious to us or others until we undertook our studies.

This interdisciplinary investigation is continuing with a current focus on the accuracy with which people are able to determine terrain geometry. By comparing human and machine vision perceptual competence, we can better understand the relevance of expert strategies to image understanding solutions. At the same time, we can identify specific perceptual skills for which mechanized aids and/or alternative training might significantly improve human performance. Elsewhere in this proceedings we summarize two such studies [Pick *et al.*, 1993]. One demonstrates that people are poor at estimating distance and slope in environments of the scale and topography typical of outdoor navigation tasks [Melendez *et al.*, in prep]. Since passive vision systems are also poor at these estimates, the strategies people use to compensate for their perceptual limitations may also be relevant in automated systems. The second study deals with localization using visual angle. Again, people are quite poor at using this cue. On the other hand, sensors which are capable of measuring large visual angles with reasonable accuracy can be designed, suggesting both possible differences between machine and human solutions and aids that might assist people in performing the task.

Figure 11: First panorama image.

6 Database and Software.

Many recent papers addressing ground level localization have presented results obtained only from synthetic imagery. A few have used the highly calibrated data available for the Martin Marietta ALV test area. In addition to the well-known pitfalls of failing to test new image understanding algorithms on real data, the use of synthetic terrain data to generate test imagery is problematic since realistic digital elevation data is often in error.

We have produced two 360° panorama images of mountainous terrain obtained with a video camera, digitized at high resolution, and digitally photo-mosaicked. (Figure 11 shows one of them). They extend approximately 6,000 pixels horizontally by 450 pixels vertically. Viewpoint location has been registered ±30m to USGS 30m DEM data. Direction relative to UTM north and tilt are known within ±0.5°. Geometric distortions due to misalignments between the pan axis, the camera, and "true" vertical have been normalized to approximately ±0.25°. Included in the database are 4 USGS 7.5' DEM quadrangles composed together and containing the viewpoints for the panorama images. Also available is software for converting USGS format data into a useful form, mosaicking DEM quads and panorama frames, and rendering expected views given map position.

References

[Andreas *et al.*, 1978] R.D. Andreas, L.D. Hostetler, and R.C. Beckmann. Continuous Kalman updating of an inertial navigation system using terrain measurements. In *Proc. IEEE National Aerospace Electronics Conference*, pages 1263–1270, 1978.

[Bennett, 1992] B.H. Bennett. *A Problem-Solving Approach to the Localization Problem.* PhD thesis, University of Minnesota, 1992.

[Bennett, 1993] B.H. Bennett. Knowledge-based control for robot self-localization. In *Proceedings NASA Goddard Conference on AI applications to Space*, May 1993.

[Fennema *et al.*, 1990] C. Fennema, A. Hanson, E. Riseman, J.R. Beveridge, and R. Kumar. Model-directed mobile robot navigation. *IEEE Trans. on Systems, Man and Cybernetics*, 20:1352–1369, November/December 1990.

[Haralick *et al.*, 1983] R.M. Haralick, L.T. Watson, and T.J. Laffey. The topographic primal sketch. *IEEE Journal of Robotics and Automation*, 2(1):50–72, 1983.

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

[Heinrichs *et al.*, 1992] M. R. Heinrichs, K. Smith, H. L. Pick, Jr., B. H. Bennett, and W.B. Thompson. Strategies for localization. In *Proc. DARPA Image Understanding Workshop*, January 1992.

[Melendez *et al.*, in prep] P.H. Melendez, D.A. Gentile, Jr. Pick, H.L., A. Yonas, and D.J. Wegesin. Psychophysical judgments of natural terrain. (in prep.).

[Pick *et al.*, 1993] H.L. Pick, Jr., A. Yonas, P. Melendez, D. Wagner, D. Gentile, and D. Wegesin. Perceptual aspects of navigation. In this proceedings, 1993.

[Pick *et al.*, in press] H.L. Pick, Jr., M.R. Heinrichs, D.R. Montello, K. Smith, C.N. Sullivan, and W.B. Thompson. Topographic map reading. In J. Flach, P.A. Hancock, J.K Caird, and K. Vicente, editors, *Ecology of Human-Machine Systems.* Lawrence Erlbaum Associates, (in press).

[Savitt *et al.*, 1992] S.L. Savitt, T.C. Henderson, and T.L. Colvin. Feature extraction for localization. In *Proc. DARPA Image Understanding Workshop*, 1992.

[Savitt, 1992] S.L. Savitt. *A Context Sensitive Segmenation Approach for Outdoor Terrain Feature Extraction.* PhD thesis, University of Minnesota, 1992.

[Smith *et al.*, 1991] K. Smith, M.R. Heinrichs, and H.L. Pick, Jr. Similarity judgment and expert localization. In *Proceedings Thirteenth Annual Conference of the Cognitive Science Society*, August 1991.

[Stuck, 1992] E.R. Stuck. *Detecting and Diagnosing Mistakes in Inexact Vision-based Navigation.* PhD thesis, University of Minnesota, 1992.

[Sutherland and Thompson, 1993] K.T. Sutherland and W.B. Thompson. Inexact navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, May 1993.

[Sutherland, 1992] K.T. Sutherland. Sensitivity of feature configuration in viewpoint determination. In *Proc. DARPA Image Understanding Workshop*, pages 315–319, January 1992.

[Sutherland, 1993] K.T. Sutherland. Landmark selection for accurate navigation. In this proceedings, 1993.

[Thompson *et al.*, 1990] W.B. Thompson, H.L. Pick, Jr., B.H. Bennett, M.R. Heinrichs, S.L. Savitt, and K. Smith. Map-based localization: The "drop-off" problem. In *Proc. DARPA Image Understanding Workshop*, pages 706–719, September 1990.

[Thompson *et al.*, 1993] W.B. Thompson, T.C. Henderson, T.L. Colvin, L.B. Dick, and C.M. Valiquette. Vision-based localization. In this proceedings, 1993.

[Thompson, 1993] W.B. Thompson. Geometric constraints for viewpoint determination. University of Utah Computer Vision Group working paper, 1993.