Saburo Tsuji and Jiang Yu Zheng
Department of Control Engineering
Osaka Unlversity
Toyonaka, Osaka 560, Japan

ABSTRACT

Stereo-vision-based mobile robots observe their environments and acquire 3-0 data with considerable errors 1n range. Rather than to use the conventional 2-0 maps, described with these Inaccurate data 1n the absolute coordinate system, a flexible relational model of the global world can be built with local maps suited for each function of robots such as path planning or manipulation. For the path planning, a perspective representation of the local world, the image on which the scene Interpretation Is mapped, 1s proposed. Using world constraints and sensory Information of camera orientation, a stereo-image analyzer determines vertical projections of edge points on the floor in the image. A method for planning of promising paths to the specified goal using this representation is presented.

1. INTRODUCTION

Sensor-based navigation of mobile robots 1n unstructured environments has been extensively studied as one of key problems of artificial intelligence research. Host robots plan paths to the destinations using 2-0 maps, given a priori or built by themselves, represented 1n the absolute coordinate system [Horavec 83, GIralt et al. 84, Tsujl 85, Elfes 86]. Drawbacks of such maps were pointed out by Brooks [Brooks 85], who presented an interesting idea to use a rubbery and stretchy relational map, making the spatial reasoning robust to errors 1n estimating positions of objects and robots 1n the absolute coordinate system. Realization of the idea, however, has not been reported yet.

The authors' group has been also studying a vision-based robot wandering about and accomplishing the specified task In an indoor environment. We follow the above-mentioned idea 1n the sense that use the local coordinate systems with relative transforms and error estimates. Rather than to use the conventional 2-0 maps for spatial reasoning, a flexible relational model of the global world can be built with local maps suited for each function of robots, such as path planning or manipulation. For the path planning, this paper presents a new scheme, based on the properties of

vision, for representing the local world and using 1t for planning. Investigating the merits and demerits of transforming the visual data into the conventional 2-0 map* we use a perspective representation of world, the image on which the scene interpretation is mapped, for planning promising paths to the goal (or sub-goals). Using world constraints and sensory information of camera orientation, a stereo-image analyzer determines vertical projections of edge points onto the floor in the image. A method for planning of promising paths to the specified goal using this representation is presented.

2. WORLD MODEL

World Model with Local Maps

Vision of intelligent robots play the most important role among their sensors in building and maintaining models of their environments. For the effective usage of the models, their structures should be carefully designed so as to be suited for the functions of the robots. A mobile robot wandering around in an unstructured environment needs functions as follows: (1) It plans a route to the specified destination, (2) navigates safely along the planned route and (3) performs the given job such as manipulation or monitoring at the destination. We do not need a world model of a unified structure and thus same performance through the entire task, but models with different structures and different grain sizes, which represent facts about the world needed for each function, are more reasonable.

For example, the planning needs coarse representation of free spaces relative to some landmarks and the robot* but the detailed structures and shapes of objects in scene, 1n most cases, do not provide important information. Navigation uses the local world model describing the approximate direction and distance it will move, obstacles along the route and the landmarks by which the sub-goal 1s specified. Manipulation needs, on the contrary, precise 3-0 models of objects located 1n a limited area such as a workbench. We can, therefore, build a global world model covering a wide environment by combining the different representations of the local worlds with coarse geometrical relations between them.

2-D Euclidean Map or Perspective Model

Let us consider how the local world is represented for planning paths to the destination. Precise estimate of the free space geometry is, in

general, difficult because of the limited resolution of visual sensors and, if possible, they are often not so meaningful at the planning stage. Robots can not always find the route to the destinations from single view. In many cases, when they have not enough experiences of wandering about the routes, these robots plan promising routes and next observation points where the possibilities of further navigation are examined. Representation of the local worlds should be selected so as to be convenient for such planning.

Vision-based robots analyze input images taken by TV cameras and yield 2-D maps in the absolute coordinated system by transforming positions of obstacles found in the images into an Euclidean representation. The maps, on which the robots are located, are then labelled with free, occupied, or unknown regions, and the routes to the destination are planned on it. The 2-D maps have been used because (1) we can easily understand them and, therefore, (2) integration of 3-D data acquired from images taken at different positions seems straight forward (but not so easy because of uncertainties in camera locations and orientations, and much research has been done [Faugeras et al. 86, Lin et al. 86]).

Difficulties arise in such representation from inherent error due to measuring in image. Measurements of ranges to objects suffer from the minimal spatial resolution of vision sensors used. Thus, the error in depth measurement increases with proportional to the squared distance of the object from the camera. As a result, estimates of far object points, back projected from the images to the 2-D map, are very uncertain. Such uncertainty must be added as a feature of each point in the maps. Another difficulty is that the observed points are sparse in far areas in the 2-D maps; the density is inversely proportional to the squared distance from the camera. If a cellular representation of the 2-D map [Elfes 86] is used, we need to interpolate features between the observed points to fill in blank cells.
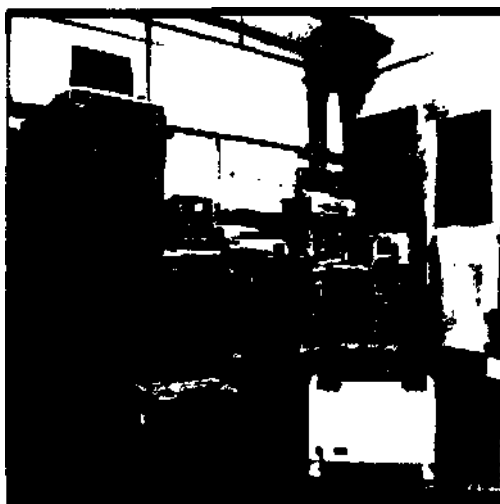
We propose to map the scene interpretation onto a

perspective world, the image itself, rather than transform the image data into the 2-D maps. We assume our robot moves on an almost flat floor, and the projection of the infinite plane containing the floor onto the image is used as a representation of the local world. This representation is natural in the sense that (1) its spatial resolution is same as that available from the sensors and (2) any interpolation and transformation is not needed while we deal with single image. We argue human beings without maps also plan the route in the field of vision rather than make reasoning in a top-view of the world.

Integration of information available from the local perspective worlds at different points is possible by estimating transformation matrices between them. A different approach, however, is proposed. We use a simple principle that an observation from a closer point yields more reliable results. Promising paths to the next observation point, where more reliable estimates will be obtained, are generated in each perspective world.

1 IMAGE ANALYSIS AND PATH PLANNING

Hardware System
We have been doing experiments with a HERO ROBOT, on which two small CCD cameras (NEC TI-22AI) separated from each other by 0.2m are mounted at a height of 0.7m. Pitch and roll angles of a camera are measured by orientation sensors (Watson flux-gate magnetometer). Images are converted into 256 by 256 8 bit digital pictures (the appoximate equivalent focal length is 300 pixel) and their edge points are found by a T0SPIX-2 image processor connected to a SUN3 workstation, which analyzes the stereo imagery and plans the promising routes. The results are fed back to the robot through a control computer MC6808. The number of rotations and steering angle of the front wheel are sent to the control computer, from which a coarse estimate of its motion is available.

Stereo Matching and Scene Interpretation
An image analyzer in the SUN3, which receives

Fig.1 An example of input images.

Figs.2 Line segments of stereo images and their correspondences.

stereo images and their edge pictures obtained by a 3x3 Sobel operator, finds 3-D structure of the viewed world and, then, determines the floor boundaries. Most indoor scenes are rich in straight edges which give important cues for scene interpretation. For segment-based stereo matching, the analyzer fits lines to edge segments of considerable lengths found by tracking edge points. We use a competitive matrix between line segments of the stereo pair, and find correspondence of lines in the images, by examining the similarity of edge strengths, locations and directions of segments and the consistency in matching. Fig.I shows an example of input left images and Figs.2 display the results of matching of the stereo pair (bold and thin lines represent matched and non-matched segments, respectively). Thus we can determine the 3-D structure of most portions of the scene.

In order to find free spaces for the navigation, we need information of height of edge points. We assume the optical axes of two cameras are parallel and horizontal. Let the coordinate system X,Y,2 be fixed at the center of the two lens centers of cameras with Z axis parallel to the optical axes. Suppose a point P at (X,Y,Z) in space is projected on an image plane of camera at (x,y), assuming the central projection. Using the camera geometry, we have the following simple equations for y#O [Tsuji et al. 86].

$$d = fL/Z = Ly/(H-H_o) \qquad (1)$$

where d is the disparity between corresponding image points, f is the focal length of the camera, L is the distance between two camera centers, and H and Ho are heights of P and the lens center from the floor, respectively. Thus, the disparity of a point in space is determined from the camera parameters (L and Ho) and its heights in image and in space.

In order to test whether or not the segments are on the floor, we examine each point in the segments that the corresponding segments contain an image point with a disparity of -Ly/Ho. If most part (more than 80%) of a non-horizontal line segments contains such image points, the line segments are considered as on the floor. For a pair of horizontal line segments, we determine whether or not the lines are on the floor by examining the disparities of their end points. Fig.3 shows the line segments on which edge points are labelled as floor edge points. Note that
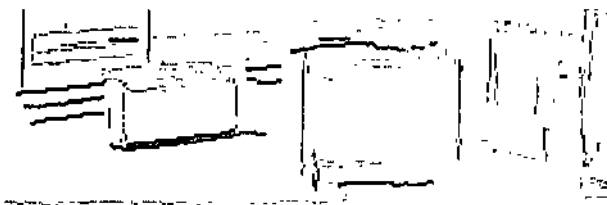
boundaries of shadows casted on the floor are labelled as the floor edge points, while the bottom edge of an object appearing in the right lower part in Fig.I is interpreted as floating above the floor (the object is supported by wheels). Thus all edge points in the image are labelled as the floor or non-floor.

Most path planning of mobile robot uses 2-D maps without height information. As a result, the planned routes may be very dangerous for collision with objects overhanged to the path. Therefore, we need 3-D information to find free paths in the world. Since the shape of our robot is simple (approximated by a cylinder) and it cannot change its shape to avoid collision, we use rather simple 3-D information that each point on the floor has an overhanged object or not. We, therefore, project dangerous non-floor edge points onto the floor in the image. We have the following simple relation between the image point (x,y) and its vertical projection (x,y') onto the floor.

$$y' = -(dH_o/L) \qquad (2)$$

Thus, we can easily project these edge points onto the floor in the image, and label the projection as obstacle edge points. Thus, the edge points have three labels; floor, obstacle and non-floor.

The above discussion assumes the invariance in height and tilt of the cameras through execution of the task. The floor on which our robot moves is almost flat and, as a result, we can consider the height is invariant. Small uneveness of the floor, however, causes the cameras to shake, and results in a considerable amount of movements of images. In order to solve the problem, we use the information from the orientation sensors of the camera, and can compensate for the movements less than one pixel.

Planning Paths
As mentioned before, the system searches the local map for promising paths to the goal or the next

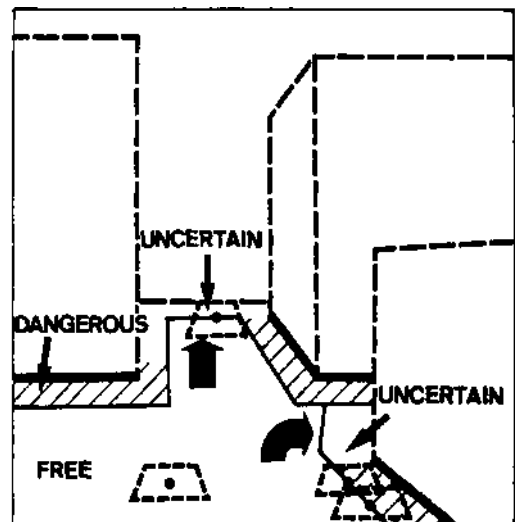FIg.3 Line segments labelled as lying on the floor.

Fig.4 Finding of free regions in the perspective world.

observation point. At first, we label the perspective world as free, dangerous, uncertain, and invisible. We move the projection of robot onto the floor in the perspective world (therefore the shape and size of the projected area change) and examine the possibility of collision. For simplicity of computation, we approximate the robot's projection by a square.

We start with the bottom row of the image assuming that it belongs to the free area, and move the robot's projection until its any part collides with edge points (see Fig.4). Then, we change the direction of movement, and continue to move it until it cannot move further. By iterating the procedure, we can find all possible movement of the projection, and areas covered by the center of the projection are labelled as free. Fig.5 shows the free area found in Fig.l. The proposed procedure is robust to the errors in finding edge segments, because missing of small portions of edge segments results in very little effects to the labelling.

We also label the narrow regions between the *free* areas and the obstacle edges as dangerous regions, and those between the free areas and non-floor edge as uncertain regions. The rest parts in the perspective world are considered as invisible regions.

Now let us consider how we can find possible routes to the goal if its approximate location is given in the perspective world. Route finding seems easy if routes to the goal exist in the free region. Objects in scene, however, often occlude parts of the routes or the goal, and we need to search the perspective world for promising routes to the goal. If the labels are correct, the routes from the free region must exist across the uncertain regions toward the goal as pointed in Fig.4, and we can determine the most promising route by using a heuristic function to estimate the possibility and cost of attaining to the goal, and set observation points along the route where the uncertain regions are easily observable.

We should note that the labels are not reliable if areas are far from the cameras, because of the limitation of sensor resolution. For example, free paths sometimes exist in far areas labelled as dangerous because of the uncertainty in estimate of 3-D distance between edge points. The uncertainty in hight estimate results in

labelling of edges of low obstacles as floor edges. Therefore, a good policy for path planning is to plan promising or possible paths to the destination, and moves along a route to obtain a closer view of the path. We propose that promising paths are generated in each perspective world and portions with more reliable estimates are merged together with a graph structure of which arcs give the approximate geometrical relations between them.

4. CONCLUSION

In this paper, we examine problems of world modelling and path planning which must be solved by a mobile robot exploring an unknown environment. In particular, we introduce the perspective model of the local world for planning promising paths to the destination. Image analysis and spatial reasoning based on this representation are also described. The major parts of the system have been already implemented and tested, and the whole system will be built in near future.

REFERENCES

[1] Brooks, R. A., Visual map making for a mobile robot, Proc IEEE Conf. Robotics & Automation, pp.824-829 (1985).

[2] Elfes, A.: A sonar-based mapping and navigation system, Proc. IEEE Conf. Robotics & Automation, pp.1151-1156 (1986).

[3] Faugeras, 0. 0., Ayache, N. and Faverjon, B,, Building visual maps by combining noisy stereo measurements, Proc IEEE Conf. Robotics & Automation, pp. 1433-1438 (1986).

[4] Giralt, G„ Chatila, R. and Vaisset, M., An integrated navigation and motion control system for autonomous multi-sensory mobile robots, in Brady, M. and Paul, R. (eds.) Robotics Research The First Symposium, MIT Press, pp. 191-214 (1984).

[5] Lin, Z. C, Lee, H. and Huang, T. S., Finding 3-D point correspondences in motion estimation, Proc. 8th ICPR, pp.303-305 (1986).

[6] Moravec, H. P., The Stanford Cart and CMU Rober, Proc. IEEE, Vol.71, No.7, pp.872-884 (1983).

[7] Tsuji, S„ Monitoring of a building environment by a mobile robot, in Hanafusa, H. and Inoue, H. (eds.) Robotics Research The Second Symposium, MIT Press, pp.349-356 (1985).

[8] Tsuji, S., Zheng, J. Y. and Asada, M„ Stereo vision of a mobile robot: World constraints for image matching and interpretation, Proc. IEEE Conf. Robotics & Automation, pp. 1594-1599 (1986).

Fig.5 *Free* regions In Fig.l.