A. Innovative claims for the proposed research (1)

We near fruition of a decades-long effort to give robots dense spatial awareness to handle the widest possible range of situations encountered in long periods of autonomous navigation in unfamiliar surroundings. Lack of long-term reliable perception has kept autonomous mobile robots out of most utilitarian applications, domestic and military. Its arrival will enable a growing competitive market, supporting rapid further development.

Our techniques were at the boundaries of feasibility, but are rapidly becoming practical with rising computer power. Arbitrarily complex surroundings are modeled in high- but constant-cost 3D grid maps of spatial occupancy evidence. The grids allow quantities of data from very imperfect inputs like sonar and stereoscopic ranging to be statistically combined into high quality spatial representations. The process is greatly enhanced by a learning technique that automatically tunes the sensor models by which the raw data is interpreted. We devised 2D versions of the approach in the 1980s: they were used in many successful research mobile robots, but fell short of utilitarian reliability. We have been developing the 1,000 times richer and much more challenging 3D version since 1992. In the last three years, under prior DARPA support, 1,000 MIPS of computer power and many innovations allowed us to produce dense near-photorealistic 3D maps of large rooms from stereoscopic traverses.

Higher perceptual functions can be built on reliable probabilistic 3D grid representations. Path selection minimizing collision probability is straightforward. Localization techniques that interpolate matches of selected cells between local and global grid maps achieve 1 mm precision in maps of cell size greater than 1 cm. We plan to identify surfaces such as floors and walls using weighted least-square plane fits. We will try recognizing objects of known shape and size by convolving maps with 3D object templates, and variable ones by combining the results of several detectors with combinatorial or trainable statistical recognizers. The techniques will also be applied to filling in gaps and hidden areas in grids, a model-based inference. By the third year we will begin to extend the grid approach into the time dimension. Four-dimensional grids, about 100 times as expensive as 3D, will enable trajectories of moving objects to be represented and extracted with techniques similar to those used to recognize and fill in static features in 3D grid maps.

Other mobile robot research groups are addressing 3D and motion. Two-dimensional mapping has appeared in some commercial products. Quality techniques to date, however, depend on clean and precise data, typically from 2D scanning laser rangefinders from Sick AG, and have limited tolerance for clutter. The grid approach gets good results from a much wider range of sensors, including low-cost stereoscopic cameras, and can map arbitrarily complex scenes. Though computationally expensive we think it is superior for reliable operational mobile robots of the near future. In February 2003 we formed a company, SEEGRID Inc, to develop and license these methods for industrial transport, cleaning and security vehicles and future applications. The company's near-term focus will support, motivate and benefit from the long-term research proposed here. The results of both should, in time, nucleate a self-accelerating industry that rapidly evolves robot capabilities, perhaps analogous to stages in our own evolutionary lineage, a proven incremental path.

B. Technical Rationale, Approach and Plan(4)

The following images are an overview and interior view of a 512x512x128 grid map (1.6 cm cell size) deduced from 100 stereoscopic views collected at one meter height in a trip down the center of the L-shaped hallway. Given 1,000 MIPS of computing our existing program processes each view in about one second. The map is ragged at the distant outer edges, and in poorly seen areas, but quite good near the camera path. Further improvements are under development, but the representation is already good enough to support mobile robot navigation and other tasks much more reliably than prior methods, and sufficiently well for many practical applications.

Our work has an exceptionally long history, with the following milestones.

1975 First use of computer vision to guide an outdoor robot (tracking horizon features to maintain heading). First "Interest Operator" to select suitable image features. (Moravec, Stanford Cart, NASA support)

1977 First use of stereoscopic vision to map obstacle fields. First multi-ocular stereoscopic vision (9 viewpoints) to reduce errors. First multi-resolution stereo system. (Moravec, Stanford Cart, ARPA, NSF support)

1979 First demonstration of robot stereoscopic indoor and outdoor obstacle avoidance, navigation and 3D mapping (maps were a sparse scattering of several dozen points on objects in the scene). (Moravec thesis, Stanford, ARPA, NSF)

1984 First occupancy evidence grid maps, in 2D, giving greatly improved reliability for robot mapping (primarily using sonar sensors, but a demonstration using stereoscopic sensing). (Moravec and Elfes, Carnegie Mellon, ONR, Denning Mobile Robotics)

1989 First learning of sensor models for 2D grid mapping, greatly improving maps, especially in mirrorlike locations where most sonar measurements were misleading. (Moravec and Blackwell, Carnegie Mellon, ONR)

1992 First very fast implementation of 3D grid map sensor evidence projection, using a combination of new techniques (integer log-odds representation of evidence, cylindrical sweep of sensor evidence cross-section, pre-calculation of generic sensor cylinder map plane intersection addressing, sorting of intersection addresses by radius so only significant cone is processed). (Moravec, Carnegie Mellon, ONR, Thinking Machines Corp.)

1996 Center of radial distortion method (image dewarping) for rectifying camera images , especially from wide-angle lenses. First use of stereoscopic vision to build 3D evidence grids. (Moravec, Carnegie Mellon, Daimler Benz)

2000 First sensor model learning by color projection of multiple scene images into trial 3D grids (low color variance indicates high grid quality). Demonstrated with binocular stereoscopic sensor, producing near-photorealistic grid maps. (Moravec, Carnegie Mellon under DARPA support)

2001 Parallel-ray reformulation of fast 3D grid map sensor evidence projection program further doubles speed and improves edge clipping (code is also simplified). (Moravec and Crosby, Carnegie Mellon, DARPA)

2002 First combination of textured-light, trinocular stereoscopic vision with 3D grids, color projection learning, vernier-search stereoscopic matching to make navigation-ready maps of a test area. The near virtual-reality quality of the maps is probably sufficient for tasks beyond navigation, up to small-object recognition. (Moravec, Carnegie Mellon, DARPA)

The grid map shown above was produced in February 2002. Since then the following new techniques were devised and are in various stages of implementation and testing:

* Route selection in 3D grids, guided by path length, collision probability and area coverage criteria, accelerated by slice selection and resolution hierarchy.

* Localization by matching of 3D grids using sampling, resolution hierarchy and interpolation. Experiments with localization by FFT convolution.

* Supplementary local least-squares image dewarping correction beyond radial fit (allows use of inexpensive, imprecise, cameras and lenses).

* Probing developing grids to get statistical occupancy priors to use in Bayes formula conversion of stereoscopic image match curve into occupancy probability curve (other components of the formula are image noise and random-match histogram. One learned parameter defines the noise value, another the “gain” of the prior -- expected to greatly reduce remaining noise and enhance cohesiveness in reconstructed grids).

* Use of dual occupied and empty thresholds to evaluate grid quality in color-projection learning (should ensure grids properly distinguish empty from unknown space for path planning and object recognition, not just occupied volume for visualization).

* Color projection and grid visualization by ray propagation through grid cells, accelerated by multi-resolution grid representation (much better scaling properties with grid size than the conventional surface-based graphics algorithms we have been using).

In the coming years we will evaluate and improve or replace these components, and extend the techniques. Refinement and ruggedizing of the existing techniques will be done largely under the wing of a new company SEEGRID, founded in February 2003 by Hans Moravec and Scott Friedman (M.D., founder and CEO of medical software company CareFlow|Net, which is being acquired by another company) to commercialize 3D evidence grid based perception for mobile robots. The near term, applications-oriented SEEGRID work should provide a solid foundation for the long term research proposed here, and a future path for bringing it into wide use.

The grid representation, using well-trained sensor models, accumulates evidence from imperfect sensors in a carefully weighted way to deduce reliable probabilistic spatial descriptions. We propose to use those dense, high-resolution, volumetric maps to achieve higher-order perception, planning and execution. A list of tasks and techniques we will address follows. It has been our experience and is our expectation that new and better ideas will emerge as we engage intensely in the development and our understanding deepens.

* Object recognition: Geometrically rigid grids offer a straightforward 3D extension of 2D pattern recognition methods. For fixed shapes, e.g. door frames and desks, simple templates may suffice, giving correlation spikes when superimposed on corresponding shapes in grid maps, despite noise and missing data. Plane templates may identify surfaces such as floors and walls. Variable objects may be found as a combination of several shapes with freedom of relative pose, grouped by combinatorial or trainable statistical recognizers.

* Unknown volume inference: Techniques similar to those developed for object recognition will be applied to filling in gaps and hidden areas in grids, a model-based inference. For instance, a wall-recognizer might be used to fill in gaps of poorly-seen unknown space in otherwise detected flat surfaces.

* 4D Grids to map motion: As computer power and memory permit, we will begin to extend the grid approach into the time dimension. Four-dimensional grids, about 100 times as expensive as 3D, will enable trajectories of moving objects to be represented and extracted with techniques similar to those used to recognize and fill in static features in 3D grid maps.

* Grid-based manipulation planning: Reliable free-range navigation makes possible many robot applications, but others, like retrieval and advanced cleaning, require that objects or tools be manipulated in 3D. Grid representation provides the same advantages for planning manipulator motion as for vehicle motion, reliably characterizing clear, occupied and unknown space despite imperfect sensors. We may first approach the problem by adapting existing configuration space planning algorithms, which are good fit for static planning.

* Grid-based dynamic vehicle and effector control: 4D grids should be usable to plan and monitor dynamic vehicle and manipulator trajectories in the presence of moving objects. This task will tax computer power storage in 64-bit processors available even five years from now, and will require clever sampling and variable resolution or other techniques.

Comparison with Current Technology

Our techniques are approaching commercialization, and it is appropriate to compare them to others at a similar development stage.

Three firms offer real-time stereoscopic systems, Tyzx a Stanford spinoff, Point Grey Research once affiliated with U. British Columbia, and Videre Design, using techniques from SRI International. All provide frame-rate depth images with existing computers (some augmented with affordable special hardware). In indoor room scenes, however all suffer from large unranged gaps, picket-fence and other range errors, and would be unsuitable by themselves for navigation. The grid approach, though many times slower, allows ambiguities in individual range measurements to be represented as a tuned probability profile that adds evidence, whether weak or strong, to an accumulating map. Multiple views gradually fill in gaps and reduce ambiguities until a navigation-ready map is achieved.

The other major navigation approach, used by many research groups and commercially offered by Siemens for service robots, relies on very clean data from a scanning range sensor from Sick, A.G., to build 2D point cloud, line or edge maps, or, by sweeping the scanning plane, to slowly build up 3D faceted maps. The cost of the Sick scanner is about $5,000, similar to the present computing cost for our 3D grid approach, and it is somewhat bulky. Computing cost is declining faster than mechano-electro-optics, and we expect our approach with stereoscopic input to become more economical than a laser range scanner within two years. In addition the grid representation is better for complex scenes with many overlapping surfaces, and can handle errors and ambiguities well. But should inexpensive high-resolution direct range sensors become available in future, whatever their limitations, they can be modeled and used effectively as input to grid maps.