Building Environmental Maps of Human Activity for a Mobile Service Robot at the ”Miraikan” Museum

Ippei Samejima, Yuma Nihei, Naotaka Hatao, Satoshi Kagami,
Hiroshi Mizoguchi, Hiroshi Takemura, and Akihiro Osaki

Abstract This paper describes environment maps that are comprised of the following three types of information, 1) 3D environmental changes that represents human activities, 2) human trajectories in 2D that represent how humans move in the environment, and 3) human posture data. These maps are utilized in order to plan safer, quicker and/or non-human-disturbing paths for a mobile service robot at the museum ”Miraikan”. Experiments are conducted within ”Miraikan” and results are shown.

1 Introduction

Autonomous carrier, cleaning, and security guard robots have recently become commercially available. The mobility function of such robots assumes quasi-static, indoor environments with few human inhabitants. Research in simultaneous localization and mapping (SLAM), mapping, localization, and path planning fields actively targets quasi-static environments. However, robots in human rich environments need further information about how humans use the environment to achieve safe and efficient behavior. Konolige et al. proposed lifelong environmental maps for daily living environments [1] to tackle this kind of problem that focused on temporal changes in map occupancy. Hamada et al. also proposed a human trajectory map that categorized human use by using moving speed and direction [2]. This information was used to bias results from robot path planning so that they could efficiently move through space cohabited by human beings.

We focused on a map used by a mobile service robot at a museum. Several such studies have been reported such as the RHINO robot at the Deutsches Museum in Bonn [3], and Minerva at the Smithsonian’s National Museum of American History

Ippei Samejima
Digital Human Research Center, National Institute of Advanced Industrial Science and Technology., 2-3-26, Aomi, Koto-ku, Tokyo, 135-0064, Japan, e-mail: i.samejima@aist.go.jp

[4]. The Minerva study demonstrated that robots in museum situations sometimes encounter crowds of people, and not just single individuals, and the typical interaction time is quite short.

We have also done studies at the ”Miraikan” Museum and identified three difficulties with mobility such as a) crowds of people moving in corridors and stopping in front of displays, b) frequently changing environments (such as fences, stools for small children, and guideboards), and c) changes in displays or exhibition areas. It is important for robots operating in museums, whatever their tasks are (such as guiding, giving tours, or cleaning), to better understand how humans use the environment. Therefore, we propose three types of maps to help robots understand the way humans use the environment: 1) a 3D environmental change map that represents human activities, 2) a map of human trajectories in 2D that represents how humans move in the environment, and 3) a human posture map to detect human interaction with museum displays.

2 Concept of Human Environment Maps

Maps are searched in the path planning phase for minimum cost paths by minimizing the accumulated costs that are assigned to each map cell. Although we can obtain minimum length paths, such a strategy might not be appropriate for museum environments. What we would like to obtain is somewhat safe, efficient, and non-human-disturbing paths. Three maps were constructed to obtain such paths: 1) a 3D environmental change map, 2) a map of human trajectories in 2D, and 3) a human posture map.

The robot could understand more about the environment by combining these three maps and generate better paths to reach its goal. This information was also useful for museum management purposes, in addition to such direct usage by the robot, such as how many people looked at displays, how long they spent looking at them, the human density of corridors, and the existence of bottlenecks.

3 3D Environmental Change Map

Static maps that include walls and columns of buildings are relatively easy to construct. However, there are objects in the environment whose locations/states can change, such as doors, chairs, fences, stools for small children, and guide boards. Also, museum displays occasionally change, and some displays are actually movable. Such objects can be avoided after they become visible to robot sensors by re-planning paths. However, when a robot becomes aware of typical changes in the environment, it can utilize that information to plan paths that traverse through areas more likely to be open space.

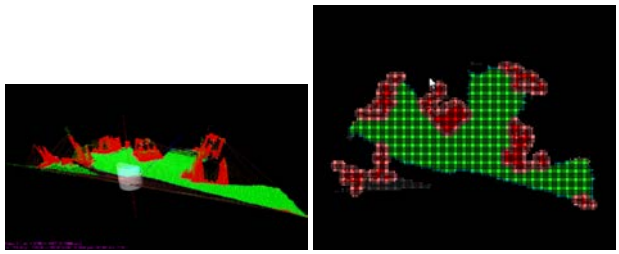

Fig. 1: DEM generation

3.1 Finding Regions with 3D Obstacles

There are obstacles with complex shapes in a museum environment that a robot needs to distinguish from the background to avoid possible collisions. A swinging laser range finder (3D LRF) was utilized to achieve this. The 3D LRF returned a point cloud that represented the surrounding environment. As the points were not distributed uniformly over 2D space, we utilized digital elevation maps (DEMs) and Delaunay triangulation to detect 3D obstacle regions [5]. A DEM is a kind of 2D grid map with each grid cell containing height information. DEMs enable the number of data to be reduced in dense sampling areas. If there are several 3D LRF points in a DEM cell, the height of the cell becomes the height of the highest 3D LRF point.

Figure 1 shows the results for DEM cell generation. Local path planning uses DEMs by converting “ground” into traversable regions and “obstacles” and “near obstacles” into non-traversable regions.

3.2 Generation of 3D Obstacle Changing Map

When the robot moves in the environment, 3D LRF data are always converted into local DEM cell classifications to avoid obstacles around the robot. The global 3D obstacle changing map is statistically updated with Eq. 1 by using this local DEM map.

$$cost = \frac{1}{t}\sum_{i=0}^{t} \|M_i - M_{i-1}\|, \tag{1}$$

where M indicates whether the cell is traversable (0), or not (1), and t is the number of updates. The path planner can utilize this 3D environmental change map to generate the path most likely to avoid potential collisions due to moving obstacles.

4 Pedestrian Trajectory Map

Useful information can be extracted by looking at human pedestrian behavior. For example, as regions that humans quickly traverse can be considered to be corridors, the robot can also utilize these regions. People walking in a museum tend to avoid going between displays and other people that are looking at the displays. Such behavior is observed and aggregated into maps, so that robots can plan similar paths.

4.1 Tracking and Identifying Pedestrian Trajectories

We often see groups of people moving together at museums. As they occlude each other in such places, it is difficult to track and identify human trajectories. We adopted cluster based sample-based joint probabilistic data association filters (SJPDAFs) [6] that use 2D LRF input to accomplish this.

SJPDAFs are robust against false positives and negatives, and they make it possible to flexibly design individual trackers using particle filters. The proposed method divides tracking targets and corresponding LRF segments into clusters, and classifies each cluster as a group of pedestrians. The number of pedestrians in each cluster is estimated independently, and each pedestrian in a cluster is tracked individually. Individual tracking enables the number of merging and splitting clusters to be estimated. Furthermore, the proposed method can remove the unwanted effects of false segmentation of LRF scans.

A SVM is adopted to classify and estimate the number of pedestrians . The method adopts time-series estimates as the shapes of LRF scan segments are not stable. The class definition of moving objects is c_0: false positive, c_1: one pedestrian, c_2: two pedestrians, and so on, where c_k indicates the label of each class. We define the feature vector of LRF scans in a cluster at time t as $z_f(t)$, and a set of feature vectors from time 0 to t as $Z_f^t = z_f(0) \cdots z_f(t)$. The value we want to estimate is $P(c_n|Z_f^t)$, and we obtain:

$$P(c_k(t)|Z_f^t) = \alpha \cdot P(z_f(t)|c_k(t)) \cdot P(c_k(t)|Z_f^{t-1}) \tag{2}$$

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

$$P(c_k(t)|Z_f^{t-1}) = \sum_n [P(c_k(t)|c_k(t-1) = c_n) \cdot P(c_k(t-1) = n|Z_f^{t-1})] \tag{3}$$

Also, from Bayes' theorem, we obtain:

$$P(z_f(t)|c_k) = \alpha \frac{P(c_k|z_f(t))}{P(c_k)} \tag{4}$$

Table 1: Numbers and frames in SVM training data set

Class	c_0	c_1	c_2	c_3	c_4	c_5
desc.	False positives	1 pedestrian	2 pedestrians	3 pedestrians	4 pedestrians	5 or more pedestrians
Numbers	212	318	54	9	5	3
Frames	868	20362	3631	352	267	58

$P(c_k|z_f(t))$ can be estimated using SVM, and $P(c_k)$ can be estimated using training sets of SVM.

The features for SVM are defined as follows:

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

Table.1 lists the total numbers of each class and the total numbers of frames detected in each class.

4.2 Aggregated Pedestrian Existence Map

An aggregated pedestrian traversal map is generated by accumulating detected pedestrian trajectories into each cell and normalizing them by time.

This map is useful for finding where humans can be reached in the entire area. If a cell and a surrounding area have no aggregated data, that may mean that humans are not allowed in the region or it is not it possible for them to traverse it, so even this is an open space for a robot, it may not be a good idea to use that region.

4.3 Aggregated Pedestrian Stopping Position Map

Slow or zero speed pedestrian trajectories are chosen to accumulate at each cell to generate an aggregated pedestrian stopping position map. If the average speed in the last N frame measurements is less than the threshold, v[m/s], this segment is treated as a stopping motion. The number obtained at each cell is normalized as previously described for the existence of aggregated pedestrians.

If the cell and surrounding area has a high stopping position score, humans tend to stay around that area. This often happens in front of displays and resting areas

in museums. A robot may not find it nice to traverse those areas since humans are concentrating on displays or are relaxing. In either case, they may not want to be distracted by the robot, or the presence of humans may cause the robot to potentially collide with them. However, as slowly moving humans in this area may need help with guidance, the robot may have to reach humans to interact with them using an additional margin for safety.

4.4 Variance in Pedestrian Velocity Map

Variations in the velocity of pedestrians at each cell were calculated from the walking speed of those who traversed each cell to generate this map.

When a group of people is moving in the corridor area, a few people first occasionally stop to wait for other people, where there are no displays. This information can be used to find corridor regions where people tend to stop and gather. This means there is enough space in the region, so that the robot can utilize such places for moving purposes, since there should be few obstacles around.

5 Human Posture Map

There are displays that can physically interact with human beings. The robot can find humans who are interacting with such displays by previously knowing where such displays are, and where interfaces are.

We adopted the Microsoft Kinect sensor as a marker-less motion capture system. This sensor was installed on a pan unit, and combined with the previously described pedestrian tracking system to control who (where) to look at. Controlling where to look at is required as the Kinect sensor has a limited working range, view angle, and maximum number of subjects that can be simultaneously tracked. We usually controlled Kinect to look at the closest human from the robot, but it could also track a specified human. Joint angles and limb lengths were obtained by utilizing the Kinect software development kit (SDK), and these data were mapped onto a 2D map.

Joint angles and limb lengths could also be used to identify where humans sat or rested.

The sitting posture was distinguished by joint angle and position conditions. Fig.2 shows three condition parameters. These values were thresholded by ratio p of observation time where all three parameter conditions were satisfied. The three parameters were:

- Hip height, h_{hip},
- Angle between the spine and the vertical direction, α_{spine}, and
- Shoulder width $d_{shoulder}$.

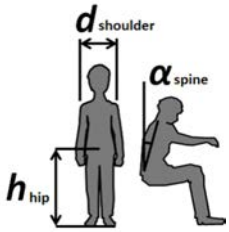

Fig. 2: Conditions to assess posture

Hip height h_{hip} was used to distinguish standing and sitting postures. α_{spine} had a maximum and minimum threshold to exclude forward and backward bent postures. Since Kinect SDK often incorrectly recognized the human skeleton when looking from the side of the subject, especially when the far side arm was hidden by the torso, $d_{shoulder}$ had a maximum and minimum threshold. Those three conditions were used to determine sitting/standing postures.

6 System and Environment Setup

Fig.3 has photographs of the robot that we used in the experiments. The mobile base was the Pioneer-III of Adept Mobilerobots. The onbody sensor for the 3D LRF was the Hokuyo UTM-30LX-F with a Pan-Unit from Sustainable Robotics, and that for the 2D LRF was the Hokuyo UTM-30LX with the Microsoft Kinect sensor. 2D LRF was not only utilized to track pedestrians racking, but also to generate 2D maps and 2D localization.

The experiments were conducted on the 3rd floor of the National Museum of Emerging Science and Innovation (Miraikan). Fig.4 shows a CAD drawing and photographs of the environment. The total area was about 120×35 [m]. There are displays with complex shapes in the environment as can be seen in Fig.4 (bottom two rows). There are also many stools for small children that can be moved.

A 2D map was generated before the experiment from 2D LRF data obtained by manually controlling the robot. Fig.4 shows a 2D map with a 5-cm grid.

7 Experiments

Three type of maps were generated and evaluated: 1) a 3D obstacle changing map, 2) a pedestrian trajectory map, and 3) a human posture map. The experiments were conducted at Miraikan as explained in the previous section.

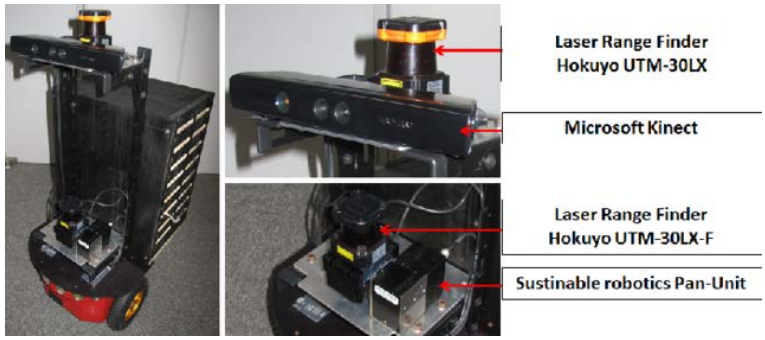

Fig. 3: Robot System

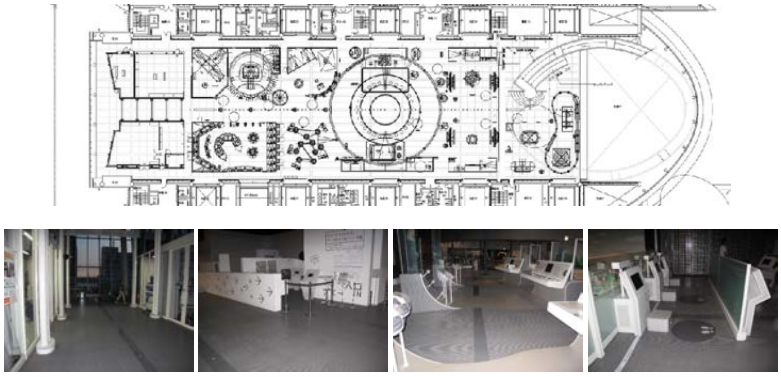

Fig. 4: CAD drawing of 3rd floor of “Miraikan” (top) and photographs of 3D obstacles (bottom)

7.1 Experiment to Generate 3D Obstacle Changing Map

The robot autonomously navigated itself by generating a path while the 3D LRF obtained data associated with localization results from a previously obtained 2D environmental map; a human operator manually provided input.

This experiment was conducted after the museum had closed and the exhibition areas were free of visitors (however, there were maintenance people present). The total length of the robot ’ s path was 1014.9 [m]. The DEM resolution was a 5-cm grid (the same as that for the 2D map) and the threshold for the Delaunay edge angle was set to 10 [deg]. DEM voting was limited to 5 [m] around the robot from the 3D LRF results since increasing the distance made sampling very sparse.

The size of the error ellipsoid area and variance in the ellipsoid direction calculated from the covariance matrix of the particle filter are summarized in Table.2 to evaluate 2D localization.

Table 2: Accuracy of localization

	Max.	Min.	Average	SD
Size of error ellipsoid area [m^2]	0.24	0.0	0.029	0.019
Variance in error ellipsoid direction [rad]	0.14	0.01	0.040	0.013

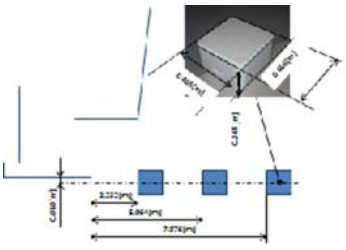

Fig. 5: Stools and their placement

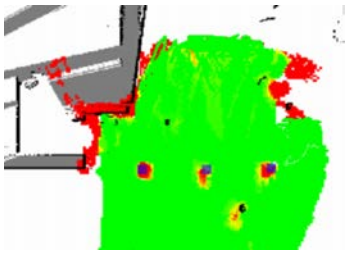

Fig. 6: Detected stools

Table 3: Accuracy of detecting 3D obstacles

	Average error	SD
Distance between centers of gravity [m]	0.19	0.06
Area size of obstacles [m^2]	0.11	0.10

7.1.1 Measurements of Accuracy of Detection

We manually moved the stools for small children into a flat area to evaluate the accuracy of detection, and manually measured their positions. Accuracy was evaluated by the size of the detected area and distance between the centers of gravity of the objects. Fig.5 outlines the experimental setup.

The same conditions as in the previous experiment were adopted, and robot behavior was also the same. The resulting map is given in Fig.6.

Table.3 summarizes the size of the detected area and the distance between stools.

Since the 2D map and DEM resolution were on 5-[cm] grids, we considered the results we obtained to have sufficient accuracy.

7.1.2 Detection of 3D Changing Environment

We manually moved the stools for small children into the display area to generate a 3D environmental change map, and the robot measured the previous stools.

Fig.7 shows the 3D environment change map. The red regions in the figure indicate changes that have happened. Because the stools were moved, there are yellow or red regions around the displays.

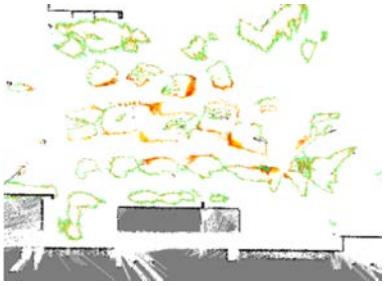

Fig. 7: Environmental change map

There are also yellow or red regions around the obstacles because of errors with localization, odometry, or 3D LRF.

7.2 Experiment to Generate Pedestrian Trajectories

7.2.1 Obtaining Pedestrian Trajectories

Measurements were conducted on one weekday from 11:00 am until 5:00 pm. Fig.8 shows the 2D LRF arrangement. Six 2D LRFs were placed on top of a tripod. Three SICK LMS200s were placed at points B, C, and D, and three Hokuyo UTM-30LXs were placed at points A, E, and F for 2D LRF. The height of the sensors was set so that it was 0.9 [m] above the ground and this was the same as the 2D LRF on the robot. The working range to detect pedestrians was limited to 20 [m]. The coverage of the sensors overlapped as can be seen from the figure.

Fig.9 shows the trajectories we obtained. Trajectories that were shorter than 5 [s] were omitted to prevent false positives from being detected caused by occlusion or noise. The total number of trajectories that remained was 126,839. The color indicates the average speed of each trajectory, where blue means slow and red means fast. As can be seen from the figure, visitors covered all possible movable regions.

7.2.2 Generation of pedestrian trajectory map

Three maps were generated from the previous experiment. Theses were 1) an aggregated pedestrian stopping map (Fig.10), 2) an aggregated pedestrian velocity map (Fig.11), and 3) a variance of pedestrian velocity map (Fig.12).

The red region in Fig.10 indicates a higher probability of pedestrians. Regions A and B in the figure are mostly red, and as they are close to displays they tend to stop there. However, region C in the figure is basically used as a corridor, and indicates

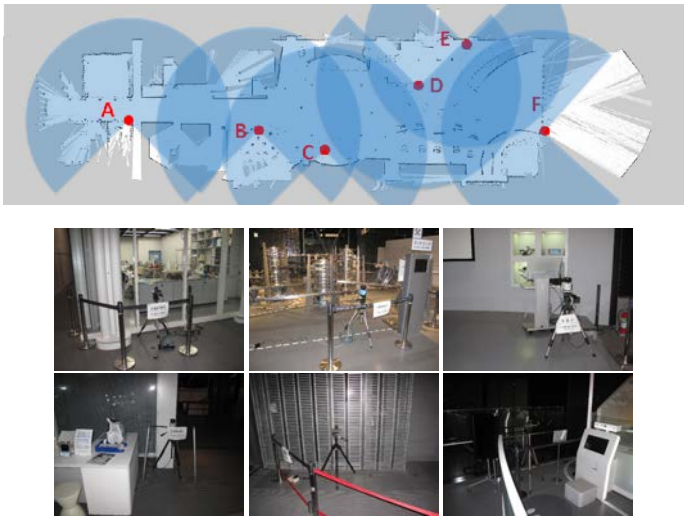

Fig. 8: Positions and photographs of sensors

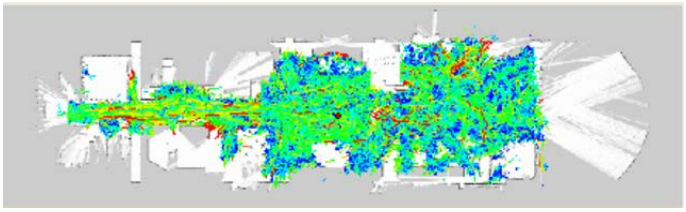

Fig. 9: Obtained pedestrian trajectories

a low probability of pedestrians. Consequently, these tendencies indicate that this map is of a real use environment.

Fig.11 indicates pedestrian speed and the red regions indicate faster motion. Region D in the figure, which is farther away from the wall, indicates faster motion. As this region has few obstacles to move around and there are no displays close by, humans there tend to move quickly.

Fig.12 indicates variance in the walking speed. The red region has higher variance. Region E is used as a corridor where there are no displays around, but sometimes people stop there to wait for other people in their group, so that variance increases. Region F is around a display and variance in speed tends to be small in such regions. However, as this region is at the entrance of a corridor, variance increases.

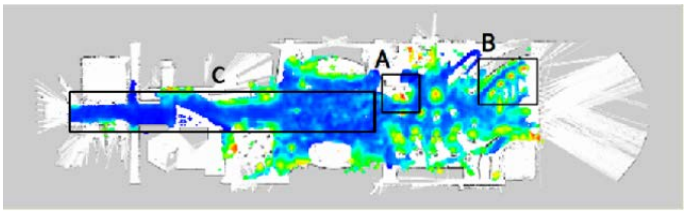

Fig. 10: Map of aggregated pedestrian stopping positions

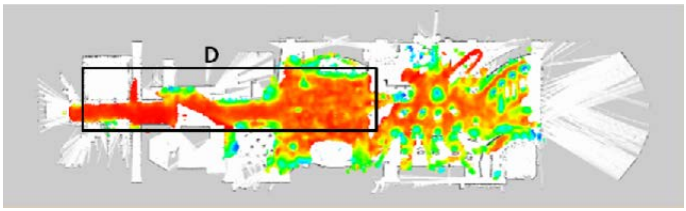

Fig. 11: Map of aggregated pedestrian velocity

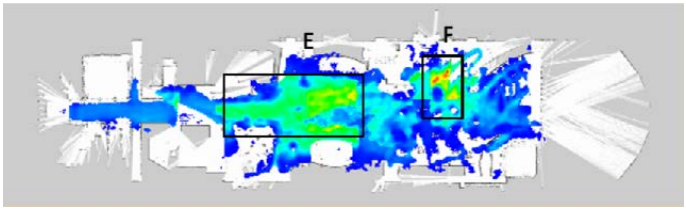

Fig. 12: Map of aggregated variance of pedestrian velocity

7.3 Experiment on generation of posture map

An experiment to detect human postures and map them onto a posture map was conducted at points A, B, and C in Fig.13. There are displays with physical interaction at points A and B, where point C is a technical information counter where visitors sit and ask the science staff of Miraikan questions.

The results are shown in Fig.14,Fig.15, and Fig.16. The yellow circle indicates the robot's position, the blue human posture indicates that the subjects were detected as standing, and the red human posture indicates that the subjects were detected as sitting.

All posture and detection results for standing or sitting postures were aggregated onto a map. Such a map can be utilized to identify where displays are with physical interactions, and where people are resting or standing.

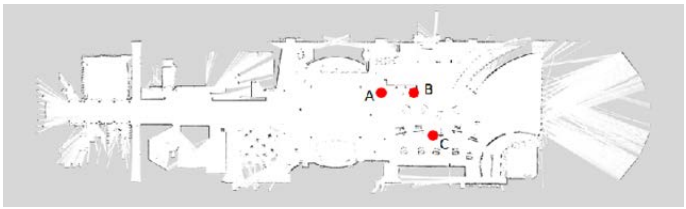

Fig. 13: Target position in experiment to generate posture map

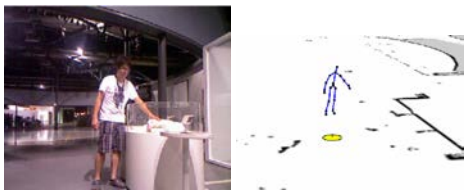

Fig. 14: Posture map experiment at point *A*

Fig. 15: Posture map experiment at point *B*

Fig. 16: Posture map experiment at point *C*

8 Conclusion

We described three types of methods of generating maps that were aimed at utilizing an autonomous robot in a museum environment, i.e., 1) a 3D environmental change map to identify human activities, 2) a pedestrian trajectory map to identify how humans use locations, and 3) a human posture map to identify how people interact with displays or the environment. Since simply a static map of environmental shapes was not sufficient in such a crowded environment, this new information should be useful in mobility based robotic services such as guidance, tours, and cleaning.

Fig. 17: A developing robot and children

This was a joint project undertaken with the National Museum of Emerging Science and Innovation (Miraikan), and the experiments were conducted on the third floor.

We are now jointly developing a mobile service robot that has an omni-directional telescopic microphone array for human interaction and is covered with a soft shell so that it does not injure visitors.

References

1. Kurt Konolige and James Bowman.Towards lifelong visual maps. Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'09),pp.1156-1163, 2009
2. Kaoru Hamada and Satoshi Kagami. Generation of Environmental Information Cost Map for Mobile Robots based on Measurement of Human Trajectory. The 28 th annual conference of the Robotics Society of Japan (RSJ 2010),3Q2-1, 2010.
3. W. Burgard and A.B. Cremers and Dieter Fox and D. H?nel and G. Lakemeyer and D. Schulz and W. Steiner and Sebastian Thrun. The Interactive Museum Tour-Guide Robot. Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), 1998.
4. Sebastian Thrun and M. Bennewitz and W. Burgard and A.B. Cremers and Frank Dellaert and Dieter Fox and D. Haehnel and Chuck Rosenberg and Nicholas Roy and Jamieson Schulte and D. Schulz. MINERVA: A second generation mobile tour-guide robot. Proc. of the IEEE International Conference on Robotics and Automation (ICRA'99), 1999.
5. Naotaka Hatao,Satoshi Kagami, Ryo Hanai, Kimitoshi Yamazaki and Masayuki Inaba. Construction of Semantic Maps for Personal Mobility Robots in Dynamic Outdoor Environments. Proc. of the Int. Conf. on Field and Service Robotics (FSR), 2012.
6. Y. Bar-Shalom.Extension of the Probabilistic Data Association Filter to Multi-Target Tracking. Proc. of the 5th Symposium on Nonlinear Estimation, pp.16-21, 1974.