Improved Proposals for Highly Accurate Localization Using Range and Vision Data

Stefan Oßwald Armin Hornung Maren Bennewitz

***Abstract*— In order to successfully climb challenging staircases that consist of many steps and contain difficult parts, humanoid robots need to accurately determine their pose. In this paper, we present an approach that fuses the robot's observations from a 2D laser scanner, a monocular camera, an inertial measurement unit, and joint encoders in order to localize the robot within a given 3D model of the environment. We develop an extension to standard Monte Carlo localization (MCL) that draws particles from an improved proposal distribution to obtain highly accurate pose estimates. Furthermore, we introduce a new observation model based on chamfer matching between edges in camera images and the environment model. We thoroughly evaluate our localization approach and compare it to previous techniques in real-world experiments with a Nao humanoid. The results show that our approach significantly improves the localization accuracy and leads to a considerably more robust robot behavior. Our improved proposal in combination with chamfer matching can be generally applied to improve a range-based pose estimate by a consistent matching of lines obtained from vision.**

I. Introduction

While climbing stairs, robots require a highly accurate pose estimate to avoid dangerous falls resulting from walking against a step, bumping into a handrail, or slipping off the stair edge after climbing up a step.

Previously, we presented a Monte Carlo localization (MCL) technique that integrates data from a 2D laser scanner, an inertial measurement unit, and joint encoders to estimate a robot's 6D pose in a given 3D model of the environment [1]. This approach yields good results when the robot is walking on flat ground. However, it is not accurate enough for safe stair climbing on complex staircases where the robot frequently has to re-align itself with the next steps. Climbing stairs usually poses higher challenges on the localization system because complex whole-body motions are executed and the feet may slip on the ground. As we showed before, an extension of the system that additionally uses visual observations can improve the pose estimation and the success rate for stair climbing [2]. In this previous approach, the laser-based pose estimate was refined by locally matching edges detected in camera images. However, incorrect data associations can lead to inconsistencies and thus to wrong pose estimates that still result in falls of the robot on a staircase.

In this paper, we present a new localization technique that uses an improved proposal distribution in the particle filter of MCL. In our new system, visual observations are directly integrated in the observation model of the particle filter in addition to range measurements. To this end, we developed a consistent observation model based on chamfer matching between a given edge model of the environment and a set of observed edges in camera images. By calculating an improved proposal based on range and vision observations, the particle filter generates highly focused particle sets that lead to an accurate pose estimate. This approach is illustrated in Fig. 1. Our approach can be applied in a similar way to augment stereo or RGBD-data (e.g., from the Kinect) in close ranges where these sensors cannot provide depth data due to the stereo baseline.

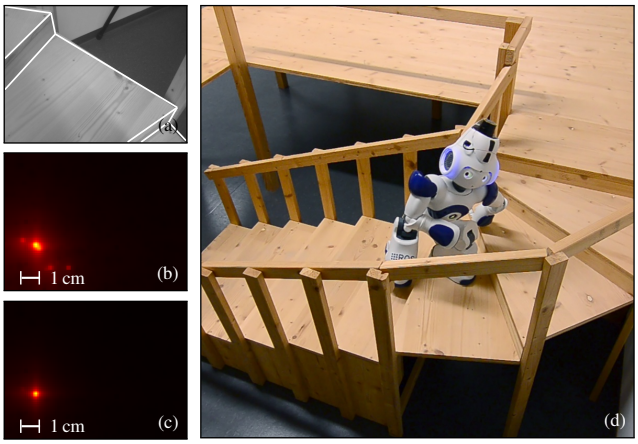


Fig. 1. Our approach uses chamfer matching between a known edge model of the environment and edges in camera images (a) to obtain the observation likelihood of the image ((b), projected into the xy-plane). Fusing this information with laser data and proprioception yields a peaked improved proposal distribution for Monte Carlo localization (c). This results in a highly accurate pose estimate, which enables a humanoid to reliably climb challenging staircases (d).

We provide experimental results obtained with a Nao humanoid robot equipped with a 2D laser range finder and a monocular camera on a challenging staircase. As the experiments show, our new approach significantly outperforms the previous techniques and leads to a highly robust navigation behavior.

II. Related Work

Previous approaches for stair climbing with humanoid robots used specialized hardware for the stair climbing task, e.g., 6D force/torque sensors in the robot's feet that allow to estimate the ground reaction forces ([3], [4]), flexible toe joints [5], impact damping mechanisms [4], or force sensors protruding from the bottom of the robot's feet [6]. In contrast to those approaches, we use a standard, low-cost humanoid

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

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

robot platform.

For sensing the staircase, Chestnutt *et al.* [6] attached a nodding laser scanner to the robot's hip that provides 3D range data of the steps ahead of the robot. Gutmann *et al.* [7] used stereo vision to reconstruct the next steps of a staircase and climb them. Both approaches incrementally build a local map of the environment containing staircases of only few steps. In contrast, our technique determines a highly accurate pose estimate in a given model of the environment where also complex staircases are present.

If the steps are small in relation to the size of the robot, the robot can plan a sequence of footsteps and use a parameterized walking controller to follow the planned footsteps [6] or use a walking pattern generator based on the zero moment point to generate a motion sequence [8]. In our environment, the step size proportionally relates to steps designed for humans and are 7 cm high. This is rather challenging to climb for the humanoid robot, so it has to execute full-body motions learned by kinesthetic teaching [2].

Cupec *et al.* [4] proposed a vision-based approach to step upon obstacles. The obstacle detection technique relies on the assumption that obstacles and floor are clearly distinguishable to simplify edge detection. The authors presented experiments in which their robot climbed two steps.

Michel *et al.* [9] track objects with the robot's monocular camera by matching detected lines to a given 3D model of the object after a manual pose initialization. While their method can only determine the pose of the robot relative to the tracked object, our localization estimates the robot's global pose within the environment.

Our technique combines observations from laser and vision for localizing the robot globally in a known environment and enables our robot to reliably climb up challenging staircases. In contrast to our previous work that locally refines the laser-based pose estimate with edge detections [2], we now present improved proposals as an extension to Monte Carlo localization and a new vision observation model based on chamfer matching. Thus, the vision data can be consistently used directly in the particle filter, which significantly improves the robot's localization accuracy and stair climbing reliability.

III. Overview

Our approach augments standard laser-based 6D MCL, described in the next section, with the integration of visual observations from multiple images. The new observation model uses chamfer matching [10] with a given environment edge model and is introduced in Sec. V. We furthermore extend the standard MCL algorithm with an improved proposal distribution in Sec. VI so that the combination of laser and vision observations leads to a highly accurate pose estimate.

IV. Standard Laser-Based 6D MCL

The complete 6D pose $\mathbf{x} = (x, y, z, \varphi, \theta, \psi)$ consists of the 3D position (x, y, z) as well as the roll, pitch, and yaw angles (φ, θ, ψ) of the robot's body reference frame in a world frame. A humanoid's body frame is usually located in its torso. We first introduce Monte Carlo localization (MCL) based on laser range data to determine the humanoid's full 6D pose in a 3D world model [1].

MCL or *particle filtering* [11] recursively estimates the robot pose at time t with a distribution of weighted particles $p\left(\mathbf{x}_t \mid m, \mathbf{u}_{1:t-1}, \mathbf{o}_{1:t}\right)$ given an environment model m, executed motions $\mathbf{u}_{1:t-1}$, and observations $\mathbf{o}_{1:t}$. The algorithm draws a new generation of particles from the proposal distribution $\pi\left(\mathbf{x}_{1:t} \mid m, \mathbf{u}_{1:t-1}, \mathbf{o}_{1:t}\right)$ and adapts the particles' importance weights to account for the discrepancy between the proposal distribution and the target distribution:

$$w_t^{(i)} = \frac{p\left(\mathbf{x}_{1:t}^{(i)} | m, \mathbf{o}_{1:t}, \mathbf{u}_{1:t-1}\right)}{\pi\left(\mathbf{x}_{1:t}^{(i)} | m, \mathbf{o}_{1:t}, \mathbf{u}_{1:t-1}\right)} \tag{1}$$

The target distribution can then be reconstructed as

$$p\left(\mathbf{x}_t \mid m, \mathbf{u}_{1:t-1}, \mathbf{o}_{1:t}\right) \simeq \sum_{i=1}^{n} w_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}\left(\mathbf{x}\right), \tag{2}$$

where δ is the Dirac delta function and $\mathbf{x}_t^{(i)}$ is a single particle pose at time t.

Commonly, the odometry motion model is used as the proposal distribution, which leads to particle weights computed based on the observation model. We compute odometry from measured leg joint angles with forward kinematics and use a Gaussian motion model.

In the observation model, laser-based 6D MCL as proposed by Hornung *et al.* [1] combines the data of the sensors into one observation $\mathbf{o}_t$: the 2D laser range measurements $\mathbf{l}_t$ corresponding to a complete scan, the height $\tilde{z}_t$ of the humanoid's torso above the current ground plane as a measurement of its joint encoders, and the angles for roll $\tilde{\varphi}_t$ and pitch $\tilde{\theta}_t$ as estimated by the IMU. Since all these measurements are independent, the observation model decomposes to the product

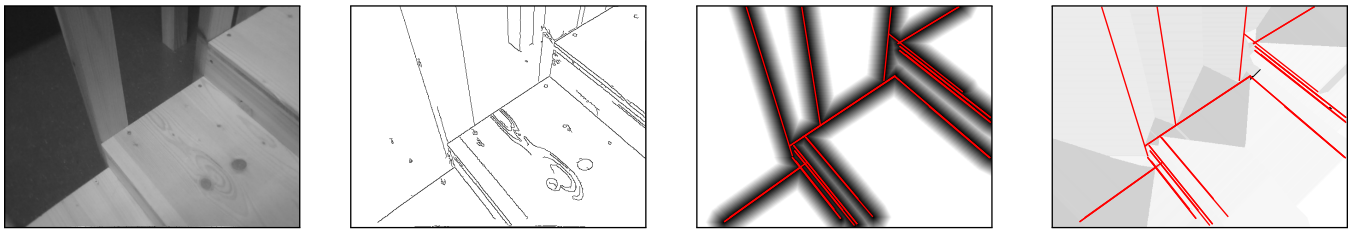
$$\begin{aligned} p(\mathbf{o}_t \mid m, \mathbf{x}_t) = p(\mathbf{l}_t, \tilde{z}_t, \tilde{\varphi}_t, \tilde{\theta}_t \mid m, \mathbf{x}_t) = \\ p(\mathbf{l}_t \mid m, \mathbf{x}_t) \cdot p(\tilde{z}_t \mid \mathbf{x}_t) \cdot p(\tilde{\varphi}_t \mid \mathbf{x}_t) \cdot p(\tilde{\theta}_t \mid \mathbf{x}_t).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

We use ray casting in a given 3D representation of the environment [12] to compute the likelihood of individual beams $p(l_{t,k} \mid m, \mathbf{x}_t)$ and apply a Gaussian noise model. Hereby, we consider the individual measurements $l_{t,k}$ to be conditionally independent and compute the product of the corresponding beam likelihoods to determine $p(\mathbf{l}_t \mid m, \mathbf{x}_t)$.

Similarly, we integrate the sensor measurements $\tilde{z}_t$, $\tilde{\varphi}_t$, and $\tilde{\theta}_t$ as Gaussian distributions based on the measured values and the predicted ones.

V. Observation Model for Vision Data

While laser-based localization is highly accurate when walking on the ground [1], translational errors can be larger while climbing a staircase due to inaccurate motion execution and slippage on the steps. This may even lead to a fall of the robot. Furthermore, the placement of the laser sensor may inhibit directly observing the area in front of the feet, which is crucial for an accurate positioning relative to the next step. Similarly, stereo and RGBD-sensors have a dead spot at close



(a) Raw camera image.

(b) Canny edge detection algorithm applied to camera image.

(c) Distance transform: Gray values represent the Euclidean distance to the nearest line.

(d) Orientation transform: Gray values represent the orientation of the nearest line.

Fig. 2. Individual steps of the chamfer matching approach: Starting with the raw camera image (a), the Canny algorithm marks edges (b) and the Hough filter extracts straight line segments shown as red lines in (c)–(d). Convolving these lines yield the distance (c) and orientation (d) transformation arrays that are used for matching edges in the observation model.

range due to their cameras' baseline. We therefore propose to augment the observation model in the particle filter with vision data in the form of detected edges, which are present in all kinds of complex environments.

The new observation model for vision data $p(C_t \mid m, \mathbf{x}_t)$ defines the likelihood of capturing the scene in a set of images $C_t = \{c_{t,1}, c_{t,2}, \dots\}$ given a 3D edge model of the environment m and the estimated pose $\mathbf{x}_t$ of the robot. We assume that the individual image observations are conditionally independent:

$$p(C_t \mid m, \mathbf{x}_t) = \prod_{c_t \in C_t} p(c_t \mid m, \mathbf{x}_t) \tag{4}$$

Our approach for estimating the observation likelihood is based on chamfer matching [10] and relies on a consistent matching of the given edge model of the staircase to lines detected in the camera images. Fig. 2 summarizes the individual steps for computing the observation likelihood based on the raw camera image shown in Fig. 2(a). First, the algorithm applies the Canny edge detection algorithm [13] and a probabilistic Hough transform [14] in order to extract line segments (Figs. 2(b) and 2(c)). Afterwards, a distance transformation is applied to the detected lines so that the value of each pixel indicates the Euclidean distance between the pixel and the nearest detected line (Fig. 2(c)). Similarly, we compute an orientation transformation that maps each pixel to the orientation of the nearest detected line (Fig. 2(d)).

To determine the observation likelihood for a given estimate of the robot's pose, the algorithm projects the edges of the given staircase model from the camera pose onto the image. By iterating over all visible pixels $\mathbf{u} = (u, v)$ of the model edges l projected onto the camera image, the algorithm computes the cost function

$$cost = \sum_{l \in m} \sum_{\mathbf{u} \in l} \left[\alpha \cdot \mathrm{dist}(\mathbf{u}, \mathrm{d}(\mathbf{u})) + \beta \cdot \angle(\mathrm{d}(\mathbf{u}), l)\right]. \tag{5}$$

Here, $\mathrm{d}(\mathbf{u})$ denotes the detected line nearest to pixel $\mathbf{u}$, $\mathrm{dist}(\mathbf{u}, \mathrm{d}(\mathbf{u}))$ denotes the Euclidean distance between the pixel $\mathbf{u}$ and the nearest detected line, $\angle(\mathrm{d}(\mathbf{u}), l)$ is the angle between the nearest detected line and the projected model line, and α, β are constant weighting factors.

We assume the observation likelihood to be distributed according to an exponential distribution over the cost function where the distribution parameter λ was determined experimentally:

$$p(c_t \mid m, \mathbf{x}_t) = \lambda \, e^{-\lambda \cdot cost} \tag{6}$$

VI. Improved Proposals for Highly Accurate Pose Tracking

During stair climbing, the odometry information gets highly unreliable and noisy, leading to a flat proposal distribution. In contrast, the observation likelihood is peaked so only a small number of particles have high weights and cover the meaningful areas of the target distribution. Hence, a large number of particles is required to sufficiently represent the posterior distribution. In order to achieve more focused particle sets, which require fewer particles to represent the posterior, we therefore use an improved proposal that takes the latest laser and vision observations into account.

A. Improved Proposal Distribution

According to Doucet *et al.* [15], the following distribution is the optimal proposal in terms of minimizing the variance of the importance weights:

$$p(\mathbf{x}_t \mid m, \mathbf{x}^{(i)}_{1:t-1}, \mathbf{o}_{1:t}, \mathbf{u}_{1:t-1}) = \frac{p(\mathbf{o}_t \mid m, \mathbf{x}_t) \cdot p(\mathbf{x}_t \mid \mathbf{x}^{(i)}_{t-1}, \mathbf{u}_{t-1})}{\underbrace{p(\mathbf{o}_t \mid m, \mathbf{x}^{(i)}_{t-1}, \mathbf{u}_{t-1})}_{=:\eta^{(i)}}} \tag{7}$$

However, computing this proposal analytically requires to evaluate the integral

$$\eta^{(i)} = \int p(\mathbf{o}_t \mid m, \mathbf{x}_t) \cdot p(\mathbf{x}_t \mid \mathbf{x}^{(i)}_{t-1}, \mathbf{u}_{t-1}) \, \mathrm{d}\mathbf{x}_t, \tag{8}$$

for which there is no closed-form solution in the general case. Grisetti *et al.* [16] introduced improved proposals in the context of grid-based SLAM (simultaneous localization and mapping) and proposed to approximate the integral in Eq. (8) as a finite sum:

$$\eta^{(i)} \simeq \sum_{j=1}^{K} p(\mathbf{o}_t \mid m, \mathbf{x}^{(i)}_j) \cdot p(\mathbf{x}^{(i)}_j \mid \mathbf{x}^{(i)}_{t-1}, \mathbf{u}_{t-1}), \tag{9}$$

where $\{\mathbf{x}^{(i)}_1, \dots, \mathbf{x}^{(i)}_K\}$ is a set of sample points drawn around the particle's pose. This sampling technique can be used to efficiently approximate the proposal if the sampled points cover meaningful regions of the proposal and these regions are sufficiently small. The proposal distribution typically

only has one mode, which we determine with scan matching of the laser scan in the 3D environment model. By sampling within a fixed radius around the computed mode, we can cover the meaningful area of the proposal distribution with a low number of samples and finally approximate the distribution by fitting a Gaussian to the weighted sample points.

B. Improved Proposals Using Laser and Vision Observations

The original algorithm [16] was used in the context of grid-based SLAM for integrating laser scans and odometry measurements, whereas we adapted the algorithm to be used for localization in known environments with additional sensors. In our case, the proposal is

$$\frac{1}{\eta^{(i)}} \cdot p(\mathbf{l}_t, C_t \mid m, \mathbf{x}_t) \cdot p(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(i)}, \mathbf{u}_{t-1}), \qquad (10)$$

and we evaluate it in points sampled in x, y, and θ direction. According to Eq. (9), this results in an approximation of $\eta^{(i)}$:

$$\begin{aligned}\eta^{(i)} &:= p(\mathbf{l}_t, C_t \mid m, \mathbf{x}_{t-1}^{(i)}, \mathbf{u}_{t-1}) \\ &\simeq \sum_{j=1}^{K} \Big[p(\mathbf{l}_t, C_t \mid m, \mathbf{x}_j^{(i)}) \cdot p(\mathbf{x}_j^{(i)} \mid \mathbf{x}_{t-1}^{(i)}, \mathbf{u}_{t-1}) \Big] \qquad (11)\end{aligned}$$

For each of the particles in the improved proposals MCL, our algorithm performs the following steps:

1) Transform the particle's pose according to the odometry information accumulated since the last integration. Starting from this initial pose estimate, a scan matcher based on gradient descent improves the particle pose by fitting the current laser scan in the given 3D environment model.

2) As the most likely pose returned by the scan matcher already provides a good estimate of the robot's pose, we assume that the meaningful regions of the proposal distribution will be in the vicinity. Hence, we sample $\{\mathbf{x}_1^{(i)}, \ldots, \mathbf{x}_K^{(i)}\}$ from a uniform distribution within a fixed range around the pose returned by the scan matcher.

3) At each sample point $\mathbf{x}_j^{(i)}$, the algorithm evaluates the observation likelihood $p(\mathbf{l}_t, C_t \mid m, \mathbf{x}_j^{(i)})$ based on the current laser scan $\mathbf{l}_t$ and the set of camera images C_t. We assume that the laser measurement and the measurements from the individual images are conditionally independent:

$$p(\mathbf{l}_t, C_t \mid m, \mathbf{x}_j^{(i)}) = p(\mathbf{l}_t \mid m, \mathbf{x}_j^{(i)}) \prod_{c_t \in C_t} p(c_t \mid m, \mathbf{x}_j^{(i)}) \qquad (12)$$

The laser observation $p(\mathbf{l}_t \mid m, \mathbf{x}_j^{(i)})$ is computed according to the raycasting model (Sec. IV) and the vision observations $p(c_t \mid m, \mathbf{x}_j^{(i)})$ according to chamfer matching (Sec. V).

4) As the algorithm can evaluate the proposal distribution only at sample points, it has to fit a continuous distribution to the sample values. Thus, a multivariate Gaussian distribution is fitted to the proposal distribution, from which it then draws the new particle poses in the resampling step.

5) The importance weights of the new particles are computed as

$$w_t^{(i)} \propto w_{t-1}^{(i)} \cdot \eta^{(i)} \cdot p(\tilde{\varphi}_t, \tilde{\theta}_t, \tilde{h}_t \mid m, \mathbf{x}_t^{(i)}), \qquad (13)$$

where $\tilde{\varphi}$ and $\tilde{\theta}$ are the current roll and pitch angles from the inertial measurement unit in the robot's chest, $\tilde{h}$ is the torso height estimated from the joint encoder values, and $\eta^{(i)}$ is the sum defined in Eq. (11) for particle i.

VII. Experiments

We use the standard laser-based MCL approach presented in Sec. IV while the robot navigates on one level within the environment. When the robot climbs a step, highly precise localization is required and we also integrate vision information using the improved proposals as introduced in Secs. VI and V. In laser-based MCL, we use 200 particles (a higher number of particles did not result in significantly higher localization accuracy). For our new approach, we use 20 particles from which we draw 500 samples to compute the improved proposal.

We apply a stratified resampling technique based on k-means clustering to reduce the number of particles when switching from laser-based MCL to the new improved proposals approach. Afterwards, the algorithm restores the original number of particles by drawing ten samples from a single proposal distribution corresponding to each particle.

A. Humanoid Robot Platform

In our experiments, we use a Nao humanoid equipped with a Hokuyo URG-04LX laser range finder. The laser is mounted on top of the head and provides 2D range data within a field of view of 240°. By tilting its head, the robot can obtain 3D range data to build a model of its environment [17]. However, due to hardware constraints, the robot cannot observe the area directly in front of its feet, which is essential for accurate and reliable stair climbing. To this end, we use data of the Nao's lower monocular camera, which points 40° downwards. While standing on a step, the robot therefore turns its head and captures three images from different camera poses to cover the whole next step and integrates the line observations in addition to laser range measurements into the particle filter as described before.

We performed our experiments on a spiral staircase of ten steps (see Fig. 1), each step has a height of 7 cm. We apply kinesthetic teaching to learn a single stair-climbing motion that enables the humanoid to climb a step [2]. To avoid falling, e.g., due to bumping into the next step or the handrail of the staircase, the robot needs to accurately position itself in front of each step.

B. Qualitative Localization Result with Improved Proposals

As an illustrating example, Fig. 3 shows the robot's view of the staircase when standing on the second step and looking to the left, center, and right with its bottom-facing camera. Fig. 4 shows the resulting observation likelihoods in the horizontal plane for these camera observations and Fig. 5 (left) shows the likelihood of the corresponding laser observation. The coordinate system is identical for all distributions and is centered at the odometry pose. The distributions of the left and right camera image are both focused and have similar modes while the distribution of the center

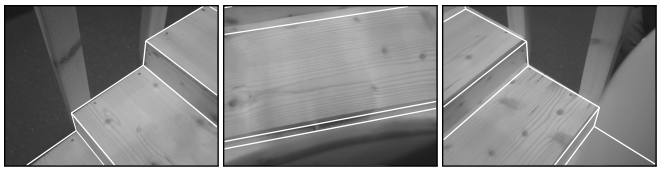


Fig. 3. Camera view of the robot when looking to the left, center, and right. Overlayed in red is the projected stair model from the best particle pose after integrating all observations. The localization is highly accurate.

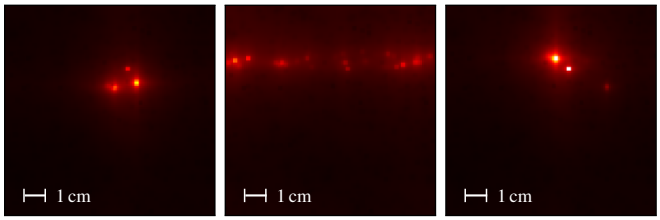


Fig. 4. Observation likelihood $p(c_t \mid m, \mathbf{x}_j)$ for left, center, and right camera image in the horizontal plane as returned by chamfer matching. Brighter areas correspond to a higher likelihood.

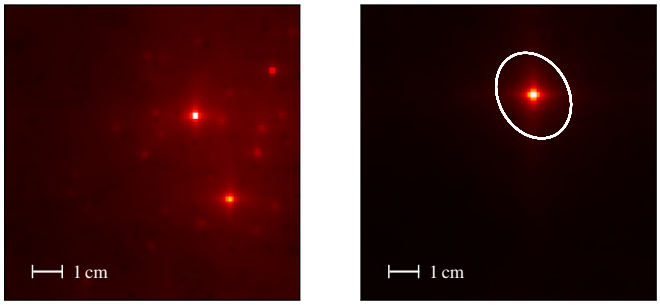


Fig. 5. *Left*: Observation likelihood $p(\mathbf{l}_t \mid m, \mathbf{x}_j)$ of the laser data. *Right*: Final improved proposal distribution by combining laser and vision observations (Fig. 4) with the 95% confidence ellipse of the Gaussian approximation (white). Brighter areas correspond to a higher likelihood.

camera image is spread in horizontal direction due to high uncertainty in lateral direction. The final improved proposal and fitted Gaussian distribution resulting from combining the distributions for vision and laser is shown in Fig. 5 (right). As can be seen, the pose estimate is highly focused. Fig. 3 shows the edge model projected from the best particle's pose after localizing on this step. The projected lines on the left and right camera image closely fit the corresponding edges of the staircase; the errors in lateral direction and in the orientation are only small. The error in forward direction is slightly higher, which is caused by the strong lines in the wood texture parallel to the stair edge.

C. Comparison to Previous Techniques

We quantitatively compared our new approach with standard laser-based MCL [1] and our previous system [2]. In this previous approach, line segments are matched to model edges in a local area around the pose estimate from the laser-based MCL. From the intersections of the matched edges the front face of the next step is computed, and from this the relative pose of the robot to the step can be inferred. In contrast to that, our new approach with improved proposals fits the edge model as a whole to the observed edges and integrates the observation back into the particle filter. In this way, the improved proposals approach avoids data association failures that can lead to inconsistent edge matchings and a degradation of the pose estimation.

TABLE I

MEAN AND STANDARD DEVIATION OF THE ERROR BETWEEN ESTIMATED POSE AND GROUND TRUTH AFTER LOCALIZING ON EACH STEP.

Error	Laser only	Local edge matching	Impr. proposals w. chamfer
Translation [mm]	25.6 ± 15.5	13.3 ± 8.8	10.9 ± 5.6
Yaw [°]	1.3 ± 1.1	1.4 ± 1.9	0.6 ± 0.6

For the quantitative evaluation of the localization algorithms, we used a *MotionAnalysis Raptor-E* motion capture system to provide ground truth data for six runs carried out with each technique on a spiral staircase with ten steps. In each run, the robot climbs the staircase until it either falls or reaches the top level.

As a performance measure for the quality of the localization, we evaluate the pose returned by each of the three localization approaches and compare it to the ground truth on each single step. In Table I, we evaluate the mean and standard deviation of the translational error in the plane as well as the absolute yaw error. The other values (roll, pitch, height above ground level) were omitted for clarity since they are directly observable from the sensors in all three algorithms. The results are aggregated over the number of successful steps the robot climbed in six runs ($N = 23$ for laser-only localization, $N = 60$ for the other two approaches).

For the standard laser-based localization, the mean translational error in the horizontal plane is 25.6 mm, which is clearly outside the tolerance range for safe stair climbing.

In comparison to that, refining the pose estimate by locally matching edges in the camera images decreases the mean translational error to 13.3 mm. This difference is statistically highly significant (t-test with 99.9% confidence). However, the standard deviation of the angular error is slightly higher because of inconsistently estimated models of the next step. The data association between detected edges and model edges is done individually for each camera image, so inconsistent assignments cannot be detected. When inconsistently assigned edges are projected back into 3D space, the corners of the front face of the step calculated by intersections of edges are incorrect and lead to errors both in the orientation and in the translation within the horizontal plane.

Our approach, Monte Carlo localization with improved proposals and consistent edge matching based on chamfer matching, outperforms the other algorithms. The mean error in the orientation of the robot is 0.6°, so the robot is able to align itself accurately with the next step. The difference of the angular error is highly significant (t-test, 99.9%). The improved proposals approach also decreases the mean translational error from 13.3 mm to 10.9 mm, which is statistically significant (t-test, 95%). Note that the accuracy is in many cases higher than the map resolution (1 cm) used for laser-based MCL, which additionally demonstrates the advantages of integrating vision information.

D. Success Rates

We now compare the success rate of climbing the complete staircase with ten steps when relying on the different localization approaches.

Using only laser-based MCL, the robot succeeded in climbing all ten steps of the staircase in only one out of six runs and fell five times. During these six experiments, the robot successfully climbed 23 steps in total. This corresponds to a success rate of 82% for climbing a single step.

Using our previous approach of matching detected edges in the camera images, the robot successfully climbed the complete staircase in seven out of ten runs. In these runs, successfully climbing 86 out of 89 single steps results in a success rate of 97%. The robot also touched handrails and bumped into the next step several times, in some cases leading to critical situations where the robot almost fell.

In contrast to that, our new approach with improved proposals leads to more accurate localization results and less interference with the environment. The robot always positioned itself precisely in front of the next step and reliably executed the stair climbing motion. Over all, the robot had a success rate of 100%, i.e., it successfully climbed the whole staircase in ten out of ten runs. A video in which the robot climbs the spiral staircase using our approach is available online at `http://hrl.informatik.uni-freiburg.de`.

E. Robustness of the Approach

Our approach is robust against unmodeled objects and texture on the surfaces, as the vision observation model only tests if the predicted edges are present in the camera images. The robot can detect self-occlusions and exclude the affected regions from the observation model. However, the localization performance may degrade if other unmodeled objects occlude significant parts of the predicted model edges, or if the texture contains strong lines running parallel to the modeled edges (such as the horizontal lines in the wood texture in Fig. 2).

The visual observations originate from the edges of the next step. Integrating this information enables the robot to align accurately with the next step. Our approach can compensate for small errors in the model of the staircase and still leads to reliable stair climbing in these cases. However, the localization accuracy clearly depends on a sufficiently accurate model of the environment.

VIII. Conclusions

In this paper, we presented a new approach for highly accurate Monte Carlo localization based on informed proposals. Our approach uses 2D laser data and monocular vision to estimate the robot's 6D pose in a given 3D model of the environment. We developed a new observation model based on globally consistent chamfer matching between an edge model of the 3D world and lines detected in camera images. We evaluated our localization technique on a Nao humanoid in a two-level environment connected by a spiral staircase with ten steps. The experiments show that our approach leads to a highly accurate localization of the robot and, thus, to robust navigation capabilities. We furthermore carried out comparative experiments that demonstrated the superior performance of our localization in comparison to previous techniques. Our method based on improved proposals and chamfer matching can be generally applied to robot localization whenever lines detected in images can be used to refine an initial pose estimate. For example, current RGBD-sensors such as the Microsoft Kinect or Asus Xtion yield no depth data at close ranges due to the stereo baseline. Localization with these sensors could be enhanced with our approach.

References

[1] A. Hornung, K. M. Wurm, and M. Bennewitz, "Humanoid robot localization in complex indoor environments," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2010.

[2] S. Oßwald, A. Görög, A. Hornung, and M. Bennewitz, "Autonomous climbing of spiral staircases with humanoids," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2011.

[3] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, "The development of Honda humanoid robot," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1998.

[4] 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.

[5] K. Nishiwaki, S. Kagami, Y. Kuniyoshi, M. Inaba, and H. Inoue, "Toe joints that enhance bipedal and fullbody motion of humanoid robots," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2002.

[6] 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.

[7] J.-S. Gutmann, M. Fukuchi, and M. Fujita, "Stair climbing for humanoid robots using stereo vision," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2004.

[8] H. Hirukawa, S. Kajita, F. Kanehiro, K. Kaneko, and T. Isozumi, "The human-size humanoid robot that can walk, lie down and get up," *The International Journal of Robotics Research*, vol. 24, no. 9, pp. 755–769, 2005.

[9] 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.

[10] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: two new techniques for image matching," in *Proc. of the Int. Conf. on Artificial Intelligence (IJCAI)*, 1977.

[11] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT-Press, 2005.

[12] 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/.

[13] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.

[14] J. Matas, C. Galambos, and J. Kittler, "Robust detection of lines using the progressive probabilistic Hough transform," *Computer Vision and Image Understanding*, vol. 78, pp. 119–137, Apr. 2000.

[15] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[16] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with Rao-Blackwellized particle filters," *IEEE Transactions on Robotics*, vol. 23, no. 1, pp. 34–46, 2007.

[17] S. Oßwald, J.-S. Gutmann, A. Hornung, and M. Bennewitz, "From 3D point clouds to climbing stairs: A comparison of plane segmentation approaches for humanoids," in *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2011.