

ND voxel localization using large-scale 3D environmental map and RGB-D camera

Shuji Oishi, Yongjin Jeong, Ryo Kurazume, Yumi Iwashita, and Tsutomu Hasegawa

Abstract— We propose an efficient 3D global localization and tracking technique for a mobile robot in a large-scale environment using 3D geometrical map and a RGB-D camera. With the rapid development of high-resolution 3D range sensors, high-speed processing of a large amount of 3D data is becoming an urgent challenge in robotic applications such as localization. To tackle this problem, the proposed technique utilizes a ND (Normal Distributions) voxel representation. Firstly, a 3D geometrical map represented by point-clouds is converted to a number of ND voxels, and local features are extracted and stored as an environmental map. In addition, range data captured by a RGB-D camera is also converted to the ND voxels, and local features are calculated. For global localization and tracking, the similarity of ND voxels between the environmental map and the sensory data is examined according to the local features or Kullback-Leibler divergence, and optimum positions are determined in a framework of a particle filter. Experimental results show that the proposed technique is robust for the similarity in a 3D environmental map and converges more stable than a standard voxel-based scan matching technique.

I. INTRODUCTION

For a mobile robot, a high-precise and reliable localization is a fundamental and indispensable requirement. More precisely, robot localization problem generally falls into the following categories:

- a) Tracking with prior knowledge such as a previous position or local travel distance
- b) Global positioning without prior knowledge

In addition, according to sensors used for the localization, they can be classified into the following types[1], [2].

- 1) Positioning using on-board internal sensors such as odometry or IMU
- 2) Positioning using on-board external sensors such as camera, LIDAR, or GPS
- 3) Positioning using off-board sensors installed in an environment[3]

The aim of this paper is to propose a new efficient localization technique in an indoor environment which is available not only for tracking but also for global positioning using an external 3D sensor. Especially we focus on a technique which can handle a huge amount of 3D range data taken by a 3D range sensor within a realistic processing time.

This work was supported by a Grant-in-Aid for Scientific Research (B) (23360115).

S. Oishi and J. Yongjin are with Graduate School of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan oishi@irvs.ait.kyushu-u.ac.jp
R. Kurazume, Y. Iwashita, and T. Hasegawa are with Graduate Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

A 2D laser range finder and an ultrasound sensor are very popular devices for mobile robot localization. For example, the most of modern 2D localization techniques utilize a 2D laser range finder. Robot positions are determined based on the similarity evaluation between a 2D environmental map and 2D measured range data[4]. However, owing to the development of a 3D laser scanner in recent years, a precise 3D environmental map is becoming available with a considerably low cost[5], [6], [7], [8], [9], [10]. In addition, a low cost range sensor such as Microsoft Kinect or Swiss Ranger SR4000 has been available on the market. These sensors called RGB-D cameras can capture 3D range data for a wide field of view in real-time ($\geq 30\text{Hz}$). Therefore, 3D localization using 3D range data is expected to be popular more and more in the near future.

In this paper, we propose a 3D localization technique using a large-scale 3D environmental map measured by a 3D laser scanner and a 3D range data captured by a RGB-D camera on a mobile robot. Conventional 3D localization techniques using 3D environmental information utilize a registration method based on point-to-point correspondence such as Iterative Closest Point (ICP) algorithm[5], [6], or voxel-to-voxel correspondence such as occupancy voxel counting[11], [12]. However, these techniques are computationally expensive or low accuracy due to the costly nearest point calculation or the discrete voxel representation, and hard to be applied for a global localization using a large-scale environmental map.

To tackle this problem, the proposed technique utilizes the idea of NDT (Normal Distributions Transformation)[13] for expressing a point distribution in a compact but information-rich form. Firstly, point-clouds in an environmental map are converted to the ND voxel representation. Then, to handle the characteristics of point distribution more efficiently, representative planes called eigen planes are extracted and registered as a new environmental map representation. Next, when a mobile robot scans the surrounding environment using a on-board RGB-D camera, an obtained 3D point-cloud is also converted to the ND voxel representation and eigen planes are extracted in a same way for the environmental map. In addition, seven representative points (six sigma points and a center point) are extracted and registered as additional features. Finally, the similarities between the environmental map and the measured data are examined based on plane-and-plane and point-and-plane correspondences. Using the obtained similarities, a particle filter is applied to find an optimum position which shows the maximum similarity between the environmental map and the measured data. In addition to the technique mentioned above, we examine a

statistical approach based on Kullback-Leibler divergence for the similarity calculation of ND voxels in the environmental map and the sensory data.

The NDT was originally proposed as an efficient alignment technique for two point-clouds[13]. It can reduce the computational cost for the alignment drastically compared to the ICP. On the other hand, the proposed technique focuses on another characteristic of the NDT, that is, the local point distribution can be stored efficiently in a compact and information-rich form. We utilize this characteristic for the localization of a mobile robot with low calculation cost.

Since wide-range and high-resolution 3D spatial data will be captured by a high-speed 3D sensor in the near future, the high-speed processing of a large amount of 3D data is an urgent challenge. The proposed technique enables to process a large 3D environmental map consisting of more than ten million 3D points and a 3D measured data with hundreds of thousands of 3D points by RGB-D camera in realistic processing time.

The remainder of the paper is organized as follows; After describing related researches in Section 2, the details of the proposed technique are introduced in Section 3. In Section 4, experimental results in an indoor corridor of $70 \times 35 \times 3$ [m] are shown using a high resolution and precise 3D model measured by a laser scanning robot[8] and a RGB-D camera (Kinect, Microsoft).

II. RELATED WORKS

2D (x and y coordinates and yaw angle) localization using a laser range finder or an ultrasound sensor is a standard technique for a mobile robot localization. In general, the obtained cross-sectional range information is compared with a 2D plane map which is measured and stored previously. 2D scan matching using Maximum Likelihood method or Iterative Closest Point (ICP) method, and Monte Carlo Localization such as a particle filter are very popular techniques[14], [15]. Konolige et al.[16] proposed a fundamental localization technique using a maximum likelihood method. In this method, by comparing a measured distance toward an arbitrary direction and the distance to the nearest object along this direction in an environmental map, the current position is estimated as the position in which the sum of differences is minimized.

On the other hand, owing to the recent development of a 3D laser scanner and a RGB-D camera, 3D (x , y , z coordinates and $roll$, $pitch$, yaw angles) localization is becoming popular more and more. Nüchter et al.[5], [6] applied the ICP method for sequential point-clouds measured by a laser range finder and achieved 3D (or 6Dof) slam.

Wülfing et al.[17] proposed a 3D localization by comparing a range data measured by a RGB-D camera and virtual range data synthesized from 3D environmental model using the ICP method. Fallon et al.[18] also proposed a similar technique using Z-buffer and Monte Carlo localization. Some researchers proposed a 3D localization by comparing 2D range data measured by a 2D laser range finder and cross-sectional shapes of a pre-scanned 3D environmental model

measured by a 3D laser scanner[9], [10]. Biswas et al.[19] proposed a 2D localization using a RGB-D camera and a 2D map. In their method, plane segments are extracted from a RGB-D image and the likelihood of robot location is examined by comparing them with a 2D vector map.

On the other hand, direct comparison between 3D voxels has also been presented so far. Olson et al.[11] proposed a localization technique which compares the distribution of occupancy voxels created by a stereo camera and a digital elevation map, and determines a position candidate using the shape correspondence by Most Likelihood method. Ryde et al.[20] also proposed a similar localization technique based on a multi-resolution occupancy voxels.

NDT[13] proposed by Biber et al. is an efficient registration technique which does not require a costly retrieval of nearest points in the ICP method. Though the accuracy is slightly lower than the ICP, the calculation cost can be reduced drastically and it can converge from a wide range of initial positions[21]. Magnusson et al. proposed a 3D localization of a mobile robot using NDT in a tunnel[22]. They also proposed a loop detection technique using histograms of local features obtained by NDT[23].

III. LOCALIZATION USING 3D ENVIRONMENTAL MAP AND RGB-D CAMERA

This section introduces the proposed technique for robot localization using a large-scale 3D environmental map and a RGB-D camera.

A. 3D environmental map

We assume that a 3D geometrical model has been created by a laser scanner such as the system proposed in [8]. This system is able to construct a large-scale 3D model consisting of more than ten million 3D points by an on-board laser range finder from multiple viewpoints. Firstly, we convert this geometrical model to a discretized 3D map which is constituted by a number of ND voxels[13] as shown in Fig.1. ND voxel is a discrete representation of point-clouds by approximating a point distribution with a multi-dimensional normal distribution. To convert a point-clouds to ND voxels, the average and the covariance matrix of the point distribution are calculated and the eigen values and the corresponding eigen vectors are extracted by the eigenvalue decomposition. Then, in each ND voxel, a representative plane named “eigen plane” is extracted. Eigen plane is a small facet which has a normal vector toward the eigen vector with minimum eigen value.

The geometrical model in Section IV consists of 40,000,000 points for an area of $70 \times 35 \times 3$ [m] as shown in Fig.2. After converted to the ND voxels with the size of 800 [mm] on a side, we obtained the 3D environmental map consisting of 50,000 voxels. The ND voxels and the eigen planes are shown in Fig.3. During this process, we also extract floor regions according to the directions of the normal vectors of eigen planes and create a 2D floor map in which the robot can move without the collision with walls

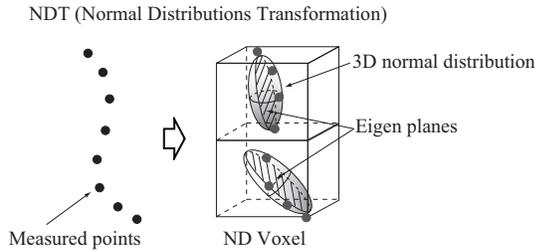


Fig. 1. Concept of NDT and ND voxels[13]

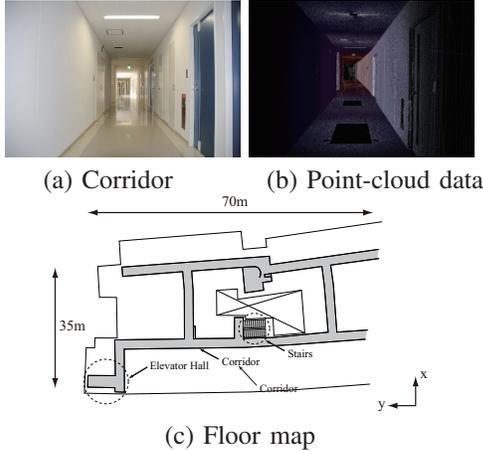


Fig. 2. Photo and 3D model of corridor

and static obstacles. These regions are used for initializing and updating particles in a particle filter.

Note that to suppress the discretization error, we adopted a similar technique with Biber's one[13], that is, adjacent voxels are overlapped each other so that the centers of voxels are displaced with a half of the voxel size as shown in Fig.4. As a result, every point in 3D space is involved with eight adjacent voxels.

B. Measured data

In this paper, we focus on a localization problem for a mobile robot equipped with a RGB-D camera such as Microsoft Kinect shown in Fig.5. An example of a point-cloud taken by a RGB-D camera in one shot is shown in Fig.6. This point-cloud consists of 300,000 3D points.

To accelerate the localization process, we again apply the NDT to this measured data and obtain ND voxels. Then, as shown in Fig.7, six points (sigma points, see appendix) and one center point of the normal distribution are extracted as seven representative points in each ND voxel. In addition, the eigen plane is also extracted in the same way as for the environmental map. Note that adjacent ND voxels in the measured data are also overlapped to suppress the effect of discretization error.

C. Hierarchical ND voxels

As described above, the environmental map and the measured data are represented by overlapped ND voxels. Therefore, if we would like to describe more complex shapes of the environment, smaller size of voxels should be adopted, but

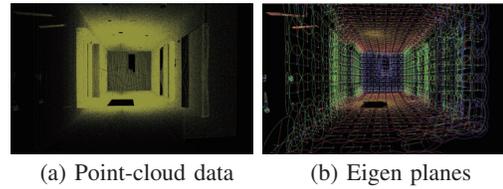


Fig. 3. Point-cloud data and eigen planes by RGB-D camera

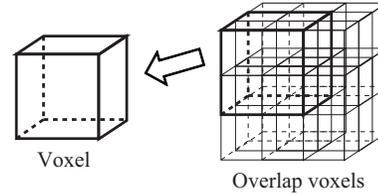


Fig. 4. Overlapped ND voxels

the computational cost will be increased. On the other hand, if we use larger voxels, the processing speed will be faster but the positioning accuracy will be reduced. Therefore, we adopt hierarchical ND voxels consisting of different sizes of voxels. The level of the voxel size is adjusted according to the convergence of the position estimation.

D. Global positioning and tracking by particle filter

Global positioning and tracking are both performed by a particle filter using the ND voxel representation for the environmental map and the measured data. Here, each particle holds a candidate position \mathbf{t} and orientation \mathbf{R} of the robot, and the likelihood of the particle is calculated according to the following procedure (Fig.8).

Firstly, the position of the ND voxels in the measured data, which is represented by a local coordinate frame, are transformed according to the candidate position and orientation in the particle and compared with the ND voxels in the environmental map.

The score of each ND voxel is calculated in two ways, that is, the similarity calculation using eigen planes and Kullback-Leibler divergence.

1) *Similarity calculation using eigen planes*: The score is calculated based on the distance from the seven representative points in the measured data and the eigen plane in the environmental map, and the angular difference of normal vectors in both eigen planes as shown in Fig.9.

More concretely, suppose $\mathbf{S}_{ik} = (S_{ikx}, S_{iky}, S_{ikz})^T$ as the k^{th} ($k = 1 \sim 7$) representative point and $\mathbf{N}_i = (N_{ix}, N_{iy}, N_{iz})^T$ as the normal vector of the eigen plane in the measured data. The representative point $\tilde{\mathbf{S}}_{ik}$ and the normal vector of the eigen plane $\tilde{\mathbf{N}}_i$ after the coordinate transformation are given as

$$\tilde{\mathbf{S}}_{ik} = \mathbf{R}\mathbf{S}_{ik} + \mathbf{t} \quad (1)$$

$$\tilde{\mathbf{N}}_i = \mathbf{R}\mathbf{N}_i \quad (2)$$

Then, the overlapped ND voxel m ($m = 1 \sim 8$) in the environmental map which contains the transformed representative point is selected, and the distance $d_{ik \rightarrow m}$ from



Fig. 5. Mobile robot equipped with four RGB-D cameras (Kinect)



Fig. 6. Point-cloud data by a RGB-D camera

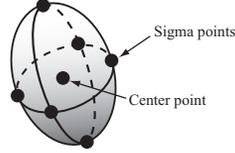


Fig. 7. Seven representative points

the representative point in the measured data to the eigen plane of the voxel in the environmental map is calculated.

$$d_{ik \rightarrow m} = | N_{mx}(\tilde{S}_{ikx} - \mu_{mx}) + N_{my}(\tilde{S}_{iky} - \mu_{my}) + N_{mz}(\tilde{S}_{ikz} - \mu_{mz}) | \quad (3)$$

Where N_m is the normal vector of the eigen plane and μ_m is the center of the normal distribution in the ND voxel m in the environmental map.

Finally, the score $\alpha_{ik \rightarrow m}$ of the representative point \tilde{S}_{ik} is given as

$$\alpha_{ik \rightarrow m} = \frac{1}{\sqrt{2\pi}\sigma_d} e^{-d_{ik \rightarrow m}^2 / \sigma_d^2} \quad (4)$$

Where, σ_d is a parameter of the variance of the distance $d_{ik \rightarrow m}$.

In the same way, the score $\beta_{i \rightarrow m}$ of the angular difference of normal vectors in both eigen planes is given as

$$\beta_{i \rightarrow m} = | N_{mx}\tilde{N}_{ix} + N_{my}\tilde{N}_{iy} + N_{mz}\tilde{N}_{iz} | \quad (5)$$

Then, the score γ_{ik} of the representative point S_{ik} is calculated as the maximum value of the product of the scores $\alpha_{ik \rightarrow m}$ and $\beta_{i \rightarrow m}$ among the overlapped voxels.

$$\gamma_{ik} = \max_{1 \leq m \leq 8} \alpha_{ik \rightarrow m} \beta_{i \rightarrow m} \quad (6)$$

Next, by summing up the scores for seven representative points, the final score δ_i of the voxel i in the measure data are obtained.

$$\delta_i = \sum_{k=1}^7 \gamma_{ik} \quad (7)$$

Finally, the likelihood λ of the particle is calculated as the sum of the scores in all voxels in the measured data.

$$\lambda = \sum_{i=1}^W \delta_i \quad (8)$$

Where, W is a number of voxels in the measured data.

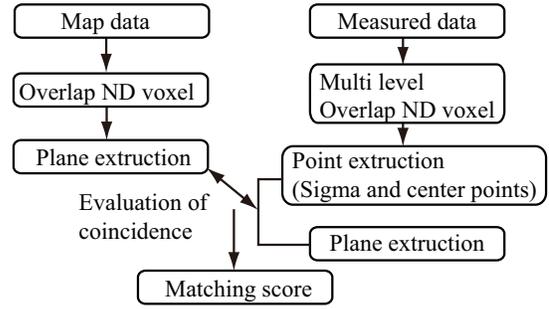


Fig. 8. Flow of likelihood calculation

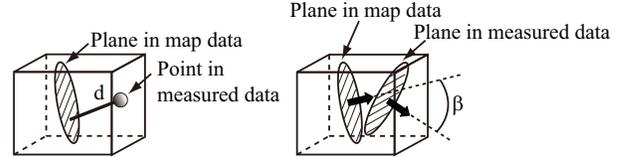


Fig. 9. Evaluation of distance and orientation

2) *Similarity calculation using Kullback-Leibler divergence:* In this method, the score of each ND voxel is calculated according to the Kullback-Leibler divergence.

Suppose $\mathcal{N}(\mu_i, \Sigma_i)$ is the normal distribution in the measured ND voxel V_i . Here, μ is the average of the points in the ND voxel and Σ is the covariance matrix. Then, the overlapped ND voxel V_m ($m = 1 \sim 8$) in the environmental map which contains the transformed center point $\tilde{\mu}_i = \mathbf{R}\mu_i + \mathbf{t}$ is selected. We denote the average and the covariance matrix of the overlapped ND voxel as $\mathcal{N}(\mu_m, \Sigma_m)$. The Kullback-Leibler divergence between these ND voxels is obtained as follows:

$$D_{KL}(\mathcal{N}(\mu_i, \Sigma_i) || \mathcal{N}(\mu_m, \Sigma_m)) = \frac{1}{2} \left(\log\left(\frac{\det \Sigma_m}{\det \Sigma_i}\right) + \text{Tr}(\Sigma_m^{-1} \Sigma_i) + (\mu_m - \mu_i)^T \Sigma_m^{-1} (\mu_m - \mu_i) - 3 \right) \quad (9)$$

Then, the score δ_i of the measured ND voxel is calculated as the minimum value among the overlapped ND voxels of the model.

$$\delta_i = \min_{1 \leq m \leq 8} D_{KL}(\mathcal{N}(\mu_i, \Sigma_i) || \mathcal{N}(\mu_m, \Sigma_m)) \quad (10)$$

Finally, the likelihood λ of the particle is calculated as the sum of the scores in all voxels in the measured data.

$$\lambda = \sum_{i=1}^W \delta_i \quad (11)$$

Where, W is the number of voxels in the measured data.

Note that the number of initial particles is 72,000 for 72 different directions at 1,000 points in 3D selected randomly on floor regions. After the first iteration, the number of particles is adjusted according to the KLD sampling technique[24] from 1,000 to 5,000 adaptively.

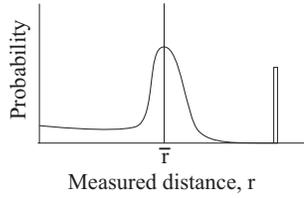


Fig. 10. Probability of sensor reading

E. Scan matching using Maximum Likelihood method based on beam model

The scan matching based on the Maximum Likelihood method is a popular localization technique[11],[20]. In general, the environmental map represented by a voxel map and the measured distance by a range sensor are compared directly.

Suppose the distance from a robot position s to a measured point as r , and the distance from a robot position s to a nearest obstacle in the environmental map along with the same direction as \bar{r} in general. The observation probability $p_s(r|\bar{r})$ (beam model) of the range sensor is given as shown in Fig.10 [25].

By assuming the observation is independent, the robot position is estimated by the maximum likelihood method as the following equation.

$$\arg \max_s p(s|r) = \arg \max_s \prod_i p_s(r_i|\bar{r}_i) \quad (12)$$

Where r_i is the measured distance in i^{th} observation.

This method can be extended for 3D voxel data. Suppose the robot position as s and the measured point such as the center of the voxel as r_j . The estimated position can be obtained by Eq.(12) if some voxels exist along the line connecting s and r_i and the center of the closest voxel is given as \bar{r} .

In addition, when we apply a particle filter to estimate a robot position, the likelihood λ of each particle can be defined as

$$\lambda = \prod_i p_s(r_i|\bar{r}_i) \quad (13)$$

In the experiments in Section IV, we adopted a simplified model of Fig.10 as

$$p_s(r_i|\bar{r}_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(r_i-\bar{r}_i)^2/\sigma^2} \quad (14)$$

IV. LOCALIZATION EXPERIMENTS

Localization experiments were conducted using a mobile robot shown in Fig.5. In this experiments, the robot moved in a corridor of $70 \times 35 \times 3$ [m] shown in Fig.2 while measuring range data by a RGB-D camera (Kinect, Microsoft). Two kinds of experiments were carried out.

- 1) Global positioning using single range data
- 2) Tracking using sequential range data and dead reckoning information

The environmental map used in the experiments consists of 51,395 ND voxels and the size of each voxel is 800[mm]

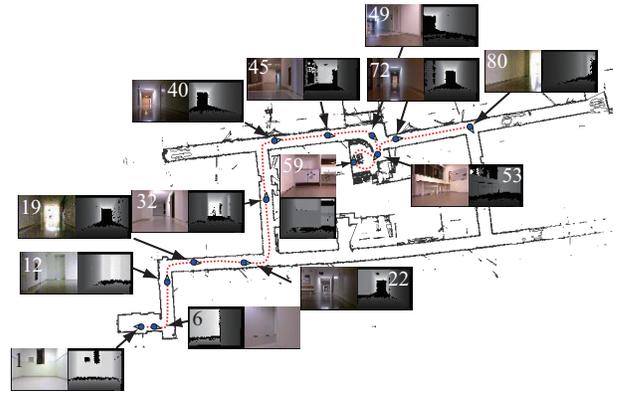


Fig. 11. Measuring positions for global positioning

on a side. On the other hand, the range data taken by the RGB-D camera consists of 307,200 points and are converted to 121 and 670 overlapped ND voxels of 1600 [mm] and 800 [mm] on a side, respectively.

A. Global localization experiments

Firstly, the robot captured an range image by the RGB-D camera and estimated the current global position without any knowledge about the position in the corridor shown in Fig.2(c).

To compare the performance of the proposed and conventional scan matching methods, we adopted the following techniques for the calculation of the likelihood in the particle filter.

- 1) Scan matching technique in Section III-E[11],[20]
- 2) Proposed technique using the ND voxels

In this experiment, we assumed that the robot is equipped with an acceleration sensor and roll and pitch angles are measured directly by this sensor. Thus each particle holds four kinds of information (position x, y, z and orientation yaw).

Firstly, the robot moved the corridor along the dotted line in Fig.11. Then the robot stopped at 80 positions along the line and captured 80 range images by the RGB-D camera. Fig.11 shows the examples of the range and color images and the positions where the robot captured the images. As seen in this figure, quite similar range and color images are taken in different positions in the corridor. Therefore, this environment is quite difficult for the global localization since perceptual aliasing tends to be happened frequently.

Next, the obtained range data was transformed to the overlapped ND voxels with a size of 1600 [mm] on a side. Then the global positions of 80 range images were estimated “independently” by the particle filter. The number of initial particles is 72,000 but it was decreased from 1000 to 5000 adaptively after the first iteration. The final position was determined after four iterations by the particle filter as a particle’s position with the highest likelihood.

Among the 80 positions, we determined that the position was correctly estimated if the positioning error is less than 500 [mm] in x, y , and z directions and the orientation error

TABLE I
CORRECTNESS OF GLOBAL POSITIONING

σ, σ_d [mm]	100	300	500	800	1000
Proposed (Sec. III-D.1) [%]	23.8	27.5	28.8	18.8	23.8
Proposed (Sec. III-D.2) [%]	16.3				
Scan matching [%]	2.5	1.3	1.3	2.5	1.3

is less than 10 [deg.]. The correctness of the proposed and the scan matching techniques (Section III-E) is compared in Table I.

From Table I, it is clear that the correctness of the global localization by the proposed technique using sigma planes outperforms the ones by the Kullback-Leibler divergence and the scan matching technique. This is partially due to the fact that the point distribution in each voxel is considered and evaluated efficiently by the overlapped ND voxel representation. On the other hand, the scan matching technique just evaluates the distances between the centers of voxels, and the point distribution is not taken into account. In addition, the Kullback-Leibler divergence is calculated only for a voxel, not overlapped ND voxels, in the map corresponding to a ND voxel in measured data.

Note that, for the scan matching technique, we evaluated the performance for various voxel sizes from 1600[mm] to 400[mm] on a side. The calculation time increased from 20 minutes to 8 hours, however, the correctness did not changed so drastically.

The average of the total calculation time (Intel(R) Xeon(R) CPU 2.67GHz Quad core 4GB) was 75.4 [sec.] for four iterations of the particle filter. The processing time for the likelihood calculation in one particle is 0.39 [msec.] for the scan matching technique and 1.15 [msec.] for the proposed technique, respectively. Since the overlapped ND voxels are adopted in the proposed technique, eight voxels are always evaluated for the likelihood calculation in each particle. The correctness without the overlapped ND voxels is, however, decreased from 28.8 % to 17.5 % in the case of $\sigma_d = 500$ [mm].

B. Tracking experiments

Next, we carried out the tracking experiments by combining the proposed localization technique using eigen planes (Section III-D.1) and the odometry information. Initial position in the environmental map was determined by the global localization mentioned above.

Firstly, the tracking result using the odometry information is shown in Fig.13. It is clear that the tracking error is gradually accumulated if only the odometry information is available.

Next, we evaluated the tracking performance by combining the proposed technique and the odometry information. At first, the robot was placed in position 1 in Fig.11, then the global position was estimated using the overlapped ND voxel with 1600 [mm] on a side. After the initial position was estimated, the size of the overlapped ND voxel was changed to 800 [mm] and the robot position was tracked

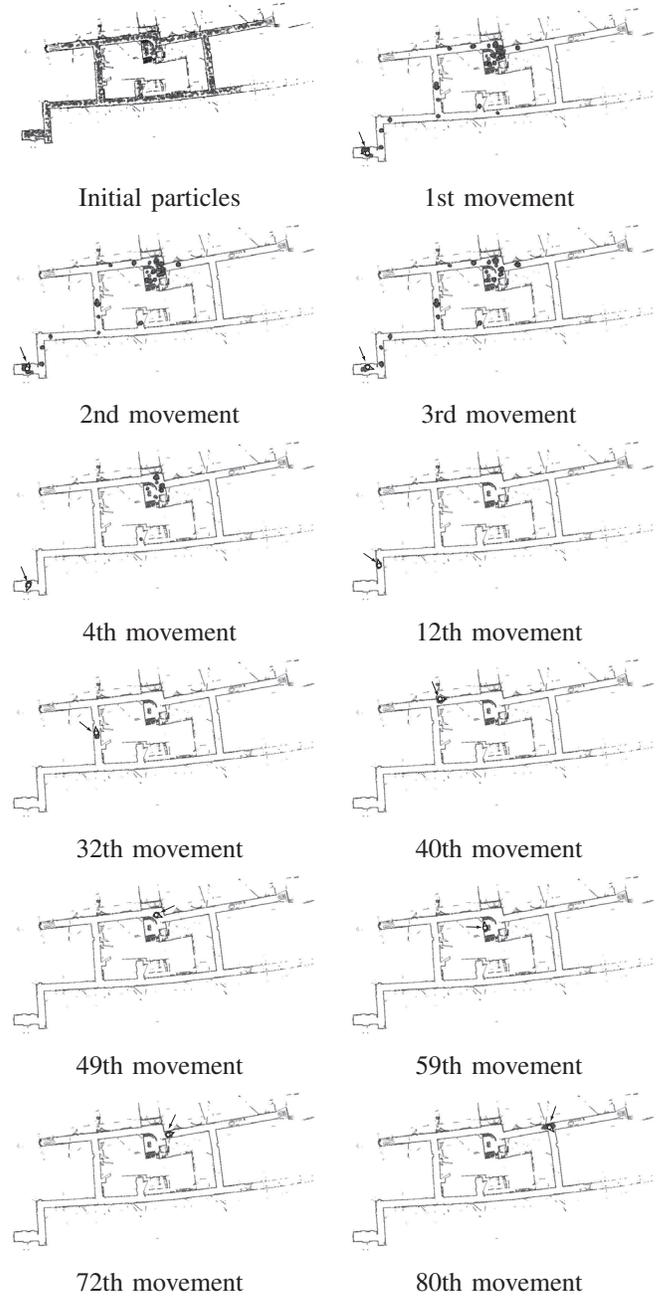


Fig. 12. Convergence of particles and estimated trajectory

with the particle filter consisting of up to 5000 particles. The experimental results are shown in Fig.12.

Figure 13 shows the tracking paths estimated by the proposed technique and the odometry information. Paths for x, y, and z direction are shown in Fig.14. Especially, at the beginning of the tracking, the proposed technique can estimate the position more stable than the scan matching technique since the ND voxels can provide richer information such as sigma planes than the normal voxels. Moreover, it is clear that the estimated height (z direction) of the robot by the proposed technique is quite stable. Since the floor of the environment was flat, the proposed technique identified the robot position more stably.

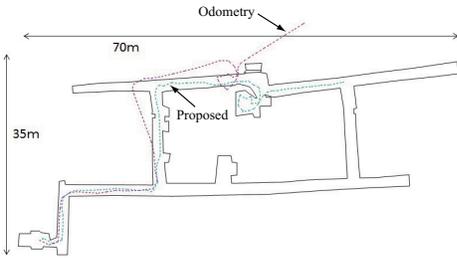


Fig. 13. Estimated trajectory by odometry and the proposed method

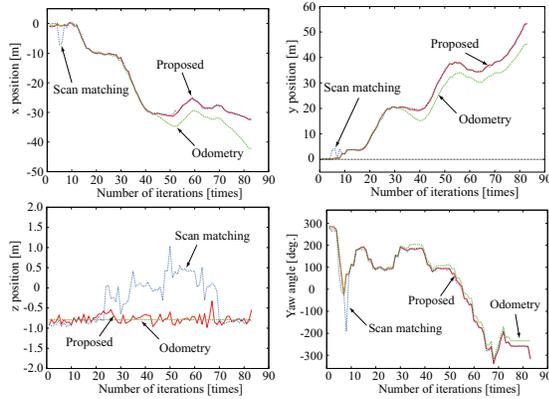


Fig. 14. Comparison of estimated trajectories

Finally, the computational time of the position estimation by the particle filter was 7.2 [sec.] on average (1379.3 particles, the size of ND voxel 800 [mm]) and 2.5 [sec.] on average (1385.2 particles, the size of ND voxel 1600 [mm]) for Intel(R) Xeon(R) CPU 2.67GHz Quad core 4GB. Note that, if we apply the ICP in Point Cloud Library[26] for two range data (307,200 points) taken by the RGB-D camera, the processing time was 118.5 [sec.] on average for one iteration of each particle. Thus if we apply the ICP for 5000 particles, unrealistic processing time such as 164 hours will be required for ONE update by simple arithmetic.

V. CONCLUSIONS

The paper presented a global localization and tracking techniques which can be performed in realistic processing time using a large 3D environmental map consisting of more than ten million 3D points and a 3D measured data with hundreds of thousands of 3D points by a RGB-D camera.

Owing to an outstanding feature of the ND voxel representation, that is, the ND voxel is able to express a point distribution in a compact but information-rich form, the proposed technique is able to evaluate the likelihood of the estimated position efficiently with a small number of voxels comparing point-clouds.

By combining the ND voxel representation and the particle filter, the proposed technique is capable of performing global localization and tracking in a large-scale 3D environmental map and 3D range data in realistic processing time.

Future works include the development of more efficient representation of point-clouds than the NDT, real-time pro-

cessing by implementing in a hardware, and comparison of accuracy and computational time with state-of-the-art ICP algorithms.

REFERENCES

- [1] Guilherme N. DeSouza and Avinash C. Kak. Vision for mobile robot navigation: A survey. *IEEE Trans. on Pattern analysis and machine intelligence*, 24(2):237–267, 2002.
- [2] Sebastian Thrun. Robotic mapping: A survey. *CMU-CS-02-111*, 2002.
- [3] Tsutomu Hasegawa, Kouji Murakami, Ryo Kurazume, Yosuke Senta, Yoshihiko Kimuro, and Takafumi Ienaga. Robot town project: Sensory data management and interaction with robot of intelligent environment for daily life. In *2007 International Conference on Ubiquitous Robots and Ambient Intelligence (URAI07)*, 2007.
- [4] Jean-Arcady Meyer and David Filliat. Map-based navigation in mobile robots: Ii. a review of map-learning and path-planning strategies. *Cognitive Systems Research*, 4(4):283 – 317, 2003.
- [5] Andreas Nüchter, Hartmut Surmann, Kai Lingemann, Joachim Hertzberg, and Sebastian Thrun. 6d slam with an application in autonomous mine mapping. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, pages 1998–2003, 2004.
- [6] Andreas Nüchter, Kai Lingemann, Joachim Hertzberg, and Hartmut Surmann. 6d slam - 3d mapping outdoor environments. *J. Field Robot.*, 24(8-9):699–722, August 2007.
- [7] Ryo Kurazume, Yusuke Noda, Yukihiro Tobata, Kai Lingemann, Yumi Iwashita, and Tsutomu Hasegawa. Laser-based geometrical modeling using cooperative multiple mobile robots. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 3200–3205, 2009.
- [8] Yukihiro Tobata, Ryo Kurazume, Yusuke Noda, Kai Lingemann, Yumi Iwashita, and Tsutomu Hasegawa. Laser-based geometrical modeling of large-scale architectural structures using co-operative multiple robots. *Autonomous Robot*, 32(1):46–62, 2011.
- [9] Satoshi Kagami, Ryo Hanai, Naotaka Hatao, and Masayuki Inaba. Outdoor 3d map generation based on planar feature for autonomous vehicle navigation in urban environment. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1526–1531, 2010.
- [10] Taro Suzuki, Mitsunori Kitamura, Yoshiharu Amano, and Takumi Hashizume. 6-dof localization for a mobile robot using outdoor 3d voxel maps. In *Proceedings of the IEEE/RSJ International conference on Intelligent Robots and Systems*, pages 5737–5743, 2010.
- [11] Clark F. Olson and Larry H. Matthies. Maximum likelihood rover localization by matching range maps. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 272–277, 1998.
- [12] Dirk Hähnel and Wolfram Burgard. Probabilistic matching for 3d scan registration. In *Proc. of the VDI-Conference Robotik 2002 (Robotik)*, 2002.
- [13] Peter Biber and Wolfgang Straber. The normal distributions transform: a new approach to laser scanmatching. In *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2743–2748, 2003.
- [14] Frank Dellaert, Dieter Fox, Wolfram Burgard, and Sebastian Thrun. Monte carlo localization for mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 2, pages 1322 –1328, 1999.
- [15] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, 2005.
- [16] Kurt Konolige and Ken Chou. Markov localization using correlation. In *Proceedings of the International Joint Conference on AI (IJCAI)*, 1999.
- [17] Jan Wülfing, Joachim Hertzberg, Kai Lingemann, Andreas Nüchter, Thomas Wiemann, and Stefan Stiene. Towards real time robot 6d localization in a polygonal indoor map based on 3d tof camera data. In *Proceedings of the 5th IFAC Symposium on Intelligent Autonomous Vehicles (IAV '10)*, 2010.
- [18] Maurice F. Fallon, Hordur Johannsson, and John J. Leonard. Efficient scene simulation for robust monte carlo localization using an rgb-d camera. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1663–1670, 2012.

- [19] Joydeep Biswas and Manuela Veloso. Depth camera based indoor mobile robot localization and navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1697–1702, 2012.
- [20] Julian Ryde and Huosheng Hu. 3d mapping with multi-resolution occupied voxel lists. *Autonomous Robots*, 28(2):169–185, 2010.
- [21] Martin Magnusson, Andreas Nuchter, Christopher Lorren, Achim J. Lilienthal, and Joachim Hertzberg. Evaluation of 3d registration reliability and speed a comparison of icp and ndt. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3907–3912, 2009.
- [22] Martin Magnusson, Achim Lilienthal, and Tom Duckett. Scan registration for autonomous mining vehicles using 3d-ndt. *Journal of Field Robotics*, 24(10):803–827, 2007.
- [23] Martin Magnusson, Henrik Andreasson, Andreas Nuchter, and Achim J. Lilienthal. Automatic appearance-based loop detection from 3D laser data using the normal distributions transform. *Journal of Field Robotics*, 26(11–12):892–914, November 2009.
- [24] Dieter Fox. Adapting the sample size in particle filters through kld-sampling. *International Journal of Robotics Research*, 22(12):985–1004, 2003.
- [25] Bruno Siciliano and Oussama Khatib. *Handbook of Robotics*. Springer, 2008.
- [26] Point cloud library. <http://pointclouds.org/>.

$(0, \sqrt{-2\log_e r}, 0)$, $(0, -\sqrt{-2\log_e r}, 0)$, $(0, 0, \sqrt{-2\log_e r})$, and $(0, 0, -\sqrt{-2\log_e r})$, their corresponding points on an ellipse called sigma points are obtained by the above transformation.

APPENDIX

A sigma point is a point on a surface of an ellipse representing a multi-dimensional normal distribution. If 3D points are distributed according to a 3D normal distribution, the existence probability of a point in a position x_s is given by

$$P(x_s) = (2\pi \det(\Sigma))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_s - \mu)^T \Sigma^{-1}(x_s - \mu)\right) \quad (15)$$

Where μ is a center of a point-cloud and Σ is a covariance matrix. The position x in which the existence probability is r ($0 < r \leq 1$) times smaller than the probability at the center position μ is given as

$$\exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) = r \quad (16)$$

$$(x - \mu)^T \Sigma^{-1}(x - \mu) = -2\log_e r \quad (17)$$

Suppose a point p on a surface of a sphere with a radius of $\sqrt{-2\log_e r}$. The equation above can be expressed as

$$(x - \mu)^T \Sigma^{-1}(x - \mu) = p^T p \quad (18)$$

When the covariance matrix Σ is a non-singular matrix, it can be decomposed with the matrixes of the eigen vectors V and the eigen values D by the eigenvalue decomposition.

$$(x - \mu)^T (VDV^T)^{-1}(x - \mu) = p^T p \quad (19)$$

$$(x - \mu)^T (VD^{-\frac{1}{2}}V^TVD^{-\frac{1}{2}}V^T)(x - \mu) = p^T p \quad (20)$$

Therefore, since V is an orthonormal matrix, the position x can be written as

$$D^{-\frac{1}{2}}V^T(x - \mu) = p \quad (21)$$

$$x = VD^{\frac{1}{2}}p + \mu \quad (22)$$

x is a projected point on an ellipse with a covariance matrix Σ from an arbitrary point p on a sphere with a radius of $\sqrt{-2\log_e r}$. Therefore, by choosing the six surface points on a sphere as $(\sqrt{-2\log_e r}, 0, 0)$, $(-\sqrt{-2\log_e r}, 0, 0)$,