Humanoid Robot Localization in Complex Indoor Environments

Armin Hornung Kai M. Wurm Maren Bennewitz

***Abstract*— In this paper, we present a localization method for humanoid robots navigating in arbitrary complex indoor environments using only onboard sensing. Reliable and accurate localization for humanoid robots operating in such environments is a challenging task. First, humanoids typically execute motion commands rather inaccurately and odometry can be estimated only very roughly. Second, the observations of the small and lightweight sensors of most humanoids are seriously affected by noise. Third, since most humanoids walk with a swaying motion and can freely move in the environment, e.g., they are not forced to walk on flat ground only, a 6D torso pose has to be estimated. We apply Monte Carlo localization to globally determine and track a humanoid's 6D pose in a 3D world model, which may contain multiple levels connected by staircases. To achieve a robust localization while walking and climbing stairs, we integrate 2D laser range measurements as well as attitude data and information from the joint encoders. We present simulated as well as real-world experiments with our humanoid and thoroughly evaluate our approach. As the experiments illustrate, the robot is able to globally localize itself and accurately track its 6D pose over time.**

I. Introduction

In this paper, we consider the problem of humanoid robot navigation in complex indoor environments, possibly consisting of different levels connected by steps and staircases. The capability to robustly navigate in such an environment is the prerequisite for robots to fulfill high-level tasks such as delivery or home-care.

In the last few years, humanoid robots have become a popular research tool as they offer new perspectives compared to wheeled vehicles. For example, humanoids are able to access different types of terrain and to climb stairs. However, compared to wheeled robots, humanoids also have several drawbacks such as foot slippage, stability problems during walking, and limited payload capabilities. In addition, the flat world assumption is violated, i.e., humanoids usually cannot be assumed to move on a plane to which their sensors are parallel due to their walking motion. The main problems which have to be dealt with to solve reliable localization for humanoids are the following. First, there is often serious noise in the executed motion commands depending on ground friction and backlash in the joints, i.e., odometry can only be estimated very inaccurately. Second, the observations provided by the lightweight sensors which typically have to be used with humanoids are rather noisy and unreliable. As a result, accurate localization, which is considered to be mainly solved for wheeled robots, is still a challenging problem for humanoid robots.

All authors are with the Dept. of Computer Science, University of Freiburg, Germany.

This work has been supported by the German Research Foundation (DFG) under contract number SFB/TR-8.

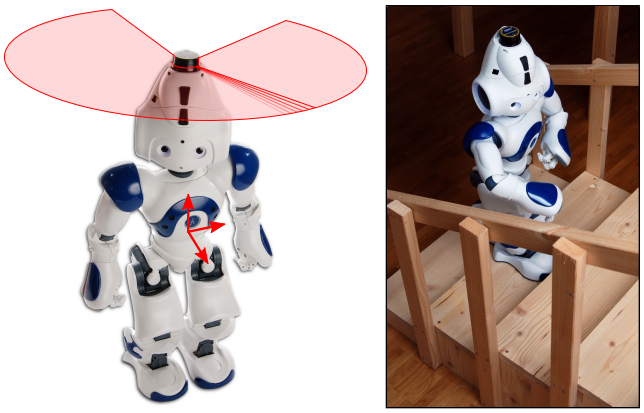

Fig. 1. Our laser-equipped *Nao* robot navigates in a multi-level indoor environment. The robot uses a head-mounted 2D laser range finder, attitude and joint angle sensors, as well as odometry information to perform Monte Carlo localization of its 6D torso pose while walking and climbing stairs.

A 2D grid map representation, which stores occupancy information of the environment discretized in cells, is in general not sufficient for navigation with humanoid robots. In complex environments containing obstacles on the floor or staircases, the capability of humanoid robots to step over or onto objects needs to be taken into account. Accordingly, 2.5D models which also represent the height of objects are often used in the humanoid robotics community (e.g., [1], [2], [3]). However, for arbitrary environments containing several levels, a 3D representation is needed which stores free and occupied areas in a volumetric way.

For reliably completing navigation tasks, a robot must be able to globally determine its pose in such a model and accurately track it over time. When operating in non-planar, multi-level environments, a robot needs to estimate a 6D state: In addition to the 2D position and yaw angle, the height of the robot's torso above the ground plane has to be estimated. As our experiments demonstrate, also the torso's roll and pitch angles are relevant since they improve localization accuracy in the 3D model.

So far, only noisy foot step odometry has been used to locally track a humanoid's pose in a 3D model (e.g., [4]). The contribution of this paper is a robust localization system for humanoid robots navigating in complex, multi-level indoor environments. We apply Monte Carlo localization (MCL) to estimate the robot's 6D torso pose in a 3D environment representation using 2D laser range data. Note that our approach does not require the robot to stop and obtain a 3D scan. As further measurements, we integrate data provided by an attitude sensor and information from the joint encoders of the robot. For our experiments, we use a modified version of

the humanoid robot *Nao* [5] which is additionally equipped with a Hokuyo laser scanner (see Fig. 1). As we show in simulated as well as in real-world experiments, the robot is able to determine its global 6D pose and accurately track it while walking and climbing stairs.

The remainder of this paper is structured as follows. We first discuss related work in the next section. Section III describes the humanoid used for our experiments, followed by a description of our 3D environment representation in Sec. IV. Our 6D localization approach is detailed in Sec. V. Finally, Sec. VI illustrates the robustness and accuracy of our localization approach in experiments.

II. Related Work

In the last few years, many approaches for tracking the pose of humanoids in the two-dimensional space have been presented. For example, Ido *et al.* [6] apply a vision-based approach and compare the current image to previously recorded reference images in order to estimate the location of the robot. Oßwald *et al.* [7] and Bennewitz *et al.* [8] compare visual features to a previously learned 2D feature map during pose tracking. Pretto *et al.* [9] track visual features over time for estimating the robot's odometry. Cupec *et al.* [10] detect objects with given shapes and colors in the local environment of the humanoid and determine its pose relative to these objects. Seara and Schmidt [11] proposed to control the gaze direction of a humanoid's stereo camera in such a way that the error of the robot's estimated foot positions is minimized. Furthermore, techniques using laser range data have also been developed. Stachniss *et al.* [12] presented an approach to learn accurate 2D grid maps of large environments with a humanoid equipped with a Hokuyo laser scanner. Such a map was subsequently used by Faber *et al.* [13] for humanoid localization in 2D. Similarly, Tellez *et al.* [14] developed a navigation system for such a 2D environment representation using two laser scanners located in the feet of the robot.

Since a 2D map is often not sufficient for humanoid motion planning, several methods use 2.5D grid maps which additionally store a height value for each cell. Thompson *et al.* [1] track the 6D pose of a humanoid equipped with a Hokuyo URG-04LX laser scanner in such a representation. However, they assume that the robot is only walking on flat ground, constraining height, roll, and pitch within fixed thresholds. In further approaches, authors use only odometry data to estimate the robot's pose while constructing a local 2.5D height map from 3D laser range measurements [2] or a combination of a local height map and a 3D grid from stereo data [15], [16]. To avoid problems resulting from the accumulated error, old data is discarded after a short period of time in these approaches.

Michel *et al.* [3] localize the robot with respect to a close object. The authors apply a model-based approach to track the 6D pose of a manually initialized object relative to the camera. Stasse *et al.* [17] proposed an approach to simultaneously localizing the robot and mapping the environment. The authors combine vision and motion information to estimate the pose and velocities of the camera as well as visual feature positions in 3D while the robot is walking on a small circle.

Finally, there exist navigation systems for humanoid robots which use external sensors to track the robot's pose, e.g., as proposed by Michel *et al.* [18], [19].

In contrast to all of these approaches, we present a system which is able to accurately determine the complete 6D pose of a humanoid robot in a 3D representation of a complex, multi-level environment using only on-board sensors.

Note that Kümmerle *et al.* [20] presented a localization technique for *wheeled* robots in a multi-level surface (MLS) map. MLS maps allow to store multiple levels per 2D grid cell. However, they do not provide a volumetric representation of the environment which is needed for humanoid navigation and they are not completely probabilistic.

III. The Humanoid Robot Nao

The humanoid robot *Nao* is 58 cm tall, weighs 4.8 kg and has 25 degrees of freedom [5]. In addition to the default sensors such as ultrasound and cameras, our humanoid is equipped with a Hokuyo URG-04LX laser range finder. While the measurements of this sensor are relatively noisy [21], it is small and lightweight. The 2D range finder is mounted in a modified head of the humanoid, providing a field of view of 240° (see Fig. 1).

In order to obtain measurements of its joint positions, Nao is equipped with Hall effect sensors which measure the angle of each joint. Using the joints of the support leg, an estimate of the robot's torso position and orientation can be obtained through forward kinematics at any time. Additionally, an inertial measurement unit (IMU) yields an estimate about the robot's orientation. Measurements from a two-axis gyroscope and a three-axis accelerometer are integrated in order to obtain an estimate of the robot's torso orientation around the world's x and y-axis (roll and pitch, respectively). The measurements of this small and lightweight IMU are quite noisy compared to the IMUs often used in robotics. However, especially while walking, these values are more accurate than the roll and pitch obtained through kinematics of the measured support leg joint angles, because the robot's feet may not always precisely rest on the ground.

IV. 3D Environment Representation

Humanoid robot navigation in complex environments requires an adequate representation of the environment. In non-planar multi-level environments, a full 3D occupancy grid map is necessary since the map needs to encode both occupied and free volumes.

In our system, we employ an octree-based mapping framework that models occupied as well as free and unknown areas in the environment in a probabilistic and memory-efficient way. This enables our humanoid to use map resolutions as small as 2 cm for a complete 3D indoor map. Our map representation is available as an open-source library [22].

Note that we do not address the problem of simultaneous localization and mapping (SLAM) in this work. We assume that a volumetric 3D map of the environment has been created beforehand.

V. 6D Localization for Humanoid Robots

For humanoid localization in complex multi-level environments, we need to determine the complete six-dimensional pose $\mathbf{x} = (x, y, z, \varphi, \theta, \psi)$ of the robot. Accordingly, we estimate the 3D position (x, y, z) and the roll, pitch, and yaw angles (φ, θ, ψ) of robot's body reference frame in the global 3D map of the environment. This reference frame is located in the center of the humanoid's torso, which is also the origin of all of its kinematic chains. For estimating the robot's 6D state, we apply Monte Carlo localization [23].

A. Monte Carlo Localization (MCL)

MCL is a Bayes filtering technique which recursively estimates the posterior about the robot's pose $\mathbf{x}_t$ at time t:

$$p(\mathbf{x}_t \mid \mathbf{o}_{1:t}, u_{1:t}) = \eta \cdot \overbrace{p(\mathbf{o}_t \mid \mathbf{x}_t)}^{\text{sensor model}} \cdot \int_{\mathbf{x}_{t-1}} \underbrace{p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, u_t)}_{\text{motion model}} \cdot \underbrace{p(\mathbf{x}_{t-1} \mid \mathbf{o}_{1:t-1}, u_{1:t-1})}_{\text{recursive term}} d\mathbf{x}_{t-1}$$

Here, η is a normalization constant resulting from Bayes' rule, $u_{1:t}$ denotes the sequence of all motion commands executed by the robot up to time t, and $\mathbf{o}_{1:t}$ is the sequence of all observations. The term $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, u_t)$ is called motion model and denotes the probability that the robot ends up in state $\mathbf{x}_t$ given it executes the motion command u_t in state $\mathbf{x}_{t-1}$. The sensor model $p(\mathbf{o}_t \mid \mathbf{x}_t)$ denotes the likelihood of obtaining observation $\mathbf{o}_t$ given the robot's current pose is $\mathbf{x}_t$.

In MCL, the belief distribution over the robot's current state is approximated by a set of n weighted samples or pose hypotheses $X_t = \{\langle \mathbf{x}_t^{(1)}, w_t^{(1)} \rangle, \ldots, \langle \mathbf{x}_t^{(n)}, w_t^{(n)} \rangle\}$. Here, each $\mathbf{x}_t^{(i)}$ is one pose hypothesis and $w_t^{(i)}$ is the corresponding weight, which is proportional to the likelihood that the robot is in the corresponding state. The update of the belief, also called particle filtering, consists of the following steps:

1) **Prediction**: The current set of particles is propagated forward according to the motion model $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, u_t)$.
2) **Correction**: The importance weight of each particle is computed according to the sensor model $p(\mathbf{o}_t \mid \mathbf{x}_t)$ given the map.
3) **Resampling**: New particles for X_{t+1} are drawn with replacement from X_t proportional to the particle weights $w_t^{(i)}$. Afterwards, their weights are reset to $w_{t+1}^{(i)} = \frac{1}{n}$. This step ensures that the filter converges to pose hypotheses with high likelihoods.

The filter is initialized with a distribution of equally weighted samples around the initial pose estimate ("*tracking*"), or with a uniform distribution over all possible hypotheses ("*global localization*").

B. Motion Model

In the prediction step of MCL, a new pose is drawn for each particle according to the motion model $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, u_t)$. In the approach presented in this paper, the motion command u_t corresponds to the incremental motion of the humanoid's torso while walking, turning, or climbing stairs. It is represented as a 6D rigid body transform that can be computed as

$$u_t = T(\tilde{\mathbf{x}}_{t-1})^{-1}\, T(\tilde{\mathbf{x}}_t)\ , \tag{1}$$

where $T(\tilde{\mathbf{x}}_t)$ denotes the transform from the origin to the estimated odometry pose $\tilde{\mathbf{x}}_t$ in an arbitrary odometry coordinate frame at time t. These estimated odometry poses originate from forward kinematics of the measured leg joint angles, as described in Sec. III.

To account for motion noise, the particle prediction step adds multivariate Gaussian noise to the motion command for each particle i:

$$\mathbf{x}_t^{(i)} = T\left(\mathbf{x}_{t-1}^{(i)}\right)\ u_t\ T\left(\mathcal{N}\left(\mathbf{0}, \delta \cdot \mathbf{\Sigma}\right)\right)\ , \tag{2}$$

where the scalar δ corresponds to the length of the translational part of u_t and $\mathbf{\Sigma} \in \mathbb{R}^{6\times 6}$ is the covariance of the motion noise. Thus, we scale the motion noise so that longer incremental torso trajectories imply higher motion noise. Note that the torso also covers a distance while turning because the humanoid constantly shifts its center of mass from one foot to the other.

In practice, odometry and other sensor data do not arrive at discrete timesteps but asynchronously and with different update rates. To solve this problem and achieve time synchronization, we update the MCL filter based on laser sensor data, interpolating odometry and IMU sensor data between two valid measurements. A second problem stems from the fact that a full laser scan is not generated instantaneously but over a certain amount of time in which the robot may be moving. To overcome this problem, we apply temporal uncertainty sampling as introduced by Thompson *et al.* [1]. For each particle, odometry transforms are interpolated to a timestamp which is sampled uniformly around the current laser timestamp in an interval corresponding to the time needed for a complete scan.

C. Sensor Model

The belief about the humanoid's 6D state is updated based on three different sources of sensor information contained in one observation $\mathbf{o}_t$. First, the laser range measurements $\mathbf{l}_t$ provided by the Hokuyo URG-04LX are integrated. Second, we regard the height $\tilde{z}_t$ of the humanoid's torso above the current ground plane as a measurement of its joint encoders and also integrate the angles for roll $\tilde{\varphi}_t$ and pitch $\tilde{\theta}_t$ as estimated by the noisy IMU. Since all these measurements are independent, the sensor model decomposes to the product

$$\begin{aligned} p(\mathbf{o}_t \mid \mathbf{x}_t) &= p(\mathbf{l}_t, \tilde{z}_t, \tilde{\varphi}_t, \tilde{\theta}_t \mid \mathbf{x}_t) = \\ &\quad p(\mathbf{l}_t \mid \mathbf{x}_t) \cdot p(\tilde{z}_t \mid \mathbf{x}_t) \cdot p(\tilde{\varphi}_t \mid \mathbf{x}_t) \cdot p(\tilde{\theta}_t \mid \mathbf{x}_t)\ . \end{aligned} \tag{3}$$

1) Laser Measurements: To integrate laser range readings, we use the endpoint model proposed by Thrun [24]. Here, the likelihood of a single range measurement $l_{t,k}$ depends on the distance d of the corresponding hypothetical beam

endpoint to the closest obstacle represented in the map, i.e., $p(l_{t,k} \mid \mathbf{x}_t) = \phi(d, \sigma_l)$, with

$$\phi(d, \sigma) = \exp\left(-\frac{d^2}{2\sigma^2}\right). \tag{4}$$

Here, σ is the standard deviation of the sensor noise and d is a distance. Note that for a given 3D map, the distances to the closest obstacles can be precomputed for all 3D cells.

Since a laser measurement consists of K beams $l_{t,k}$, the integration of a full scan $\mathbf{l}_t$ is computed as the product of the beam likelihoods

$$p(\mathbf{l}_t \mid \mathbf{x}_t) = \prod_{k=1}^{K} p(l_{t,k} \mid \mathbf{x}_t) \tag{5}$$

with the common assumption of conditional independence between the beams.

2) Roll, Pitch, and Height Measurements: Furthermore, we need to integrate the torso height $\tilde{z}_t$ above the ground plane as computed from the values of the joint encoders and the roll $\tilde{\varphi}_t$ and pitch $\tilde{\theta}_t$ provided by the IMU. Here, we evaluate the difference between the quantities predicted according to the motion model and their measured values. Similar to the laser measurements, their likelihoods are computed using a distance function as in Eq. 4:

$$p(\tilde{z}_t \mid \mathbf{x}_t) = \phi(z_{t,\text{ground}} - \tilde{z}_t, \sigma_z) \tag{6}$$

$$p(\tilde{\varphi}_t \mid \mathbf{x}_t) = \phi(\varphi_t - \tilde{\varphi}_t, \sigma_\varphi) \tag{7}$$

$$p(\tilde{\theta}_t \mid \mathbf{x}_t) = \phi(\theta_t - \tilde{\theta}_t, \sigma_\theta)\,, \tag{8}$$

where σ_z, σ_ϕ, and σ_θ are given by the noise characteristics of the joint encoders and the IMU, and $z_{t,\text{ground}}$ is computed from the height difference between z_t and the closest ground level in the map.

Finally, the complete measurement update step of the localization can be combined to the product of Eq. (5)-(8) according to Eq. (3) by choosing appropriate weighting factors for the individual likelihoods.

D. Global Localization in Multi-level Environments

When the robot has no initial guess about its pose, it needs to estimate the pose globally. In MCL, this is done by initially distributing the pose hypotheses over possible robot poses in the whole environment, i.e., in our case also over all levels. To efficiently draw robot pose hypotheses, we sample x, y, and ψ uniformly within free areas of the map and z from the probability distribution given by Eq. (6). For each sampled 2D position (x, y), we hereby consider the initial height measurement $\tilde{z}_0$ at all represented ground levels. Similarly, roll and pitch are sampled according to Eq. (7) and (8) for an initial IMU measurement $(\tilde{\varphi}_0, \tilde{\theta}_0)$.

Obviously, global localization requires more particles than pose tracking. However, once the initial particle cloud has converged, the robot's pose can be tracked using fewer particles.

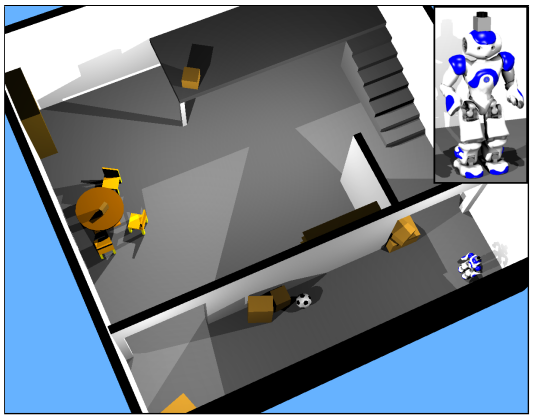

Fig. 2. The simulation environment with the humanoid at its initial position.

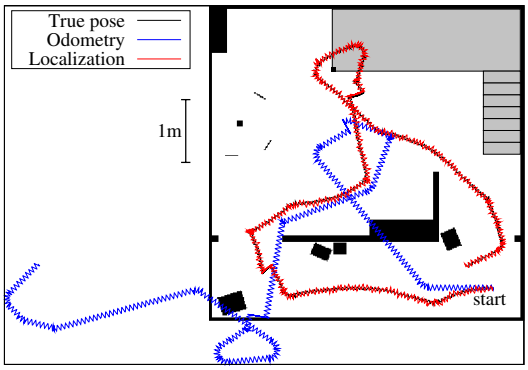

Fig. 3. The pose estimated by our localization system compared to odometry and ground truth for tracking in simulation. The 6D poses are projected on the ground plane. The particle filter accurately tracks the robot's pose, while the odometry estimate quickly diverges.

VI. Experiments

We now present localization experiments carried out with our humanoid in a real environment as well as in the *Webots* robot simulator [25]. The simulator realistically models the Nao humanoid robot, its physical interaction with the environment, and the data obtained by the sensors. The simulated environment which we designed for our experiments is shown in Fig. 2. It contains various levels, obstacles, and manipulable objects in an area of 5 m × 5 m × 2 m. The environment for the real robot consists of two levels connected by a staircase as can be seen in Fig. 5, and has a size of 6.6 m × 4.2 m × 2 m.

As the estimated localization pose of the robot, we use the weighted mean of the particle distribution. Since we only have a true pose available in the simulation, we take the 2D position of the torso as estimated by an external laser-based tracking system [26] as ground truth in our real experiments. Thus, we only have ground truth values for the x and y position of the torso with an accuracy of up to a few centimeters.

In all our experiments, Nao was teleoperated using the default behaviors for walking straight, on arcs, sideways, and for turning. For climbing stairs, we used a manually designed behavior. While walking, the robot does not move its head to ensure a stable gait. The laser sensor plane is parallel to

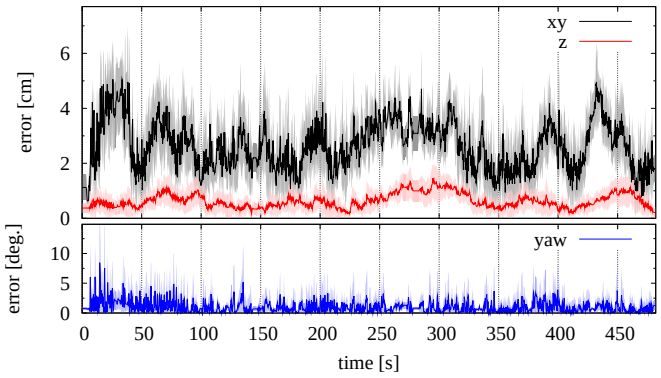

Fig. 4. Mean and standard deviation of the tracking error over 10 runs while following the trajectory depicted in Fig. 3. The errors for roll and pitch are not shown due to space constraints. See text for details.

the ground when standing still.

A. Pose Tracking

First, we evaluate the performance of the proposed localization technique for a simulated tracking experiment. Figure 3 shows the trajectory of the robot projected on the xy-plane. In this experiment, the robot was walking on the ground floor. As can be seen, foot slippage and joint backlash quickly lead to a drift in the odometry estimate. Contrary to that, our localization accurately tracks the robot's movements.

Since any Monte Carlo method is susceptible to the effects of pseudo-random number generators, we evaluate the errors as mean and standard deviation over ten differently seeded localization runs of the same recorded sensor data. The tracking error of the localization is plotted in Fig. 4. The average translational xy-error over the whole trajectory is $2.6\,\text{cm} \pm 0.8\,\text{cm}$, the average absolute yaw error is $1° \pm 0.9°$. The corresponding values are $0.3° \pm 0.3°$ for the roll and $0.3° \pm 0.2°$ for the pitch error, and $0.6\,\text{cm} \pm 0.2\,\text{cm}$ for the error of the torso height. This illustrates that our localization method is able to accurately track the 6D pose of the humanoid's torso while it is navigating in the environment.

Ignoring the roll and pitch angles during localization results in a larger error of the robot's pose: The average translational error increases to 5 cm, the yaw error to 2°, and the z error to 2 cm. This demonstrates that the humanoid's swaying motion of up to 5° in each direction needs to be considered and a full 6D pose localization is required.

Furthermore, we evaluate the pose tracking performance in a real multi-level environment which is depicted in Fig. 5. As can be seen in Fig. 6 and 7, our localization system reliably estimated the real robot's torso pose while it was walking through the environment and climbing stairs. Opposed to that, accumulated errors from foot slippage quickly lead to an erroneous odometry estimate.

Throughout all tracking experiments, $n = 500$ particles were sufficient to track the robot's pose. More particles did not lead to a significant increase in accuracy. Note that our proposed localization method efficiently runs online on a standard desktop computer, without any special optimization towards computing speed.

Fig. 5. The real-world experimental environment for the humanoid robot, consisting of two levels connected by a staircase.

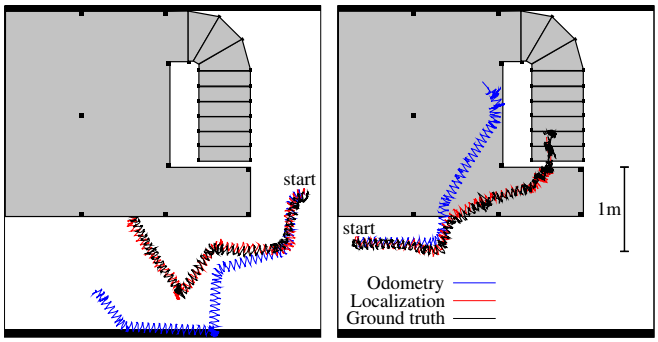

Fig. 6. Experiments carried out in our real-world environment (*Left*: ground floor only; *right*: ground floor and two stairs). Ground truth was obtained through a laser-based external tracking system. While odometry quickly diverges, our localization keeps track of the correct pose.

B. Global Localization

The plots in Fig. 8 display the evolution of $20,000$ particles during a global localization experiment in the environment shown in Fig. 2. After initialization, the particles were distributed uniformly in the free space on different levels. As can be seen, the distribution quickly converged to two clusters of hypotheses on different height levels, i.e., on the ground floor and on the top platform of the staircase. After integrating the sixth observation, the particle cloud finally converged to the true pose.

VII. Conclusions

In this paper, we proposed a probabilistic localization technique for humanoid robots using a 3D representation of arbitrary complex environments. Our system deals with all challenges occurring during humanoid robot navigation. This includes only very rough odometry information, inherently noisy sensor data, and the violation of the flat

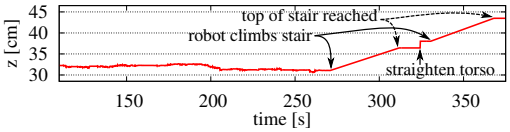

Fig. 7. The robot reliably tracks its torso height during the stair-climbing trajectory in Fig. 6 (right). Only the last part of the trajectory is shown. Note that during stair-climbing, the motion model is only applied after completing a step.

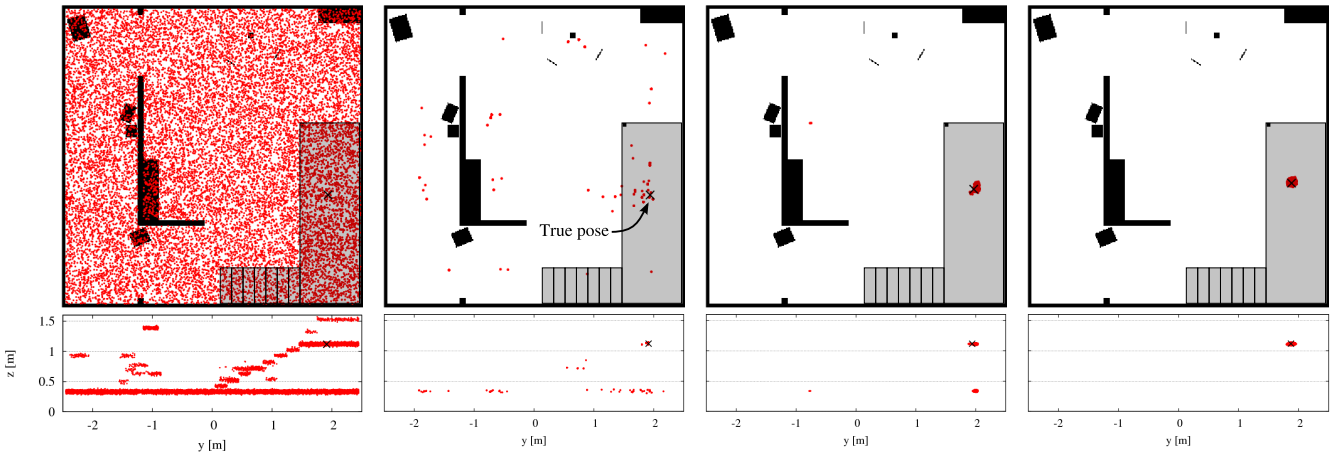

Fig. 8. Particle distribution for global localization in the simulated environment (see Fig. 2) after initialization, and after the first, third, and sixth update of the filter (left to right). The robot is standing on the top level. The top plots show a projection into the ground plane, while the bottom plots show a view from the side. Localization quickly converges to two clusters of particles on different height levels, which can be disambiguated after the sixth update.

world assumption. We apply Monte Carlo localization to globally determine and reliably track a humanoid's 6D pose, consisting of the 3D position and the three rotation angles. During localization, we integrate 2D laser range data, as well as attitude estimates and measurements from the joint encoders.

As we show in our experiments in simulation and with a real humanoid robot, our method is able to accurately estimate the 6D pose of the humanoid's torso while walking and climbing stairs. To the best of our knowledge, the presented technique is the first approach to localization of humanoids in such complex, multi-level environments.

References

[1] S. Thompson, S. Kagami, and K. Nishiwaki, "Localisation for autonomous humanoid navigation," in *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2006.

[2] J. Chestnutt, Y. Takaoka, K. Suga, K. Nishiwaki, J. Kuffner, and S. Kagami, "Biped navigation in rough environments using on-board sensing," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2009.

[3] P. Michel, J. Chestnutt, S. Kagami, K. Nishiwaki, J. Kuffner, and T. Kanade, "GPU-accelerated real-time 3D tracking for humanoid locomotion and stair climbing," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2007.

[4] J.-S. Gutmann, M. Fukuchi, and M. Fujita, "A floor and obstacle height map for 3D navigation of a humanoid robot," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2005.

[5] Aldebaran Robotics, http://www.aldebaran-robotics.com/, 2010.

[6] J. Ido, Y. Shimizu, Y. Matsumoto, and T. Ogasawara, "Indoor navigation for a humanoid robot using a view sequence," *Int. Journal of Robotics Research (IJRR)*, vol. 28, no. 2, pp. 315–325, 2009.

[7] S. Oßwald, A. Hornung, and M. Bennewitz, "Learning reliable and efficient navigation with a humanoid," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2010.

[8] M. Bennewitz, C. Stachniss, W. Burgard, and S. Behnke, "Metric localization with scale-invariant visual features using a single perspective camera," in *European Robotics Symposium 2006*, ser. STAR Springer tracts in advanced robotics, vol. 22, 2006.

[9] A. Pretto, E. Menegatti, M. Bennewitz, W. Burgard, and E. Pagello, "A visual odometry framework robust to motion blur," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2009.

[10] R. Cupec, G. Schmidt, and O. Lorch, "Experiments in vision-guided robot walking in a structured scenario," in *Proc. of the IEEE Int. Symp. on Industrial Electronics*, 2005.

[11] J. Seara and G. Schmidt, "Intelligent gaze control for vision-guided humanoid walking: methodological aspects," *Robotics & Autonomous Systems*, vol. 48, no. 4, pp. 231–248, 2004.

[12] C. Stachniss, M. Bennewitz, G. Grisetti, S. Behnke, and W. Burgard, "How to learn accurate grid maps with a humanoid," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2008.

[13] F. Faber, M. Bennewitz, C. Eppner, A. Goeroeg, A. Gonsior, D. Joho, M. Schreiber, and S. Behnke, "The humanoid museum tour guide Robotinho," in *Proc. of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2009.

[14] R. Tellez, F. Ferro, D. Mora, D. Pinyol, and D. Faconti, "Autonomous humanoid navigation using laser and odometry data," in *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2008.

[15] J.-S. Gutmann, M. Fukuchi, and M. Fujita, "3D perception and environment map generation for humanoid robot navigation," *The International Journal of Robotics Research (IJRR)*, vol. 27, no. 10, pp. 1117–1134, 2008.

[16] ——, "A modular architecture for humanoid robot navigation," in *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2005.

[17] O. Stasse, B. Verrelst, A. Davison, N. Mansard, F. Saidi, B. Vanderborght, C. Esteves, and K. Yokoi, "Integrating walking and vision to increase humanoid autonomy," *Int. Journal of Humanoid Robotics (IJHR), special issue on Cognitive Humanoid Robots*, vol. 5, no. 2, pp. 287–310, 2008.

[18] P. Michel, J. Chestnutt, J. Kuffner, and T. Kanade, "Vision-guided humanoid footstep planning for dynamic environments," in *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2005.

[19] P. Michel, J. Chestnutt, S. Kagami, K. Nishiwaki, J. Kuffner, and T. Kanade, "Online environment reconstruction for biped navigation," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006.

[20] R. Kümmerle, R. Triebel, P. Pfaff, and W. Burgard, "Monte Carlo localization in outdoor terrains using multilevel surface maps," *Journal of Field Robotics (JFR)*, vol. 25, pp. 346–359, 2008.

[21] L. Kneip, F. Tâche, G. Caprari, and R. Siegwart, "Characterization of the compact Hokuyo URG-04LX 2D laser range scanner," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2009.

[22] K. M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems," in *Proc. of the ICRA 2010 Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation*, 2010, software available at http://octomap.sf.net/.

[23] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte Carlo localization for mobile robots," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1998.

[24] S. Thrun, "A probabilistic online mapping algorithm for teams of mobile robots," *Int. Journal of Robotics Research*, vol. 20, no. 5, pp. 335–363, 2001.

[25] O. Michel, "Webots: Professional mobile robot simulation," *Journal of Advanced Robotics Systems*, vol. 1, no. 1, pp. 39–42, 2004.

[26] K. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2008.