N93-25963

# KNOWLEDGE-BASED CONTROL FOR ROBOT SELF-LOCALIZATION

Bonnie Kathleen Holte Bennett, PhD

Honeywell Systems and Research Center
Minneapolis, MN

University of St. Thomas
Department of Computer Science
St. Paul, MN

## ABSTRACT

Autonomous robot systems are being proposed for a variety of missions including the Mars rover/sample return mission. Prior to any other mission objectives being met, an autonomous robot must be able to determine its own location. This will be especially challenging because location sensors like GPS, which are available on Earth, will not be useful, nor will INS sensors because their drift is too large. Another approach to self-localization is required.

In this paper, we describe a novel approach to localization by applying a problem-solving methodology. The term "problem-solving" implies a computational technique based on logical representational and control steps. In this research, these steps are derived from observing experts solving localization problems. The objective is not specifically to simulate human expertise but rather to apply its techniques where appropriate for computational systems. In doing this, we describe a model for solving the problem (Ref. 1) and a system built on that model, called localization control and logic expert (LOCALE), which is a demonstration of concept for the approach and the model. The results of this work represent the first successful solution to high-level control aspects of the localization problem.

Keywords: Knowledge-based control, robotics

## INTRODUCTION

Interest has been growing in the development of autonomous mobile robot systems. For example, autonomous mobile robots have been proposed for the Mars rover/sample return mission. In addition, applications for such systems are being proposed for military, industrial, and scientific endeavors. Missions include advanced reconnaissance, battle damage/contamination assessment, and exploration for cartographic, geographic, and geologic concerns. In each of these missions, an autonomous mobile robotic agent would be used in place of a human agent for cost savings and safety reasons. In order for a robotic agent to perform the above missions, it must be able to perform navigation tasks. These tasks generally include locating oneself on a map, determining a route to a specified location, performing some operation at that location, and continuing on to other locations or returning. The first of these tasks, locating oneself on a map, is the most critical because all the other functions rely on the agent having and maintaining accurate knowledge of self-location. The environments for these tasks are usually large outdoor spaces where environmental features are much larger than the robot, and the entire environment cannot be observed all at one time from the robot's sensors. Unambiguous, human-made landmarks and other location tools are not available.

There are several systems used by aircraft and other navigational systems to perform localization. They include global positioning systems (GPS) and inertial navigation systems (INS). GPSs use radio signal returns from orbiting satellites to determine an agent's current position on the Earth. The resolution of these systems is quite good and would preclude the need to solve the localization problem for Earth-based scenarios. However, localization is a major problem for space exploration. No GPS satellites exist for Mars. It will not be cost-efficient to put a GPS system in place for this relatively low usage, so in the near term, autonomous systems on Mars will need the capability to localize. While INSs also provide localization information, they unfortunately experience drift on the order of feet per hour over the long run and meters per second in the short run, making these systems inadequate for localization in ground-based robot systems.

## THE LOCALIZATION PROBLEM

### Problem Description

The objective of the localization problem is specifying the current viewpoint and viewing direction in the map coordinate system. Knowledge of self-location is essential to any agent that will interact with an external environment. If self-location is defined in terms of the map coordinate system, then knowledge of it makes all other map data accessible. Given the constraints of current technology (e.g., videocameras, digital maps), self-localization becomes a translation from one input domain into another. For our research, two data sources were explored: visual information and map information.

At an abstract level, localization can be modeled as three interacting processes (see Figure 1). Two of the processes are perceptual: they identify the pertinent information from the view of the image and from the map. The inputs from a videocamera are a series of pixels, each defining a grey level or color. These need to be preprocessed to determine meaningful symbolic labels like hill, valley, saddle, etc. The inputs from a digital map are elevation points in a grid pattern over the map area. These, too, need to be preprocessed into meaningful symbolic labels. Ideally, both of these processes are able to operate in both data-driven and hypothesis-driven modes. In the data-driven mode, they reason bottom-up from the input data, gleaning all they can from new data and integrating it with old data. In the hypothesis-driven mode, they reason top-down and search for specified data of a certain type or in a specific location. The third process determines the correspondence between the features in the map and the features in the view. Correspondence is determined by matching features from the map and the view. This matching should be able to occur in both directions: map to view and view to map. This capitalizes on the results of data-driven reasoning in each domain and uses those results to drive hypothesis-driven reasoning in the other. The search for matches should be guided by knowledge of the environment and heuristics that reduce the computational complexity of the search. The correspondence process mediates between the two perceptual processes. For example, it translates between the map's plan-view (down-looking) representation, where elements are north or west of each other, and the image's lateral (side-looking) view where elements are left and right or in front of each other.

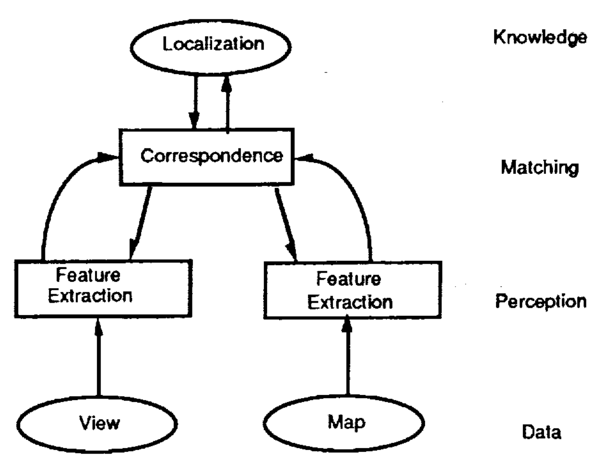


**Figure 1. Top-level Model of the Localization Process** (The perception process extracts features from the map and the view of the image. Matching determines the correspondence between the view and map features. Knowledge is used to determine the localization of the agent on the map.)

## Problem Approach

Formally, the localization problem is matching image features to map features and using that information to hypothesize a current viewpoint. The goal of localization is to determine an estimate of the location where the image was shot and the direction from which it was shot (i.e., to derive a viewpoint hypothesis). In the case where one unambiguous estimate cannot be derived, a list of prioritized viewpoint hypotheses is generated. These viewpoint hypotheses constitute the best estimates derived along with rank-order preference for them.

Because the objective of this research was to develop a model to provide high-level control for localization, it determined strategies for effectively and efficiently generating and evaluating viewpoint hypotheses.

The rationale for using feature-matching techniques is that there is simply too much data to deal with individually. This is essentially an argument of granularity. Both raw map and image data are digitized for input to a computational system; however, the granularity of this digitization is extremely small in order to provide the computer with as much data as possible. The prospect of matching each picture element, or pixel, in the visual sensor input data to a point on the map is daunting. The approach of combining individual map and sensor data elements into features reduces the search required for matching. In this approach, many data elements are combined into geographic features and are dealt with on the level of hills, valleys, gaps, and so forth. Humans performing this task use data elements on the level of geographic features. It is therefore a natural representation level to communicate the computer system's abilities to its human builders and observers.

## Demonstration Constraints

For this research, test cases with specific map and sensor data have been explored. In these test cases there are two available inputs: a topographic map and a single video sensor image. These inputs are assumed to be processed by a low-level processing system, which is not part of this research. Figure 2 shows an example view. Figure 3 shows the area of the topographic map used in this problem.

The rationale for limiting the inputs is that they are a minimal set of inputs. If a system can be built that works effectively with this constrained environment, it can likely be expanded to work in domains with richer inputs. The limit on the visual sensor to one input frame is quite severe. This means that no stereo or image-to-image information is available. The limits of a normal camera are also quite tight—the angle of view is limited. So, while a panoramic or preferably a full-circle view would give more data, we chose to explore what can be gained from the standard limited camera view. In addition to limiting the viewing angle from side to side, the standard camera also limits the viewing angle from top to bottom. So the data about the location on which the camera is standing, which could be quite useful, is unavailable. The main limitations on the map data are the resolution and the fact that it is limited to elevation data. Our goal was to focus on large outdoor environments, so we eliminated human-made features from our scenarios and picked areas where their effect was minimal. Thus, the elevation data in the digital map is essential and was readily available.

This work assumes that a low-level image and map processing system processes the raw image signal and map elevation data and sends processed information to LOCALE. The result of this processing is the location and classification of features in the map and image. Map features are peaks, valleys, ridges, etc. Image features are peaks, valleys, gaps, ridges, saddles, and inclines. Figure 4 shows the processed map information. The image and map processing system was simulated for this work because computational systems are only just being developed to this effect (Refs. 2 and 3). LOCALE can query the simulated image

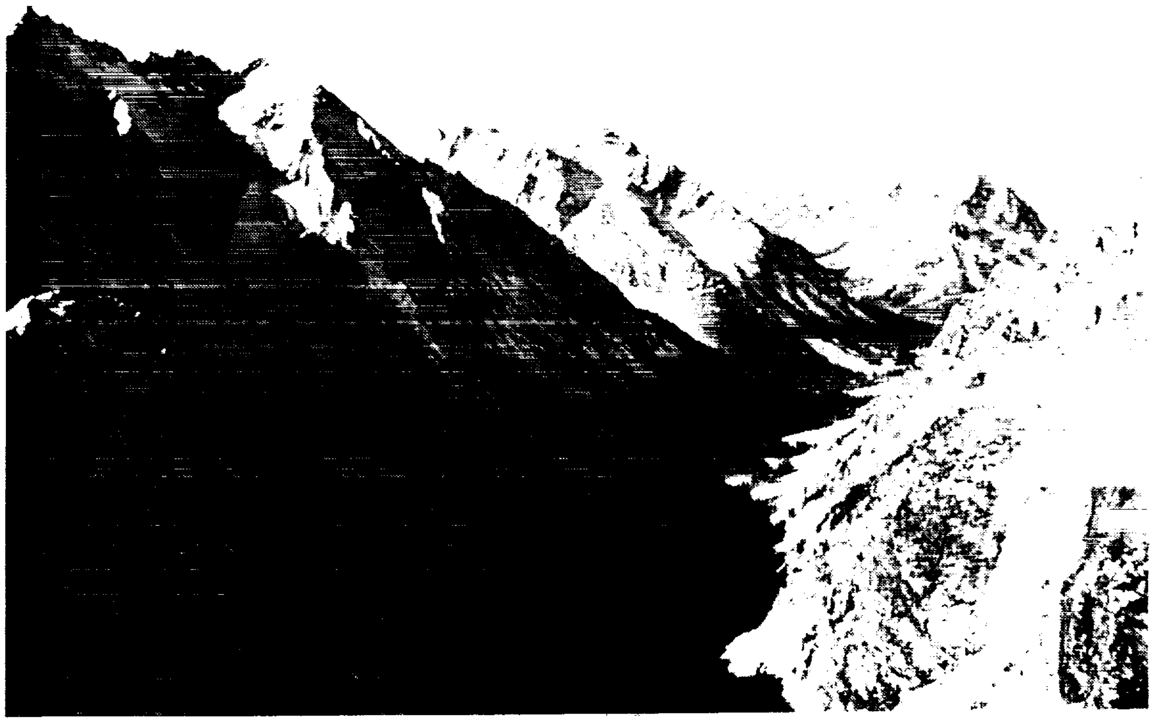

**Figure 2. Example Videocamera View** (In this example of a videocamera view, the most prominent features are the large valley in the middle and the two protrusions on either side of it in the front. Other valleys and peaks also appear in the view.)

and map processing system for specific data as required. The simulated image and map processing system replies by describing the map and image features (e.g., hills, valleys, etc.) at varying levels of detail.

Finally, the localization problem is actually a class of problems that fall on a spectrum determined by the amount of *a priori* information available to the system. Figure 5 shows the localization spectrum. Near one end of the spectrum are update problems where a lot of *a priori* information exists. In this region the typical problem is verifying one's location after a short move from a known location. Update problems are easier than dropoff problems because the agent has an indication of current location in an update problem. The agent needs to test actual sensor data against expected sensor data based on estimated location. In the dropoff scenario the agent must determine the estimated location in addition to testing its validity. In dropoff problems the agent has no *a priori* knowledge of where it is on the map. The research we have done addresses the dropoff problem and works with no *a priori* knowledge, not even a compass heading.

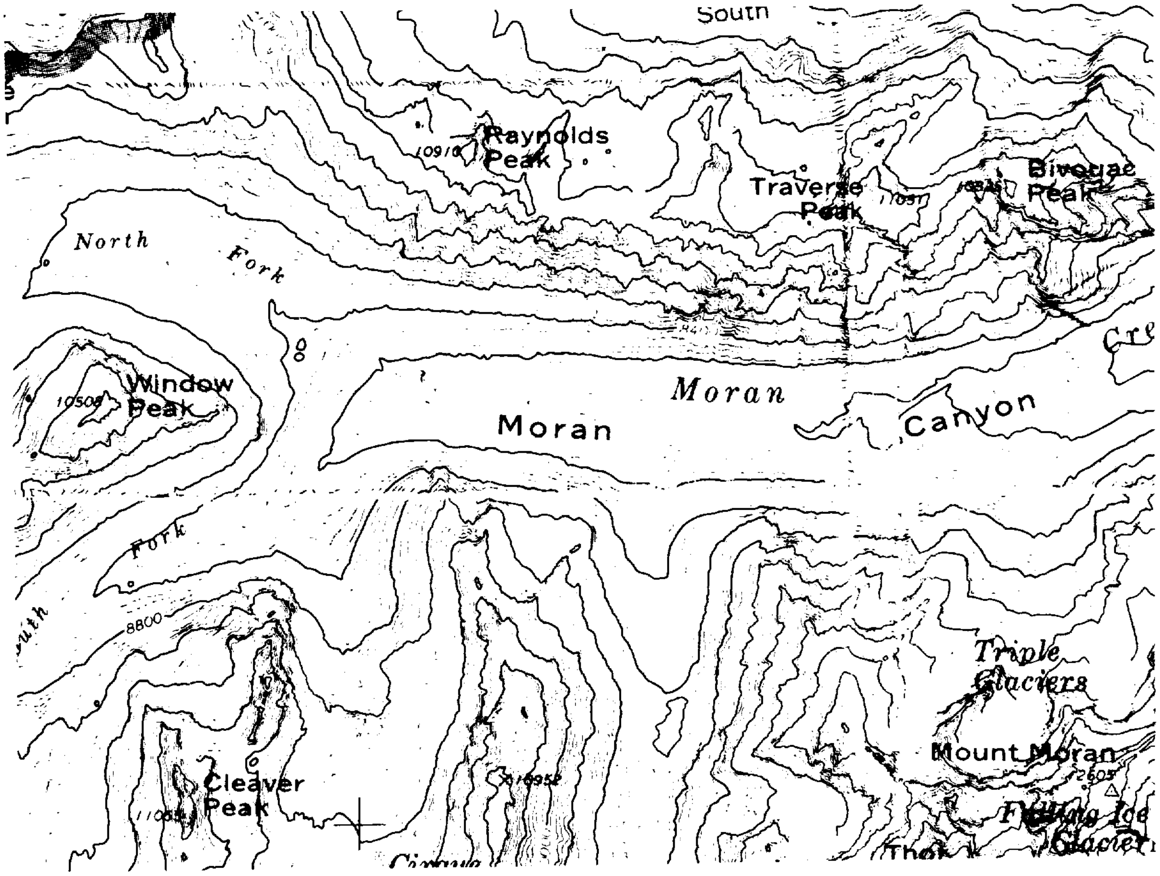


**Figure 3. Example Topographic Map of Teton Region**

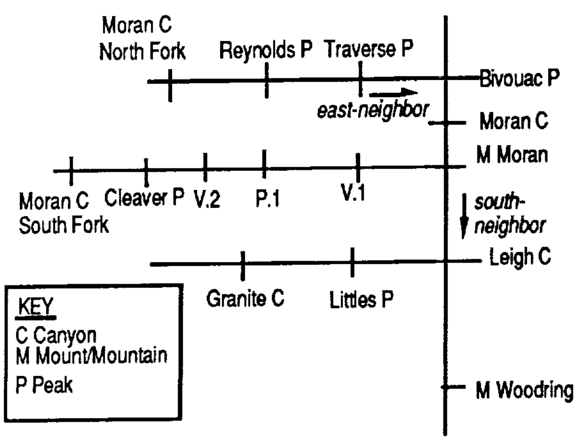


**Figure 4. Processed Information from the Map** (The processed map information is represented in a semantic network with proximity links between adjoining features.)

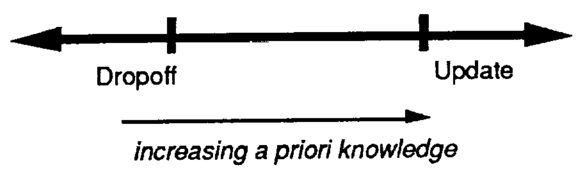


**Figure 5. The Localization Spectrum** (Problems with no *a priori* knowledge are dropoff problems. Problems with more *a priori* knowledge are update problems. This research focuses on dropoff problems.)

## RELATED WORK

Traditional computational approaches to the localization problem and related problems have developed in several areas: pattern recognition, control and representation systems, and computer vision research.

Classic pattern recognition approaches to the localization problem have differed from this work in two aspects: their reliance on low-level matching and their reliance on *a priori* knowledge.

Past work has explored low-level signal matching techniques as opposed to frame-based approaches for correlating images with maps. There are two signal domains in which this work can be pursued: the image domain and the map domain. More work has been done in the image domain. Ernst and Flinchbaugh (Ref. 4) matched estimated features with sensed features and required a known sensor location within a small neighborhood. Stein and Medioni (Ref. 5) explored localization using panoramic horizons as the features. This approach requires extensive pre-computation of indexed synthetic horizon maps and then matches the actual horizon to these. This approach also requires a full 360° view. As for the map domain, Lavin's work (Ref. 6) centered around determining what depth map could cause a two-dimensional (2-D) projection. It requires multiframe moving images.

The HILARE project (Ref. 7) sought to develop an experimental testbed on which to study general robotics, and robot perception and planning. The position referencing subsystem on HILARE used infrared triangulation operating in areas where fixed beacons were installed. This allowed for position determination either relative to objects and specific environment patterns or in a constructed frame of reference.

Beyond the low-level matching, some attention has been paid to control for low-level image processing. Arkin et al.(Ref. 8) explored an integrated system for the interpretation of visual data in a mobile robot testbed. This work essentially explored the low-level processing tasks and relied heavily on *a priori* knowledge of expected location. In related work, Fennema, et al. (Refs. 9, 10, and 11) use a hierarchy of representation and control techniques to solve the planning concerns for control uncertainty but do not examine it in light of specific localization problems. In addition, some research has explored advanced representational structures. Binford (Ref. 12) and Kriegman, et al. (Ref. 13) explore a hierarchical representation model for robot navigation focusing on interior environments. Smith and Strat (Ref. 14) begin to explore a frame hierarchy and a community of independent processes for solving outdoor problems with human-made landmark recognition. Andress and Kak (Ref. 15) explore knowledge-based control for accumulating evidence and controlling reasoning in a hierarchical spatial reasoning system with a computer program called production system environment for integrating knowledge with images (PSEIKI) that reasons about interior environments.

Traditionally, vision system approaches have only examined the update problem. Update implies *a priori* knowledge, an accurate estimate of current location. Examples of such systems include the work by Davis, et al. (Ref. 16) on DARPA's Autonomous Land Vehicle (ALV) program, Carnegie-Mellon University's Navlab project [17], and Lawton, Levitt, et al. (Refs. 18, 19, 20, 21, 22 and 23).

Thompson, et al. (Refs. 24 and 25) define the aspects of the localization problem and specifically the dropoff problem in large-scale environments.

The research described here uses a different approach where abstract representations of both the map and image were generated by extracting high-level features from each domain. The correspondence between these features is then computed in this higher-level abstract domain.

The work of Thompson, Pick, et al. (Ref. 25, 26 and 27) is closely related to this research. Here, protocol analyses of experts indicated that humans solving localization problems benefit from the following strategies:

1. Concentrate on the view first.
2. Landmark features should be organized into configurations.
3. Information about terrain at the viewpoint is important.
4. Multiple hypotheses need to be generated and examined.
5. Hypotheses should be compared using a disconfirmation strategy.
6. The ability to move to alternate viewpoints is important.

From work with experts, we made the following general observations:

- **Grouping things into configurations is important**—These configurations are linear and contain relationships among the constituent entities. This serves to constrain the search because the more complex a feature is the more specific the search can be. And, configurations are more complex than the features that compose them.
- **Working at various levels is important**—At times it is useful to take an overall view of the area or the map. At other times it is important to focus on increasingly minute details of an area. It is important to be able to swap back and forth between these levels, too.
- **Heuristic generation and testing of hypotheses is important**—For example, humans use the fact that a great deal of information is required to fully accept a hypothesis, while very little is required to reject one.
- **Data-driven and hypothesis-driven reasoning is used**—Early on, data about the viewpoint are gathered and interrogated—this is data-driven reasoning. Once enough data are present to construct sufficiently interesting hypotheses, they can drive the reasoning.

## THE MODEL

From the discussion on human experts in the previous section, two principles stand out:

- Grouping objects into composite entities focuses attention and reduces search.
- Representing data and working at multiple levels allows opportunistic and agenda-driven reasoning to work cooperatively.

### Grouping Objects

From a purely mathematical perspective, grouping objects into composites for matching has clear significance. If one is trying to match two sets of features (e.g., trying to match image features to map features) and there are five features in the first set and 40 in the second, then the number of possible matches is 90,536,361.

This calculation is

$$\sum_{j=0}^{\min(m,n)} \frac{n!}{j!\,(n-j)!} \frac{m!}{(m-j)!}$$

where m and n are the cardinality of the sets (in this case 5 and 40). If, however, the first set is actually grouped into two groups: one of three and one of two, and the second set is divided into eight groups of three and some singletons, then the number of possible matches between the groups of three in each set is only nine. The group of three from the first set could match any of the eight, or none at all. So, from a mathematical perspective, grouping clearly assists matching. In computational terms, grouping objects into composites and then working with the composites reduces the search space of the problem.

Grouping is observed in expert performance in the localization problem. Successful experts group individual features into configurations. The configurations observed and used are linear and generally radial from the subject. The expert realizes that there are fewer groupings of hill-valley-hill in a straight line on the map than there are individual hills or valleys. So the expert chooses to reason at the configuration rather than the feature level.

As for the model, the goal is to capture the groupings that facilitate the heuristic solution to the localization problem. Practically, this means an enumeration of the terms experts use in problems of this type and a thorough understanding of the interrelationships of these terms. This understanding leads to illumination of constraints and other rules of thumb to focus matching and other reasoning processes for localization.

## Multiple Levels of Representation and Reasoning

The second major principle of the model is that working at multiple levels provides the ability for opportunistic- and agenda-driven reasoning to work cooperatively. Data required for the model fall across a spectrum of levels of complexity. The levels of data required in the model reflect the derivative nature of the data. Low-level data are the raw inputs from the simulated image and map processing system. They consist of brief statements of fact, for example, that a certain hill is at a certain location. Higher-level data, including configurations, possible configuration matches, and viewpoint hypotheses derive from them.

Data at different levels are very different. Raw data are immutable facts. Derived data are less strong. It is useful to distinguish permanent and persistent data in this context. As the system approaches a given localization problem in a given geographic area, that is one problem-solving episode; there are some data that will be permanent to this problem-solving episode, and some that will not. The permanent data are facts like, "There is a hill at coordinate 335,432." Less permanent data (we use the term *persistent* data) may fall in and out of favor. Persistent data is a specialized example of a requirement for nonmonotonic reasoning. Hypotheses are examples of persistent data. At one point in the episode a hypothesis may look very promising, it may lose credibility, then gain it again as more data are gathered, but it is not truly temporary because even when it appears unlikely, the mere fact that a hypotheses has been explored to a certain degree of detail is important and should be preserved and not discarded as one would be tempted to do with false information. Like systems requiring full nonmonotonic reasoning, persistent data requires that the logical dependencies of conclusions are maintained; however, this is not a case where data will later be retracted, per se, as in a full non-monotonic system. In contrast, persistent data will not decrease the amount of knowledge held by a system (it will always grow), but this knowledge will simply have preference values that may change (increasing and decreasing) over time; however, all of the information used to solve a given problem is temporary in the sense that it holds for only one localization episode. In the next episode, when another given problem in another given geographic area is undertaken, all of these data will be gone, unlike the domain-specific information retained from problem to problem within a given geographic area.

In addition, we observe that two approaches to reasoning are employed by successful human experts. First, they use a *data-driven* approach to the problem, where they are gathering all the information they can bring to bear on the problem at hand. In this approach the expert is building up complex representations of the world. This is bottom-up reasoning from raw data. Once these representations have been built, and the pertinent data have been gleaned from them (e.g., there is a big valley in the middle of the image with a hill on either side, therefore, the configuration hill-valley-hill is important), then *hypothesis-driven* reasoning can begin (e.g., go look for hill-valley-hill configurations in the map). This is top-down reasoning from derived information. It is important to use both data- and hypothesis-driven approaches because a data-driven approach works well when little is known about the problem at hand, but a hypothesis-driven approach focuses the search when specific hypotheses exist. And, it is important to be able to alternate between them during the course of a problem-solving episode. A strategic reasoning superstructure provides the capability for the system to assess its current state, select among alternatives for the next step, and choose the appropriate one. This is the self-conscious control of the system because the break points provided in the strata of reasoning components are the opportunities for evaluation and selection of the next course of action.

## THE APPROACH

The approach used for this research was to understand the features in the domain relevant to solving localization and then to construct the representational and control structures to work with this information.

The individual features are hills, valleys, walls, etc. Image features have properties like membership in a group of similar features (valleys, hills, gaps) and relations to other features in the image (being right or left of one another, occlusion) and height in the frame. Map features have properties like location, slope, relation to other features (north-of, south-of, etc.), and elevation. The current implementation limits features to points on an X, Y coordinate. This limitation is used for simplicity of processing. The most significant of these properties are the relations among features. These relations are used to define configurations of features. One type of configuration is a linear configuration where three or more objects are in a line. In this case the relation between the first two objects is the same as between the second and third objects.

Hypotheses are expressions of potential solutions (or partial potential solutions) to the localization problem at hand. Multiple, conflicting hypotheses may be under consideration at any one time. There are three types of hypotheses: feature-match hypotheses, configuration-match hypotheses, and viewpoint hypotheses. Feature-match hypotheses acknowledge the possibility that a particular map feature may be a particular image features. These are constrained by matching rules derived from the possible visual appearance of map features. For example, a saddle from the map may appear as either a valley, a saddle, or a gap in the image. Only possible matches need to be posited. Configuration-match hypotheses are statements of the potential correspondence between a configuration in the map and a configuration in the image. These are constrained by the feature matches. For a configuration-match hypothesis to be retained, not only must the configuration forms match (two linear and three component configurations may be matched, but a linear configuration with three components and a right-angle configuration with four components may not be matched), but the individual features must be compatible. That is, the appropriate feature-match hypotheses must exist. Finally, viewpoint hypotheses are the outgrowth of configuration-match hypotheses. If two configurations do indeed match, then there is a limited area from which they can be viewed to give the appearance in the image. The viewpoint hypotheses are the representation of this. In addition to the individual components that must match for it to be true, the viewpoint hypothesis includes a description of the area where the observer must be located. This area is constrained to be within certain map coordinates limited by the visibility and intervisibility of the features in the image as related to their potential match partners from the map.

### Representation Issues

The representation components of the model use a hierarchical semantic network. Figure 6 shows the data categories of the representation components. The lowest level data is the raw data input from the simulated

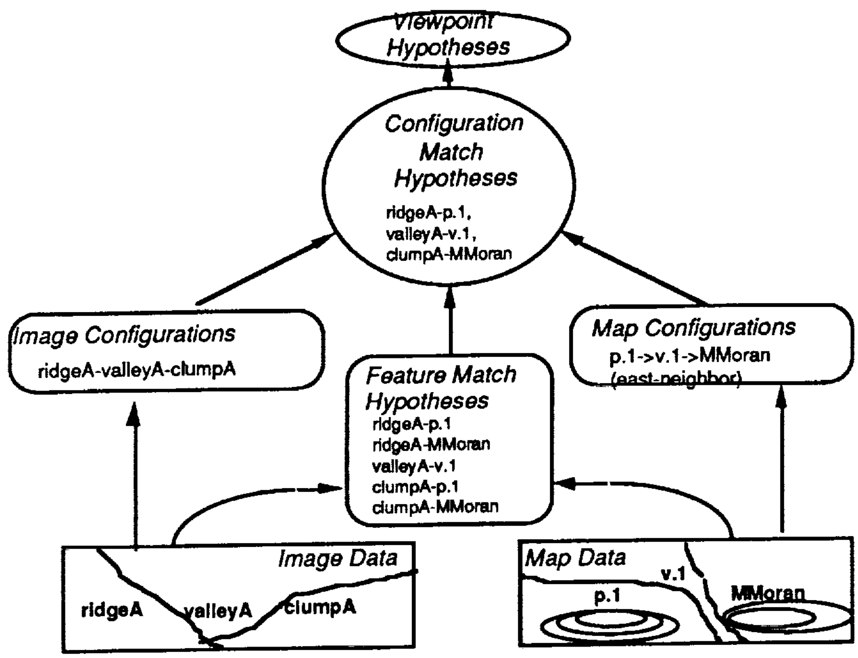


**Figure 6. Data Categories in the Computational Model of Localization**

image and map processing system. Successively higher levels of data represent abstracted, interpolated, or otherwise derived data that the system has concluded from the input data. The components of the semantic network are the objects and the relations between them. The components are represented in *frames* and the relations are represented in *slots* in the frames.

There are actually several hierarchies that are appropriate to this problem. The main data representation hierarchies are the configuration hierarchy and the feature taxonomy. Hierarchies are also used for rules and relations.

Individual map and image features are represented as instances of the classes defined in a domain-specific feature taxonomy that divides features into image features and map features. Image features are GAPS, IMAGE-RIDGES (so called to distinguish them from ridges that appear in the map), IMAGE-SADDLES, IMAGE-VALLEYS, INCLINES, and PEAKS. These are all of the elements that can be uniquely distinguished in an image. Map features are divided into BENCHES, DEPRESSIONS, PROTRUSIONS, and WALLS. DEPRESSIONS are divided into RE-ENTRANTS and VALLEYS. VALLEYS are divided into BASINS, DRAWS, GULLIES, HANGING-VALLEYS, and MAP-SADDLES. BASINS are divided into BOWLS and CIRQUES. MAP-SADDLES are divided into COLS and PASSES. PROTRUSIONS are divided into BUTTES, PEAK-PRIMITIVES, RIDGES, and SPIRES. RIDGES are divided into BUTTRESSES, SHOULDERS, and SPURS. WALLS can be distinguished into HEADWALLS.

## Control Issues

There are many types of expertise brought to bear on localization problems. High-level reasoning expertise can select from among several high-level alternatives:

- Understand the viewpoint,
- Understand the map,
- Generate and test hypotheses.

In addition, these high-level reasoning processes can call on a number of lower level subroutines to perform their functions:

- Gather map data,
- Gather image data,
- Scrutinize the incoming data and connect them to known data,
- Match features,
- Locate configuration,
- Match configurations,
- Establish viewpoint hypotheses,
- Evaluate and refine viewpoint hypotheses.

Each of these reasoning steps (both high-level and low-level) is a specialized subroutine. These subroutines can encapsulate just enough information to perform one specific function. The implementation represents them independently and weaves them together as appropriate (e.g., where a high-level function calls one or more low-level functions). And, it coordinates the actions of the multiple experts.

## THE SYSTEM

Figure 7 shows the system diagram of the computer implementation running on a Sun workstation using KEE® (by Intellicorp) and lisp. Data flow in from the simulated image and map processing system and are posted on either the map or the image knowledge bases (KBs). These KBs are built on top of the taxonomy KB, which contains the problem-specific data about the localization problem and the geographic region in general. The taxonomy is the hierarchy of geographic features that occur in this area. The control structures are the reasoners and rule bases that scrutinize the map and image information, taking into account their relationships within the taxonomy. The results of this scrutiny form the basis for the hypotheses that are posted in the hypothesis KB. Further scrutiny of the hypotheses may lead the control structures to send queries back to the simulated image and map processing system for more data.

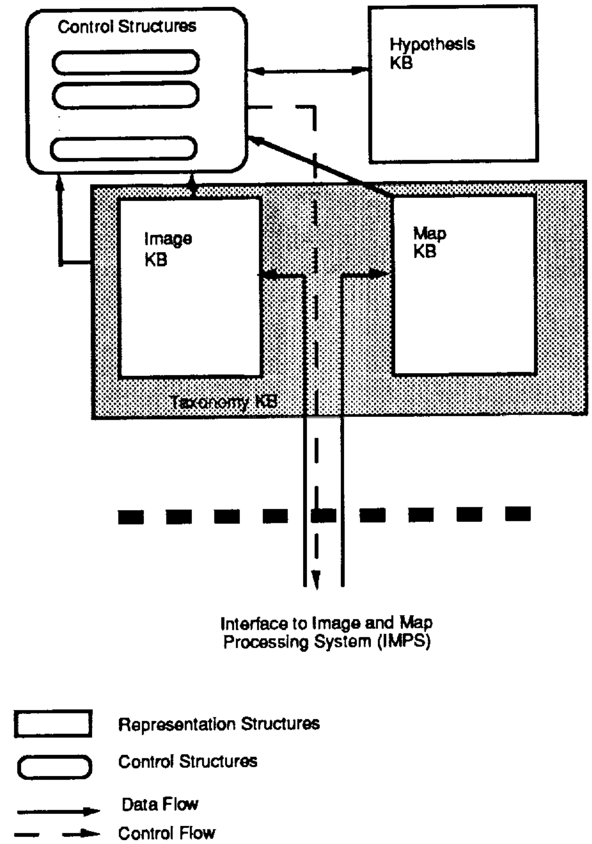


**Figure 7. LOCALE System Diagram**

These data will arrive as new postings to the image and map KBs.

The levels of representation of problem-specific information from lowest to highest are:

- Input data (map and image),
- Feature-match hypotheses,
- Configurations (map and image),
- Configuration-match hypotheses,
- Viewpoint hypotheses.

As an input datum arrives it is plugged in as an instance of one of the classes in the hierarchy. This allows it to inherit certain properties from its super classes and to be reasoned about as a member of the class.

The feature taxonomy provides the basis for feature matching. Table 1 shows a feature-

**Table 1. Feature-Match Matrix**(Potential features from the map and the image are compared for match quality.)

| *Map-Features* | Gaps | Image-Ridges | Image-Saddles | Image-Valleys | Inclines | Peaks |
|---|---|---|---|---|---|---|
| | | | *Image-Features* | | | |
| Benches | 0 | 3 | 0 | 0 | 1 | 1 |
| Depressions | 5 | 0 | 5 | 5 | 0 | 0 |
| Re-entrants | 5 | 0 | 1 | 3 | 0 | 0 |
| Valleys | 5 | 0 | 3 | 5 | 0 | 0 |
| Basins | 5 | 0 | 5 | 5 | 0 | 0 |
| Bowls | 3 | 0 | 3 | 5 | 0 | 0 |
| Cirques | 3 | 0 | 3 | 5 | 0 | 0 |
| Draws | 5 | 0 | 1 | 5 | 0 | 0 |
| Gullies | 5 | 0 | 1 | 5 | 0 | 0 |
| Hanging-valleys | 3 | 0 | 3 | 5 | 0 | 0 |
| Map-Saddles | 3 | 0 | 5 | 3 | 0 | 0 |
| Cols | 3 | 0 | 5 | 5 | 0 | 0 |
| Passes | 3 | 0 | 5 | 5 | 0 | 0 |
| Protrusions | 0 | 3 | 0 | 0 | 3 | 3 |
| Buttes | 0 | 3 | 0 | 0 | 3 | 3 |
| Peak-primitives | 0 | 3 | 0 | 0 | 3 | 5 |
| Ridges | 0 | 5 | 0 | 0 | 3 | 1 |
| Buttresses | 0 | 5 | 0 | 0 | 3 | 3 |
| Shoulders | 0 | 5 | 0 | 0 | 3 | 3 |
| Spurs | 0 | 5 | 0 | 0 | 3 | 3 |
| Spires | 0 | 3 | 0 | 0 | 3 | 5 |
| Walls | 0 | 3 | 0 | 0 | 5 | 0 |
| Headwalls | 0 | 3 | 0 | 0 | 3 | 0 |

match matrix between image and map features. Feature matches are ranked on a scale from 0 to 5, bad to good, where 0 indicates that a map feature can never appear as an image feature (for example, a gully in the map will never appear as a peak in the image), and 5 indicates a preferred match (for example, a peak in the image matches well with a peak in the map).

Reasoning is divided into task-specific subroutines and proceeds in the manner described in the approach section above. Components are high-level (strategic), and low-level (specific tasks). High-level components are the conscious reasoners of the system. They pick the strategic direction in which the system should proceed, initiate that work, evaluate its performance, and then choose the next strategic direction.

## RESULTS

In LOCALE two types of heuristics were used. The first type of heuristic was the use of configurations. By considering features in groups instead of as individuals, search was limited to those features that were parts of appropriate groups. The second type of heuristic was the use of category limitations. Only map features of the appropriate type were considered for matching with the image features. In addition, matches were prioritized based on proximity in the feature hierarchy, so that stronger matches could be considered first. Each heuristic is useful, but the real power of this approach came from the combination of both heuristics. The result was to constrain the search space to only those map features that were parts of appropriate configurations *and* were of the correct type to match with the image features. The effect of this is to determine the subset of features that meets the configuration constraints and to determine the subset of features that meets the category constraints, and then to take the intersection of those two subsets as the search space. We can quantify the benefits of this approach for an example problem. After three levels of map data detail and two levels of image data detail have been loaded into the system, there are thirty-seven map features and eight image features. The number of possible matches between these two sets is $6.48914 \times 10^{16}$. The power of this approach is that very few possible matches are actually considered and explored. Using the configuration heuristic, there are only 98 map configurations that match the current image configuration. Using the category heuristic, there are only 52 possible matches between the image features and map features that are constrained by the compatibility of their categories. Combining the results of those two heuristics, there are only twelve configuration-match hypotheses that can be developed into viewpoint hypotheses. This reduction of the search space is dramatic. Because this is a heuristic approach, its performance cannot be guaranteed in the same way an algorithm's performance can. The reduction in search depends on the uniqueness and identifiability of the feature categories and the availability of configurations; however, this magnitude of search reduction was consistently observed among all the test cases. Table 2 summarizes the state space reductions observed in both this and other test cases. The prospect of exploring and evaluating 10 to 20 test cases is reasonable. And, even if the correct solution is not always selected as the best alternative at any one time, the fact that it exists among the small, select set of alternatives is significant.

## CONCLUSIONS

This work has analyzed the components of the localization problem. The solution of this problem is a critical component to future work on autonomous mobile robot systems like those proposed for missions such as the Mars rover/sample collector. Localization has the potential to become a computationally insurmountable problem. However, heuristic strategies for high-level control can be employed to combat this challenge. Two such strategies are the use of configurations of features to control feature matching and the use of category limitations. The LOCALE system has been implemented to demonstrate these strategies.


## ACKNOWLEDGEMENT

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


**Table 2. Comparison of Test Cases**

| Test Case | Number of Map Features | Number of Image Features | State Space | Viewpoint Hypotheses Explored |
|---|---|---|---|---|
| Moran | 37 | 8 | $2.0 \times 10^{12}$ | 12 |
| Teewinot | 37 | 5 | $6.1 \times 10^{7}$ | 12 |
| Bivouac | 37 | 6 | $2.0 \times 10^{9}$ | 20 |

## REFERENCES


1. B.H. Bennett. (1992). *A problem-solving approach to the localization problem.* PhD Thesis, University of Minnesota, Department of Computer Science.

2. S.L. Savitt, T.C. Henderson, and T.L. Colvin. (1992). Feature extraction for localization. *Proceedings of the DARPA Image Understanding Workshop.*

3. S.L. Savitt. (1991). *A context sensitive segmentation approach for outdoor terrain feature extraction.* PhD Thesis, University of Minnesota.

4. M.D. Ernst and B.E. Flinchbaugh. (1989). Image/map correspondence using curve matching. *Proceedings of the AAAI Robotic Navigation Symposium*, 23-30.

5. F. Stein and G. Medioni. (1991). Map-based localization using the panoramic horizon. Proceedings of the *8th Israeli Symposium on Artificial Intelligence and Computer Vision.*

6. M.A. Lavin. (1979). Analysis of scenes from a moving viewpoint. P. H. Winston and R. H. Brown (Eds.) *Artificial Intelligence: An MIT Perspective.* Cambridge: MIT Press.

7. G. Giralt, R. Chatila, and M. Vaisset. (1984). An integrated navigation and motion control system for autonomous multisensory mobile robots. *Proceedings of the First International Symposium on Robotics Research.*

8. R.C. Arkin, E.M. Riseman, and A.R. Hanson. (1987). AuRA—An architecture for vision-based robot navigation. *Proceedings of the DARPA Image Understanding Workshop*, 417-430.

9. C.L. Fennema, Jr., E.M. Riseman, and A.R. Hanson. (1988). Planning with perceptual milestones to control uncertainty in robot navigation. *Proceedings of SPIE - International Society for Photographic and Industrial Engineering.*

10. C. Fennema, D. Hanson, E. Riseman, J.R. Beveridge, and R. Kumar. (1990). Model-directed mobile robot navigation. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(6), 1352-1369.

11. C. Fennema, and A.R. Hanson. (1990). Experiments in autonomous navigation. *Proceedings of the 10th International Conference on Pattern Recognition*, 24-31.

12. T.O. Binford. (1988). Image understanding: Intelligent systems. *Proceedings of the DARPA Image Understanding Workshop*, 17-28.

13. D.J. Kriegman, T.O. Binford, and T. Sumanaweera. (1998). Generic models for robot navigation. *Proceedings of the DARPA Image Understanding Workshop*, 453-460.

14. G.B. Smith, and T.M. Strat. (February 1987). Information management in a sensor-based autonomous system. *Proceedings of the DARPA Image Understanding Workshop*, 170-176.

15. K. M. Andress and A.C. Kak. (Summer 1988). Evidence accumulation and flow of control in a hierarchical spatial reasoning system. *AI Magazine*, 75-94.

16. L.S. Davis, D. Dementhon, R. Gajulapalli, T.R. Kushner, J. Le Moigne, and P. Vetach. (1987). Vision-based navigation: A status report. *DARPA Image Understanding Workshop*, 153-169.

17. C. Thorpe, M. Hebert, T. Kanade, S. Shafer, and the members of the Strategic Computing Vision Lab. (1987). Vision and navigation for the Carnegie-Mellon Navlab. *Annual*

*Review of Computing Science (2)*, J.F. Traub, B.J. Grosz, B.W. Lampson, and N.J. Nilsson (Eds.). Palo Alto: Annual Reviews, Inc.

18. D.T. Lawton, T.S. Levitt, and P. Gelband. (May 1989). Knowledge-based vision for terrestrial robots. *Proceedings of the DARPA Image Understanding Workshop*, 128-133.

19. D.T. Lawton, T.S. Levitt, C. McConnell, and J. Glicksman. (April 1986). Terrain models for an autonomous land vehicle. *Proceedings of the IEEE International Conference on Robotics and Automation*, 2043-2051.

20. D.T. Lawton, T.S. Levitt, C.C. McConnell, P.C.Nelson, and J. Glicksman. (1987). Environmental modeling and recognition for an autonomous land vehicle. *Proceedings of the DARPA Image Understanding Workshop*, 107-121.

21. T.S. Levitt, D.T. Lawton, D.M. Chelberg, and P.C. Nelson. (July 1987). Qualitative landmark-based path planning and following. *Proceedings of the National Conference on Artificial Intelligence AAAI-87*.

22. T. S. Levitt, D.T. Lawton, D.M. Chelberg, and P.C. Nelson. (1987). Qualitative navigation. *Proceedings of the DARPA Image Understanding Workshop*, 447-465.

23. T. S. Levitt, D.T. Lawton, D.M. Chelberg, K.V. Koitzsch, and J.W. Dye. (1988). Qualitative navigation II. *Proceedings of the DARPA Image Understanding Workshop*, 319-326.

24. W.B. Thompson. (September 1990). Vision-based navigation. *Proceedings of the DARPA Image Understanding Workshop*, 141-143.

25. W.B. Thompson, H.L. Pick, Jr., B.H. Bennett, M.R. Heinrichs, S.L. Savitt, and K. Smith. (September 1990). Map-based localization: The 'drop-off' problem. *Proceedings of the DARPA Image Understanding Workshop*, 706-719.

26. K. Smith, M.R. Heinrichs, and H.L. Pick, Jr. (August 1991). Similarity judgement and expert localization. *Proceedings of the Thirteenth Annual Conference of the Cognitive Society.*

27. M.R. Heinrichs, K. Smith, H.L. Pick, Jr., B.H. Bennett, and W.B. Thompson. (January 1992). Strategies for localization. *Proceedings of the DARPA Image Understanding Workshop*.