Construction of Semantic Maps for Personal Mobility Robots in Dynamic Outdoor Environments

Naotaka Hatao, Satoshi Kagami, Ryo Hanai,
Kimitoshi Yamazaki and Masayuki Inaba

Abstract In this paper, a construction system of outdoor semantic maps by personal mobility robots which move in dynamic outdoor environments is proposed. The maps have topological forms based on understanding of road structures. That is, nodes of maps are intersections, and arcs are roads between each pair of these intersections. Topological framework significantly reduces computer resources, and enables consistent map building in environments which include loops. Besides, Trajectories of moving objects and landmarks including entrances of buildings and traffic signs are added in the each road. This framework enables personal mobility robots to recognize dangerous points or regions. The proposed system uses two laser range finders (LRFs) and one omni-directional camera. One LRF is swung by a tilt unit, and reconstruct 3D shapes of obstacles and the ground. The other LRF is fixed on the body of the robot, and used for moving objects detection and tracking. The camera is used for localization and loop closings. We implemented the proposed system in a personal mobility robot, and demonstrated its effectiveness in outdoor environments.

1 Introduction

Personal mobility robots are required to have ability to move safely in outdoor environments. Outdoor environments are much larger than ranges of sensors and include moving objects faster than robots. Although mapping, localization and path planning methods are radically improved in recent years, safe navigation in dynamic outdoor environments will not be achieved by just improving their accuracy or reducing their computational costs. To cope with the dangerousness in dynamic environments, we introduce semantics and reasoning on maps.

Naotaka Hatao
National Inst. of AIST, 2-3-26 Aomi, Koto-ku, Tokyo, Japan e-mail: n.hatao@aist.go.jp

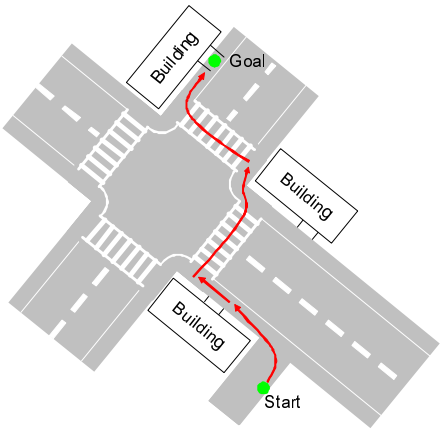

Fig. 1 A simple example of outdoor environments and the human path

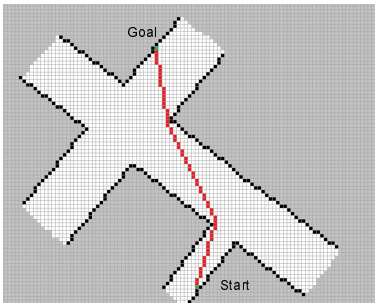

Fig. 2 A path generated from a grid map: it has higher risk of colliding cars

Several previous researches took notice of semantic maps, and proposed their applications. Nüchter et al. developed semantic labeling and object detection framework using a 3D map obtained by a robots[1]. Posnera et al. proposed an online system for semantic labeling of maps in outdoor urban environments[2]. Mozos et al. extracted semantic areas like rooms, corridors, and doors from a 2D map obtained by a robot, and built topological construction using them [3]. This framework enables to decrease computational times of localization and navigation, and to interact with robots using nature language. Galindo et al. improved performance of their planner using semantic information in maps[4].

In this paper, we focus on road structures and their semantics, and propose a system of building maps which includes those semantics and of navigation using these maps. This paper is organized as follows. We present an overview of the proposed method in section 2. In section 3, we introduce our personal mobility robot and its sensor system for outdoor navigation. In section 4, the geometric semantic map making system performed offline is described. Online navigation system using the semantic maps is presented in section 5, and a conclusion is given in section 6. In section 5,

2 Concept Description of Outdoor Semantic Maps

In this chapter, the advantages of semantic maps for autonomous navigation of mobile robots are introduced. Fig.1 shows a simple example of outdoor environment and the path, which we believe that a man traces, and Fig.2 shows the result of shortest path planning using a simple grid map. The path in Fig.2 has higher risk of colliding cars than one in Fig.1, and is difficult to accept.

In this study, we introduce a semantic map to make the mobile robot implement safer path planning as shown in Fig.1. Semantics useful for outdoor navigation

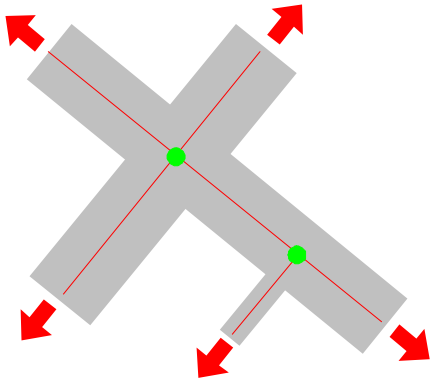

Fig. 3 A Geometrical Semantic Map; Green circles mean intersections, and red arrows indicates that the roads remains along these directions.

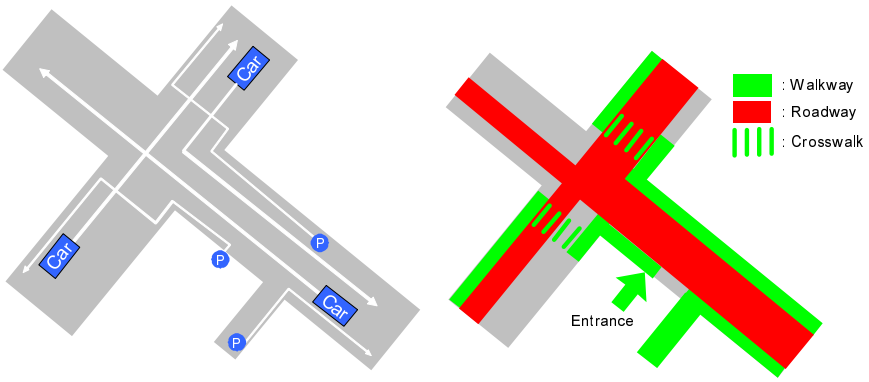

Fig. 4 Left: trajectory histories of moving objects ("P" means a pedestrian), Right: Traffic Semantics obtained from the histories. Green part indicates the area pedestrians moved around and the red part indicates the area cars moved around. The place where a men may appear (written as " Entrance ") , pedestrian crosswalks, and directions of cars in each part of roads can be found.

are classified into three categories, "Geometrical Semantics", "Traffic Semantics" and "Symbolic Semantics" in the study. Geometrical Semantics indicate geological properties, such as roads and intersections. Traffic Semantics indicate movement properties of pedestrians and cars on the roads, such as which side of the road they take and if cars can enter certain areas or not. Symbolic Semantics indicate pedestrian crosswalks, traffic signs, etc.

The Geometrical Semantic Maps are constructed as sets of roads. The maps have topological structures which connects each roads with intersections. By adopting topological structure, even if the map is not geometrically accurate, the positions of robots can be estimated as long as it can recognize intersections, landmarks, etc. Fig.3 shows an example of topographical semantic map based on road structure.

Other semantics are added in the Geometrical Semantic Maps. The Traffic Semantics can be obtained from trajectory histories of moving objects (Fig.4), and the

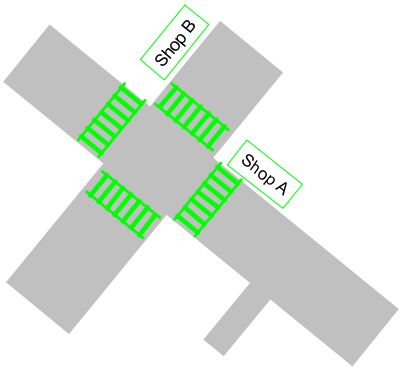

Fig. 5 Symbolic Semantics: It is possible to find pedestrian crosswalks and entrances from which pedestrians might appear.

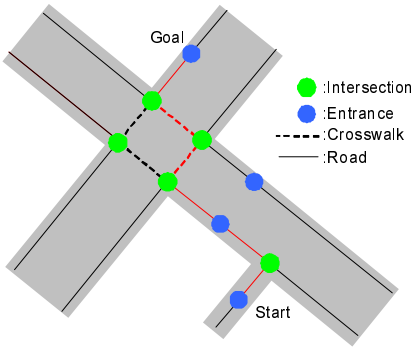

Fig. 6 Path planning result using a semantic map. Circles and lines indicate nodes and arcs respectively, and red lines are the path.

symbolic semantics are obtained from results of image processing(Fig.5). The Traffic Semantics in this example is incomplete, but the accuracy can be improved by moving in the same environment repeatedly and updating trajectory histories.

Using these semantics, path planning similar to the one shown in Fig.1 become possible (Fig.6. Arcs are placed edges of roads as possible. In addition, reasoning such as “the robot should cross roads at crossroads”, or “the robot should stop temporary before entrances or crossroads to confirm safe condition” will be applied to improve safety.

3 System Configuration of Personal Mobility Robot for Outdoor Environments

Fig.7 shows the mobile robot named “PMR”[5]. It is a single-seat two-wheeled inverted pendulum mobile robot and was originally developed by Toyota Motor Corporation as “MOBIRO”. As PMR can maintain horizontal stance on a slope, it has a small risk to fall in outdoor environments. Besides, PMR can climb over small bumps up to 50[mm]. The sensor system of PMR is constructed as follows. One LRF (Hokuyo Top-URG 100Hz version) is mounted on a pan-tilt head (TrackLabs Biclops PT), and used for 3D reconstruction described in section 4.1. The pan-tilt head swings the LRF between 0 degree (horizontal) and 50 degree downward, and it takes about 1.8[sec] each one-way swing. Another LRF (Top-URG 40Hz version) is fixed on the robot, and used for tracking of moving objects described in section 4.2. Also, PMR is equipped with an omni-directional camera (Opt NM33) for loop closing detection described in section 4.4 and a inertial measurement unit (Crossbow VG440) for accurate posture estimation. The camera is mounted on the pan-tilt head, and images are obtained only when the head becomes horizontal.

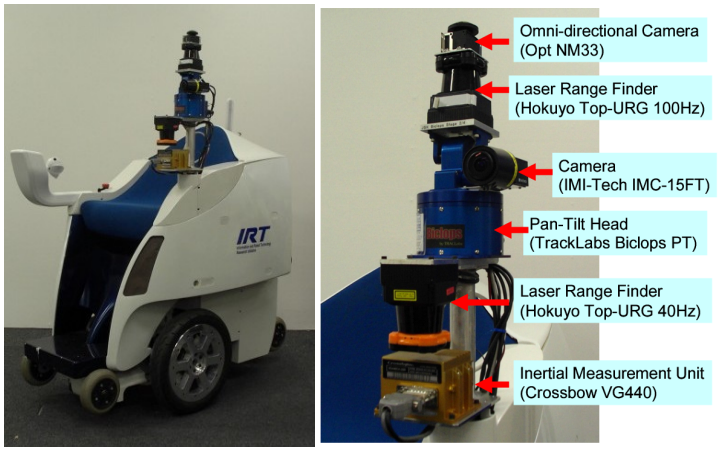

Fig. 7 Two-wheeled Inverted Pendulum Mobile Robot "PMR" and its sensor system

4 Semantic Map Making using Sensor Data Logs

This section describes a method to obtain semantic map defined in section 2 using sensors mounted on PMR. Once PMR runs manually in a targeted outdoor environment, a geometrical semantic map is obtained. From the second time, PMR can move automatically in the environment, and the map is updated with each moving.

Geometrical Semantic map making is performed as follows:

1. Local 3D maps are constructed using a swing LRF every time the LRF reaches top or bottom of swing. Local 3D maps are represented by DEMs (Digital Elevation Maps), and each DEM cell is classified as "Ground" and "Obstacle".
2. "Obstacle" cells in the DEM maps corresponding to moving objects are remove using results of tracking of moving objects.
3. Local Topological Map Patches are generated from DEMs. Topological Map Patches contain positions of intersection points and width of roads.
4. Topological Road Maps are generated from Map Patches. Once the robot turns at a intersection, a new road is added to the map.

4.1 Local 3D Map Generation using a Swing LRF

The proposed method makes local 3D maps using a swing 3D LRF for detection of static obstacles. It is capable of detecting slops and small bumps included in outdoor environments. 3D LRF continues to be swung while the robot is moving. The origin of each 3D map is the midpoint of grounding points of both wheels when the robot start swinging, and relative coordinates of the robot while moving are estimated using odometry.

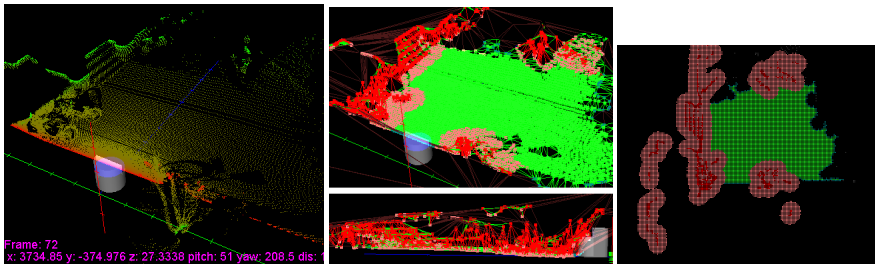

Fig. 8 A result of DEM making in a slope. Left: 3D point cloud. Center: DEM and Delaunay edges (perspective and sideways views), Right: Interpolated DEM. Green, red, pink, gray, cells mean ground, obstacle, near obstacle, and unknown respectively.

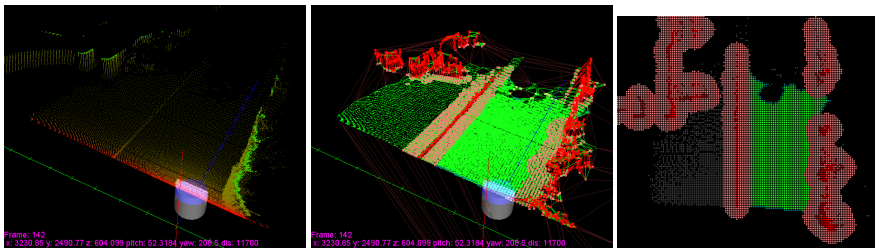

Fig. 9 A result of DEM making. The height of the bump in the center of DEM is about 80[mm]. Left side area of the bump are plane, but the robot cannot entered because of the bump. As a result, most of cells in these area classified as unknown.

As density of 3D LRF scan points are not uniform, the proposed method uses DEM (Digital Elevation Map) and Delaunay Triangulation. DEM is a kind of 2D grid map, and each grid cell has information of height. DEM enables to decrease number of points in dense area. If several LRF points exist in a DEM cell, the height of the cell becomes the height of the highest LRF point.

The resolutions of LRF scans far away from the robot are sparser than them of DEM, and Delaunay triangulation is used to deal with such situations. Points for Delaunay triangulation are center points of each DEM cell which LRF scans exist in. Heights of DEM cells which LRF scans do not exists in are estimated using Delaunay triangles which include the center points of them.

DEM cells are classified as “Ground”, “Obstacle”, “Near Obstacle”, and “Unknown”. It is impossible to classify a cell as “Ground” or “Obstacle” using only its height, because outdoor environments include slopes. Therefore, the proposed method uses gradients of Delaunay edges. If the gradient of one of edges connected to a DEM cell is larger than a threshold, that DEM cell is classified as “Obstacle”. Cells within the robot’s radius of “Obstacle” cells become “Near Obstacle” cells, which mean the center of the robot cannot enter these cells. Fig.8 and Fig.9 show the results of DEM generations. The size of each DEM cell is 100[mm].

4.2 Tracking and Identification of Moving Objects

Outdoor environments include not only static objects but also moving objects. It is difficult to track moving objects using a swing 3D LRF because of its slow swing speed. The proposed method uses another LRF fixed on the robot to detect and track moving objects.

The tracking algorithm is based on SJPDAFs (Sampling-based Joint Probabilistic Data Association Filters) [6]. SJPDAFs is a kind of multiple hypothesis tracking algorithms[7], and it is robust against false positives and negatives, and makes it possible to flexibly design individual trackers using particle filters.

In the targeted outdoor environments, several kinds of moving objects exist: pedestrians, bicycles, and cars. As the height of the fixed LRF on PMR is about 800[mm], shapes of LRF scans corresponding to pedestrians are almost same as to bicycles. However, sizes and shapes of LRF scans corresponding to pedestrians and cars are quite different from one another. Besides, as pedestrians often form groups, estimation of number of pedestrians in groups is needed.

The proposed method performs classification of clusters using SVM (Support Vector Machine), and has 7 classes: false positive, a car, a pedestrian, from 2 to 5 pedestrians. That is, the method solves estimation of number of pedestrians as a classification problem. As the shapes of LRF scan segments are not stable, the method adopts a time-series estimation.

We define the feature vector of LRF scans in a cluster at time t as $\boldsymbol{z}_f(t)$, and a set of feature vectors from time 0 to t as $Z_f^t = \{\boldsymbol{z}_f(0) \cdots \boldsymbol{z}_f(t)\}$. The value we want to estimate is $P(c_n|Z_f^t)$, we obtain

$$P(c_k(t)|Z_f^t) = \alpha \cdot P(z_f(t)|c_k(t)) \cdot P(c_k(t)|Z_f^{t-1}) \tag{1}$$

$$P(c_k(t)|Z_f^{t-1}) = \sum_n [P(c_k(t)|c_k(t-1) = n) \cdot P(c_k(t-1) = n|Z_f^{t-1})] \tag{2}$$

Also, from Bayes' theorem, we obtain

$$P(\boldsymbol{z}_f(t)|c_k) = \frac{P(c_k|\boldsymbol{z}_f(t))P(\boldsymbol{z}_f(t))}{P(c_k)} = alpha\frac{P(c_k|\boldsymbol{z}_f(t))}{P(c_k)} \tag{3}$$

$P(c_k|\boldsymbol{z}_f(t))$ can be estimated using SVM, and $P(c_k)$ can be estimated using training sets of SVM.

The features for SVM are defined as followed:

z_{f0} : Number of LRF segments
z_{f1} : Sum of lengths of LRF segments
z_{f2} : Average speed
z_{f3} : Difference between angle of directed bounding box and angle of average velocity vector
z_{f4} : Length of long side of directed bounding box
z_{f5} : Length of short side of directed bounding box
z_{f6} : Residual error between directed bounding box and LRF scan points

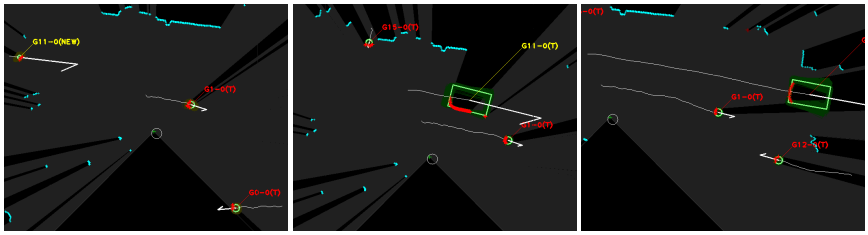

Fig. 10 Result of tracking and identification of moving objects. circles and squares indicate pedestrians and cars respectively. Although G11 was initially classified as a pedestrian when it was far away from the robot, correct classification occurred later.

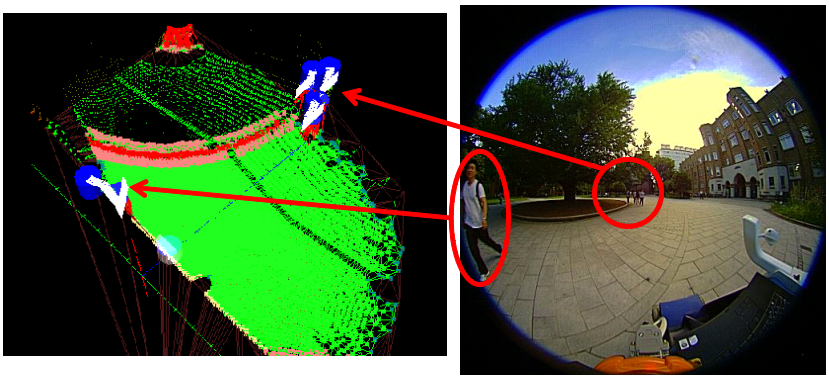

Fig. 11 Trajectory histories in a DEM and a image obtained by the front camera. White arrows indicate trajectories of moving objects.

All features are unaffected by distances and gradients of bounding boxes. Fig.10 shows a result of tracking and identification of moving objects. Fig.11 shows trajectory histories in a DEM. Cells which moving objects got through are classified as “Ground” or “Unknown”.

4.3 Making of Topological Map Patches and Reconstruction of Road Structures

The geometrical semantic maps defined in Section 2 consist of topological structures. The proposed system integrates local DEMs and builds several Topological Road Maps. Then, topological connection between roads using odometry logs and image processing.

Firstly, each DEM is converted to “Topological Map Patch”. Topological Map Patches consist of “Raster Segments”. Raster Segments are placed every 100[mm] (same as the size of DEM cell) along x axes, and parallel to their y axes. Note that the directions of x axes of DEMs and Topological Map patches are parallel to ones

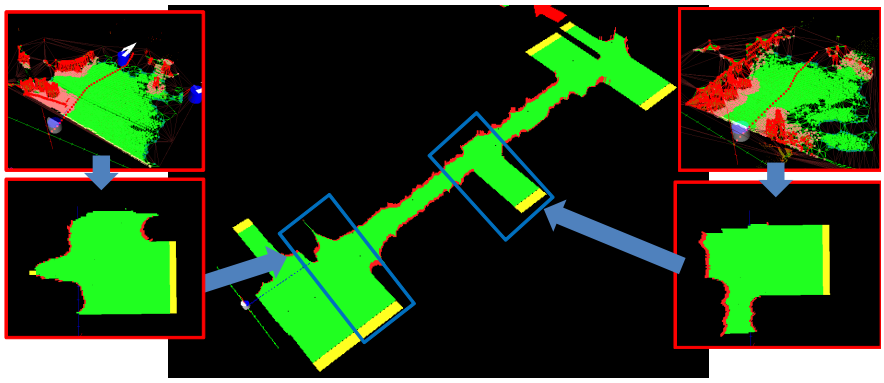

Fig. 12 A result of building of a Road Map using Topological Map Patches. The center image shows a Road Map, the lower side images show Map Patches which are parts of the Road Map, and the upper side images show DEMs before converted to the Map Patches. Red and yellow areas in the Map Patches and the Road Map indicate obstacles and open spaces (equal with intersections) respectively.

of roads. The length of a Raster Segment means the width of the road. Left and right edges of each raster have three kinds of status: “Obstacle”, “Open Space”, “Unknown”. “Open space” means that there are no obstacles within the range of the 3D LRF, and the system regard “Open space” as intersections. Intersections can become nodes which connects two roads.

Next, Road Maps are built using Topological Map Patches. Road Maps have same the structure as Topological Map Patches. Localization of Topological Map Patches are performed using the odometry of the robot, and if overlapped Raster Segments exist, newer ones are adopted. Fig.12 shows a result of Road Map Making.

Finally, topological connections between pairs of Road Maps are updated. Because map making is a offline process, the future trajectory of the robot from the time when the 3D LRF Data were obtained is known. If the future trajectory cross an intersections of the current Road Map, the system considers the robot transitions to another road, and new Road Map is created when the robot reach the coordinates of the intersection. Both Road Maps are connected with each other by a topological link. Besides, the proposed system solves loop problems using FAB-MAP[8]. FAB-MAP is the method to choose the most similar image with the current one from a sequence of images. The system makes a FAB-MAP using an omni-directional camera, and checks similarity every time the robot reaches a intersection. If FAB-MAP estimates that the robot already visited the current intersection, corresponding topological connection is created. Fig.13 shows results of semantic map building. Although the shape of the map in the left image skews because of the inaccurate odometry, the loop is processed appropriately.

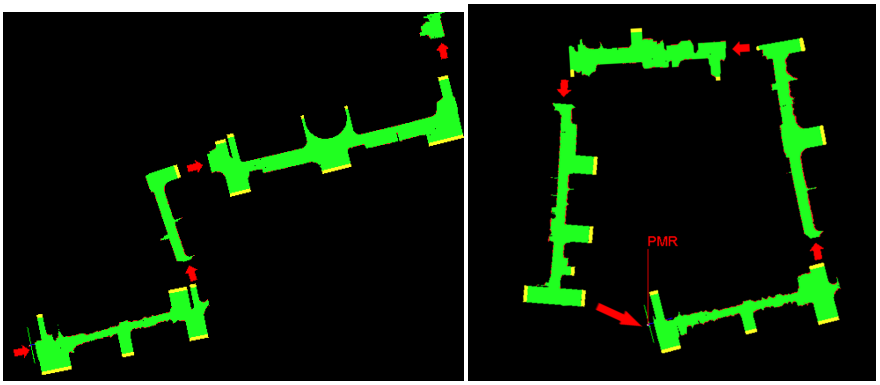

Fig. 13 The results of geometrical semantic map building. Red arrows indicate topological connections. The longest arrow in the right indicates a connection built by FAB-MAP.

5 Outdoor Navigation using a Semantic Map

This section describes navigation system in outdoor environments based on semantic maps. Although the algorithms in the previous section are performed off-line, algorithms in this section are performed on-line using a lap-top computer equipped with PMR.

5.1 Localization Algorithm Using a Geometrical Semantic Map

As mentioned above, a geometrical semantic map consists of several Road Maps. The Localization in this section means to estimate coordinates of robots in a Road Map. It is easy to estimate that which road is the road the robot exists in, because The robot is given the first road and the initial coordinates in the first road, and the localization enables to detect the intersection where the robot turn.

The localization method in the proposed system uses omni-directional images and topological map patches, and is performed in two stage. Firstly, the angle between the road which the robot currently exists in and the current direction of the robot is estimated using omni-directional images. Secondly, x and y position are estimated using a topological map patch. Note that x axis of a Road Map is parallel to the direction of the road, and the estimations of x positions mean to estimate how long the robot moved along the road.

The angle estimation is performed based on FAB-MAP mentioned in section 4.3. Firstly, FAB-MAP selects the most similar image with the current image in the image sequence obtained when the Road Map is constructed. Secondly, the angle estimation is performed using SURF feature points which are used for FAB-MAP. Only features in middle 240[pixels] (67.5[deg]) are used for the estimation, and it is assumed that the offsets along y direction are much smaller than distances the robots

Fig. 14 A result of the angle estimation using FAB-MAP. Upper image is the one selected as the most nearest to the lower one the by FAB-MAP. Yellow circles indicate SURF features used by the angle estimations, and green lines indicate the move distances between corresponding points.

and SURF feature points. The angle offsets are calculated simply using average horizontal offsets between corresponding points. Outliers are omitted by Smirnov-Grubbs test. Fig.14 shows a result of the angle estimation method.

The Next stage is the estimation of x and y positions in a Road Map using topological map patches. Before topological map patches are constructed, 3D LRF scan points are rotated in order that x axes of patches are parallel to x axis of the corresponding Road Map using the result of the angle estimation described above.

x positions are estimated using positions of intersections. Intersections in the Road Map corresponding to ones in the current topological map patch are selected, and x position of the robot is calculated using average values of positions of lower end of the intersections. If there are no intersections in the current map patch, x position is updated using the odometry of the robot.

y positions are estimated using x positions estimated above. First, we define averages values of y positions of obstacles in left and right sides of the current map patch as $\overline{y_{l,t}}$and $\overline{y_{r,t}}$ respectively. Then, corresponding areas of the Road Map to current map patch are extracted using the estimated x position of the robot. We define averages values of y positions of obstacles in left and right sides of the extracted part of the Road Map as $\overline{y_{l,g}}$and $\overline{y_{r,g}}$ respectively. Finally, the estimated y value of the robot $\overline{y_t}$ is calculated as following.

$$\overline{y_t} = ((\overline{y_{l,t}} - \overline{y_{l,g}}) + (\overline{y_{r,t}} - \overline{y_{r,g}}))/2 \tag{4}$$

5.2 Update of Semantic Information

Localization makes it possible for traffic streams or landmarks to be placed into the map. As mentioned above, semantic information is extracted from them. Fig.15 left shows trajectories of moving objects. PMR made 4 runs in that environment. The images in right side of Fig.15 show “Entrance Points”, that is, although the robot detected no intersections, pedestrians appeared out of those points. For safe

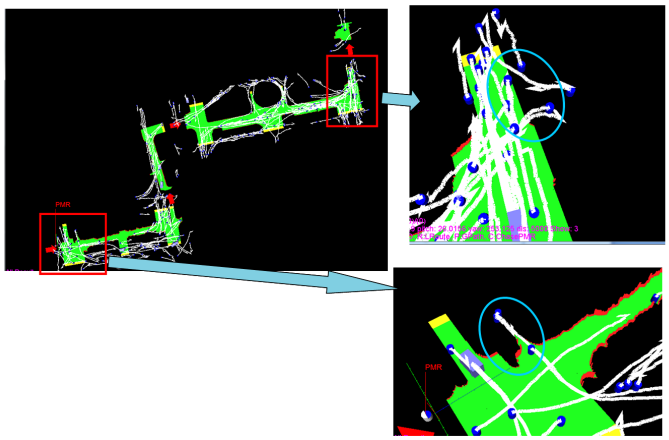

Fig. 15 Traffic Semantics : Detection of entrances

navigation, the robot should stop before those points, and check if pedestrians exist. Other than that, several kinds of semantics are extracted. For example, the robot can discern road ways and walk ways, and directions of road ways using trajectory histories of cars.

5.3 Navigation Experiment in an Outdoor Environment

The path planning in the propose system is performed in two stages, that is, global path planning and local path planning.

As mentioned in section 2, Global path planning makes rough routes to reach the destination using the semantic map. It decides that which side of each road the robot should move along and where the robot should cross roads. The proposed system places several types of nodes shown in Fig.16. All nodes are placed along right and left sides of each road. The types of nodes are following:

"Node Type 1" Starting and ending points of intersections.
"Node Type 2" Starting points of roads.
"Node Type 3" Points placed at the opposite sides of "Type 1" nodes. These nodes are not placed if the opposite sides are intersections.
"Node Type 4" "Entrance points" described in section 5.2.

The robot stops temporary in front of "Node Type 1"(only starting points) and "Node Type 4" to confirm the safety. Arcs are created as following. Firstly, adjacent nodes along x axis of each road are connected. Then, "Node Type 3" are connected to corresponding nodes in the opposite side are connected. Finally, "Node Type 2" connected to "Node Type 1" placed at the corresponding intersections. The costs of the arcs are calculated using length of arcs, but the cost of the arcs which indicate

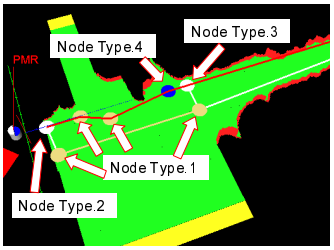

Fig. 16 Nodes in a Road Map

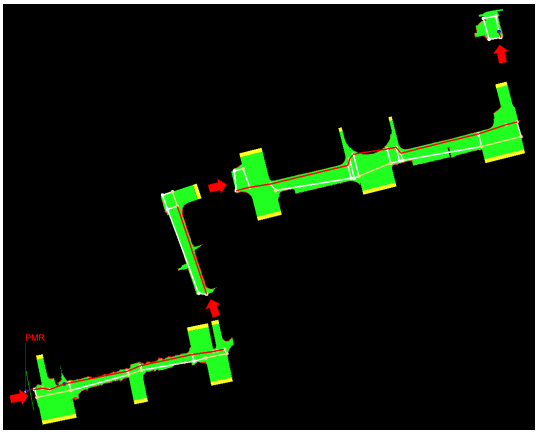

Fig. 17 A result of path planning in a semantic map (Red lines are the path)

that the robot crosses a road are doubled. This weighting intends to decrease the number of crossing of roads. Fig.17 shows a result of the global path planning.

Although the global path planning uses the semantic maps made by sensor data obtained in previous runs, local path planning uses the current 3D sensor data. This is because the outdoor environments might be changed by quasi-static objects like stopped cars or bicycles. The local path planning generates a line, along which the robot can come close to left or right edge of the road as possible.

Fig.18 shows the result of the navigation system. The semantic map used in this experiment is same as Fig.17. The moving distance of PMR was about 250[m].

6 Conclusion

In this paper, in order to realize autonomous navigation in outdoor environments which include moving objects, we proposed a system of semantic map making based on road structures. In the future, we aim to safer navigation system in consideration of human social nature, and system of conveying robot's intention to people around.

References

1. A. Nüchter and J. Hertzberg. Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, Vol. 56, No. 11, pp. 915–926, 2008.
2. I. Posnera D. Schroeter and P. Newmana. Online generation of scene descriptions in urban environments. *Robotics and Autonomous Systems*, Vol. 56, No. 11, pp. 901–914, 2008.
3. O. M. Mozos. *Semantic Place Labeling with Mobile Robots*. Springer Tracts in Advanced Robotics, 2010.

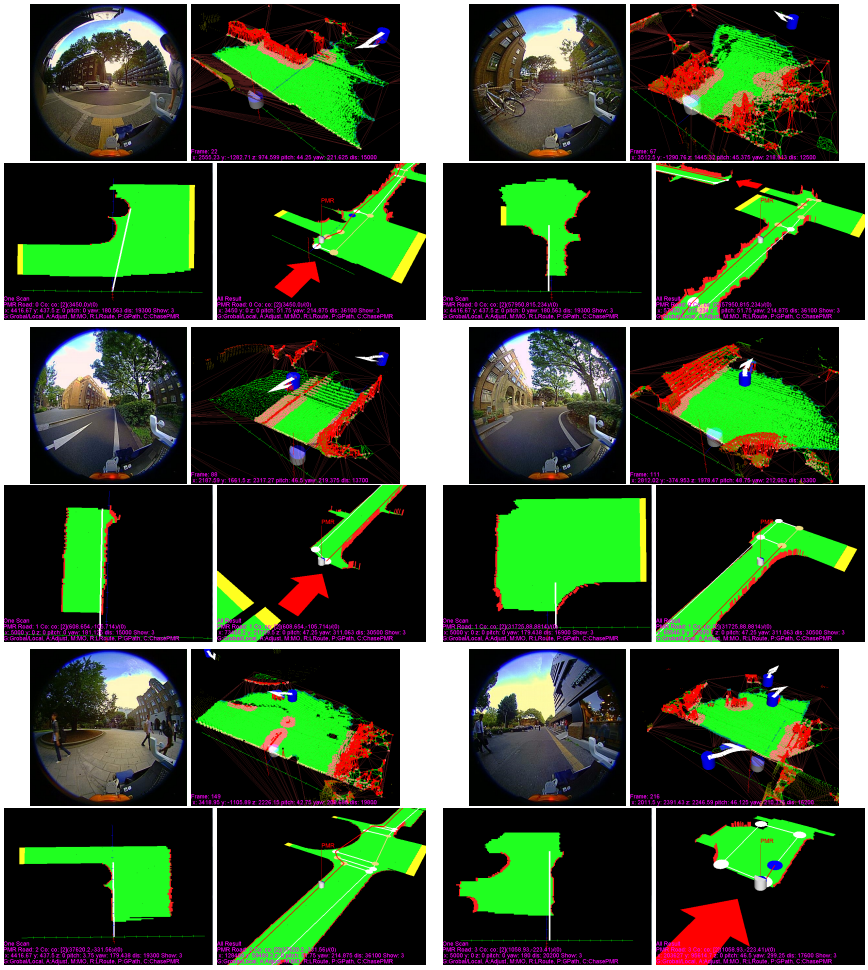

Fig. 18 Navigation experiment results. Upper left: front camera images. Upper right: DEMs, Lower left: local map patches and results of the local path planning. Lower right: results of the localization and the global path planning.

4. C. Galindo, J. Fernández-Madrigal, J. González, and A. Saffiotti. Robot task planning using semantic maps. *Robotics and Autonomous Systems*, Vol. 56, No. 11, pp. 955–966, 2008.
5. N. Hatao, R. Hanai, K. Yamazaki, and M. Inaba. Real-time navigation for a personal mobility robot in a environment with pedestrians. In *Proc. of 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'09)*, pp. 619–626, 2009.
6. D. Schulz, W. Burgard, D. Fox, and A. B. Cremers. People tracking with a mobile robot using sample-based joint probabilistic data association filters. *International Journal of Robotics Research*, Vol. 22, pp. 99–117, 2003.
7. Y. Bar-Shalom. Extension of the probabilistic data association filter to multi-target tracking. In *Proc. of the 5th Symposium on Nonlinear Estimation*, pp. 16–21, 1974.
8. M. Cummins and P. Newman. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, Vol. 27, No. 6, pp. 647–665, 2008.