

# Orthogonal Wall Correction for Visual Motion Estimation

Jörg Stückler and Sven Behnke

**Abstract**—A good motion model is a prerequisite for many approaches to simultaneous localization and mapping. Without an absolute reference, it is however difficult to prevent drift when estimating motion. To prevent orientation drift, our approach exploits typical features of indoor environments: Straight walls that are parallel or orthogonal to each other.

Our idea is to detect walls in monocular depth measurements and to correct odometry obtained from matching successive images and from inertial measurements, such that the observed walls are aligned with the main orientation estimated from the map that is being built.

The experimental results indicate that orientation drift can be prevented and orientation uncertainty can be reduced greatly when applying the proposed orthogonal wall correction. This can make the difference between reliable mapping and failure.

## I. INTRODUCTION

A good robot motion model is a prerequisite for many approaches to simultaneous localization and mapping (SLAM) [1], [2]. While for wheeled robots, odometry measurements can be obtained easily from wheel encoders, tracked and legged robots frequently lack good odometry. Moreover, even for wheeled robots, the odometry measurements for the robot heading are less reliable than for the robot displacement. When building maps, these motion measurement errors accumulate and may lead to significant pose errors that prevent the closing of loops.

While for unstructured environments, the closing of loops is the only chance to correct such orientation errors, typical indoor environments contain more structure that can be exploited for the continuous correction of orientation errors. Our idea is to detect walls in monocular depth measurements and to correct odometry, obtained from matching successive images and from an inertial measurement unit, such that the observed walls are aligned with the main orientation estimated from the map that is being built.

The key assumptions of our approach are that a main orientation of the walls exists in the environment and that it can be extracted reliably both from the map and from the current depth measurements. As our experiments show, these assumptions are realistic for typical indoor environments that follow common architectural rules. The proposed orthogonal wall correction can be applied to normally distributed pose estimates that are commonly used in EKF-based [3] approaches to localization and SLAM. Robot motion in particle filters [4] is also frequently approximated using Gaussians, resulting in Gaussian pose estimates to sample from [5].

This work was supported by German Research Foundation (DFG) under grant BE 2556/2-2

The authors are with the Computer Science Institute, Albert-Ludwigs-University of Freiburg, 79110 Freiburg, Germany, {stueckle | behnke}@informatik.uni-freiburg.de

The orientations of the poses are corrected towards the registration between visually measured main orientation and the building main orientation.

The remainder of the paper is organized as follows. After reviewing some of the related work, we describe in Sec. III the measurement of depth with monocular images. The estimation of robot motion from an inertial measurement unit and from consecutive depth measurements is detailed in Sec. IV. Sec. V covers our main contribution: The correction of orientation estimates based on matching the main orientation of the current measurement to the main orientation of the building. We evaluated our approach using real-world image sequences. The results reported in Sec. VI indicate that orientation uncertainty can be reduced greatly when applying the proposed orthogonal wall correction.

## II. RELATED WORK

The idea to exploit the typical orthogonal structure of indoor environments in order to simplify localization and SLAM has been investigated by some research groups in the last years. For example, a rather specific approach has been proposed by Schröter et al. [6]. They extract the main orientation from quadratic floor tiles in order to correct odometry towards the main orientation of the building. Also application-specific is the method by Jensfelt et al. [7], who applied geometric constraints to the trajectories generated by a wall-following robot. This post-processing effectively corrected orientation drift. Another approach has been proposed by Newman et al. [8]. In a post-processing step for SLAM they replace, in the generated map, almost parallel lines with parallel ones and almost orthogonal ones with orthogonal ones.

In contrast to the previous approaches, OrthoSLAM, proposed by Nguyen et al. [9], relies on a preprocessing step. The method detects line features in horizontal range scans and registers them to the main orientation of the building. This reduces SLAM to a linear 2D problem. The OrthoSLAM approach has been applied recently to 3D range scans, where orthogonal corners are detected and used as features [10]. Another 3D method is orthogonal surface assignment, proposed by Kohlhepp et al. [11]. This approach recovers the building coordinate system and matches planes to it. It is used for pose tracking in the inner loop of 3D SLAM that is based on elastic view graphs.

Common to the above approaches is that higher-level features, such as walls and corners, are matched to such features of the map being built. This is problematic in situations where an area of the environment is explored for the first time, because the features in the map in these areas are not

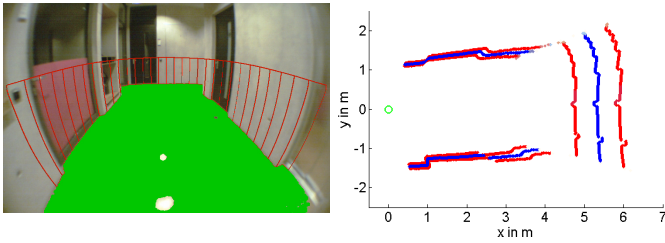


Fig. 1. Monocular depth measurement. Starting from the horizon in the wide-angle camera image (left), vertical line projections are tracked downwards until the segmented floor (green) is reached. The camera’s projective function is inverted to convert the detected wall-floor transitions into egocentric depth measurements, shown on the right in blue. The red points indicate the variance of the depth measurements.

stable yet. Our approach avoids these problems by fitting line segments to the depth measurements that are used to estimate the main orientation of the depth measurement, without adding them to the map. In our system, map construction is a separate process that relies mainly on other visual features.

### III. MONOCULAR DEPTH MEASUREMENTS

Camera-based approaches to SLAM are becoming increasingly popular as small inexpensive cameras are part of many mobile devices, and most autonomous mobile robots are equipped with cameras. Visual SLAM is particularly useful in man-made environments, as they are designed for visual navigation.

In our system, we use a wide-angle camera to capture images in frontal direction. The camera has a  $140^\circ \times 90^\circ$  field-of-view and captures WVGA color images at a rate of 10Hz. The camera moves parallel to the floor at a height of 1m. Its pose has three degrees of freedom:  $(x, y)$  position and orientation  $\theta$ . The camera is tilted in pitch-direction  $23.5^\circ$  downwards, which results in a vertical field-of-view of  $21.5^\circ$  above the horizon and of  $68.5^\circ$  below the horizon. Its projective function and radial distortion parameters are determined in advance. Fig. 1 shows a typical image captured in our test environment, the corridors of our office building.

In typical indoor environments, the floor can be easily distinguished from the walls by simple features, such as color or texture. We segment the hardwood floor in our test environment using its color in YUV space. Relaxation labeling [12] makes sure that only larger structures are segmented as floor. A connected component analysis ensures that the segmented floor is connected to the region directly in front of the robot.

Fig. 1 also shows how vertical lines are tracked downwards from the horizon to localize the wall-floor transition. The detected transitions are transformed into a camera-centric metric coordinate system by inverting the camera’s projective function. First, the intrinsic camera distortion is inverted numerically [13]. Then, angle and distance can be computed by linear projection of the detected transitions to the floor plane. Variance in the depth measurements is approximated from variance due to perturbations in the camera pitch angle, incorrect floor segmentation, and variance of depth estimates in a local neighborhood. The resulting depth measurements

are depicted on the right side of Fig. 1. An additional extraction confidence in  $[0, 1]$  for depth measurements is estimated based on floor segmentation quality and local depth variance.

## IV. MOTION ESTIMATION

In our system, we do not have control inputs or odometry sensors that could be used to estimate the camera motion. Instead, our robot is equipped with an inertial measurement unit that measures linear accelerations in the horizontal plane and turning speed around the vertical axis. Integrating linear accelerations to a velocity estimate yields a drift that we correct using visual reference measurements. We use the speed estimated from matching successive depth images as reference.

### A. Inertial Motion Estimation

An inertial measurement unit (IMU) delivers information about the attitude of the sensor and about sensor motion. Our IMU measures linear accelerations and turning rates on three orthogonal axes. As the robot moves in a horizontal plane, we can interpret turning rates around the vertical (yaw) axis and horizontal linear accelerations as motion information. When the IMU slightly deviates from the horizontal, its linear acceleration sensors measure fractions of gravity in addition to the acceleration indicating sensor motion. As accelerations are integrated to velocities, slight deviations from the horizontal would yield a significant velocity drift. To avoid this effect, after an initial calibration phase, the acceleration that is due to gravity is removed from the measured linear accelerations.

### B. Visual Motion Estimation

In addition to the inertial motion estimate, we also estimate robot motion from matching successive depth images. Let  $\mathbf{s}_t = (x_t, y_t, \theta_t)^T$  and  $\mathbf{s}_{t-1} = (x_{t-1}, y_{t-1}, \theta_{t-1})^T$  be the robot poses at times  $t$  and  $(t-1)$ , respectively. Pose  $\mathbf{s}_t$  can be obtained by translating pose  $\mathbf{s}_{t-1}$  by  $\Delta \mathbf{p}$  and then rotating by  $\Delta \theta$ , i.e.  $\theta_t = \theta_{t-1} + \Delta \theta$  and  $(x_t, y_t)^T = (x_{t-1}, y_{t-1})^T + R(\theta_{t-1})\Delta \mathbf{p}$ . Here,  $R(\theta)$  denotes a 2D rotation by  $\theta$  around the origin.

A stationary point  $Q = (q_x, q_y)$  is observed at times  $t$  and  $(t-1)$  in the depth images at camera-centric Cartesian positions  $\mathbf{q}_t$  and  $\mathbf{q}_{t-1}$ , respectively. The global position  $Q$  is then

$$Q = (x_{t-1}, y_{t-1})^T + R(\theta_{t-1})\mathbf{q}_{t-1} \text{ and} \\ Q = (x_t, y_t)^T + R(\theta_t)\mathbf{q}_t.$$

Consequently,  $\mathbf{q}_{t-1} = R(\Delta \theta)\mathbf{q}_t + \Delta \mathbf{p}$ .

We transform the depth images into Cartesian point sets  $\mathcal{Q}_{t-1} = \{\mathbf{q}_{t-1,k}\}_{k=1}^K$  and  $\mathcal{Q}_t = \{\mathbf{q}_{t,k}\}_{k=1}^K$ , where  $k$  increases from left to right. Only the points for which depth could be measured with high confidence are used. We follow the iterative closest point (ICP) approach [14] to determine Rotation  $R := R(\Delta \theta)$  and Translation  $\Delta \mathbf{p}$  that minimizes

the quadratic registration error between the two point sets:

$$E(R, \Delta \mathbf{p}) = \frac{1}{K} \sum_{k=1}^K \|R\mathbf{q}_{t,k} + \Delta \mathbf{p} - \Pi(R\mathbf{q}_{t,k} + \Delta \mathbf{p}, \mathcal{Q}_{t-1})\|^2$$

Function  $\Pi(\mathbf{q}, \mathcal{Q})$  computes a correspondence for point  $\mathbf{q}$  in the polyline described by point set  $\mathcal{Q}$ . Because of the discrete sampling of points in the depth measurements, the point sets  $\mathcal{Q}_{t-1}$  and  $\mathcal{Q}_t$  correspond to different global positions. Assuming local linearity between neighboring depth measurements, we use a point-to-line metric in function  $\Pi$  as follows. For a point  $\mathbf{q}$ , we find the closest point  $\mathbf{q}'_k$  in  $\mathcal{Q}$ . We make sure that the distance between  $\mathbf{q}$  and  $\mathbf{q}'_k$  is below a threshold. For the direct neighbors  $\mathbf{q}'_{k\pm 1}$  that are less than a maximal distance away from  $\mathbf{q}'_k$ , we project  $\mathbf{q}$  onto the line spanned by  $\mathbf{q}'_k$  and  $\mathbf{q}'_{k\pm 1}$ . If the projection  $\pi$  lies between the line-spanning points and the Euclidean distance between  $\mathbf{q}$  and its projection is below a threshold,  $\Pi$  delivers for  $\mathbf{q}$  this projection:  $\pi := \Pi(\mathbf{q}, \mathcal{Q})$ .

We denote with  $\mathcal{C} := \{\mathbf{q}_c \in \mathcal{Q}_t\}_{c=1}^C$  the set of  $C$  points  $\mathbf{q}_c \in \mathcal{Q}_t$ , for which projection  $\pi_c := \Pi(\mathbf{q}_c, \mathcal{Q}_{t-1})$  exists.  $\mathcal{P} := \{\pi_c\}_{c=1}^C$  denotes the corresponding set of projections.

To determine rotation and translation that registers the point sets  $\mathcal{C}$  and  $\mathcal{P}$ , we remove the means of the point sets:

$$\bar{\mathbf{q}}_c := \mathbf{q}_c - \underbrace{\frac{1}{C} \sum_{c=1}^C \mathbf{q}_c}_{=: \mu_q}, \quad \bar{\pi}_c := \pi_c - \underbrace{\frac{1}{C} \sum_{c=1}^C \pi_c}_{=: \mu_\pi}$$

Singular value decomposition of

$$W = \sum_{c=1}^C \bar{\mathbf{q}}_c \bar{\pi}_c^T = USV^T$$

yields the optimal rotation  $R = UV^T$  and the optimal translation  $\Delta \mathbf{p} = \mu_\pi - R\mu_q$ .

The registration procedure is iterated until the estimated transformation converges to zero or a maximal number of iterations is reached. Points in  $\mathcal{Q}_t$  are transformed with the rotation  $R$  and the translation  $\Delta \mathbf{p}$  computed in the previous iteration. The first iteration of ICP is initialized with the current velocity estimate  $\mathbf{v} = (v_x, v_y, \omega)^T$  (see Sec. IV-C).

In addition to the transformation estimate itself, we also need an estimate of its certainty in order to be able to fuse it with other measurements in a sound way. The certainty shall reflect not only the certainty of the depth measurements, but also the fact that some of the three dimensions might not be observable.

We express the certainty of the visual motion estimate as covariance for rotation  $R$  and translation  $\Delta \mathbf{p}$ . To compute the covariance, we follow the approach recently proposed by Censi [15]. Let  $\mathbf{m} := (\Delta x, \Delta y, \Delta \theta)^T$  and let  $\mathbf{n} := (\mathbf{q}_{t-1,1}, \dots, \mathbf{q}_{t-1,K}, \mathbf{q}_{t,1}, \dots, \mathbf{q}_{t,K})^T$ . The dependence of the error on  $\mathbf{m}$  is denoted  $E(\mathbf{m}) := E(R(\Delta \theta), \Delta \mathbf{p})$ . The covariance  $\Sigma_m$  is now approximated as follows:

$$\Sigma_m \approx \left( \frac{\partial^2 E}{\partial \mathbf{m}^2} \right)^{-1} \frac{\partial^2 E}{\partial \mathbf{n} \partial \mathbf{m}} \Sigma_n \frac{\partial^2 E}{\partial \mathbf{n} \partial \mathbf{m}}^T \left( \frac{\partial^2 E}{\partial \mathbf{m}^2} \right)^{-1}$$

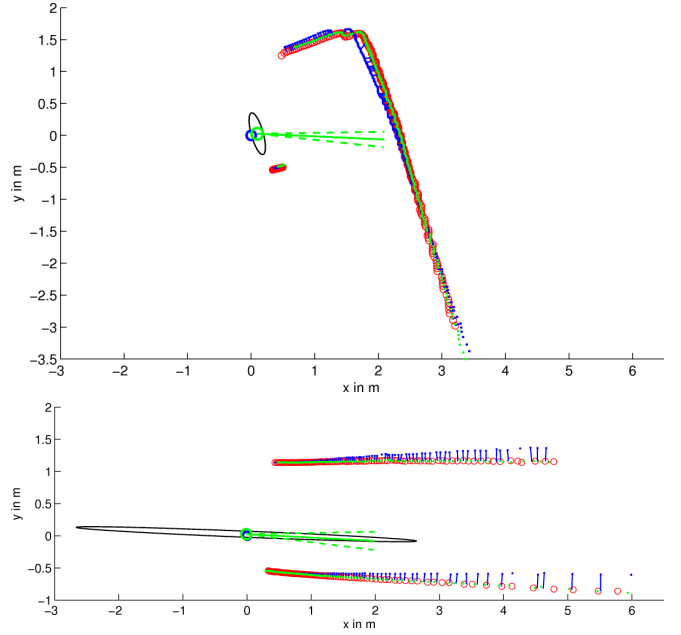


Fig. 2. Two examples of visual motion estimation. The delta between the poses of the previous time step (blue circle) and the current time step (green circle with heading indicator) corresponds to the transformation that maps depth measurements of the current time step (blue dots) onto depth measurements of the previous time step (red circles). The blue lines indicate the projections of the current measurements found in the first iteration of ICP. The final transformation is indicated by the green dots. The covariance estimate of the transformation is indicated by its  $2\sigma$  position range (black ellipses) and by its  $2\sigma$  orientation range (green hashed lines). Bottom: The elongated covariance ellipse reflects high pose uncertainty along the corridor.

The second derivatives needed above can be computed in closed form. If stochastic independence between the point observations is assumed, then  $\Sigma_n$  is diagonal, which simplifies the computations.

The output of the described method is a Gaussian estimate  $((\Delta x, \Delta y, \Delta \theta)^T, \Sigma_m)$  of the robot motion that leads from pose  $\mathbf{s}_{t-1}$  to pose  $\mathbf{s}_t$ . Fig. 2 illustrates the behavior of visual motion estimation for two typical situations. In the lower part of the figure, it is clearly visible that the covariance reflects the uncertainty of motion estimation that arises from the missing measurements orthogonal to the corridor in front of the robot.

### C. Fusion of Motion Estimates

We maintain a Gaussian velocity estimate  $(\mu_{v,t}, \Sigma_{v,t})$  that reflects both the acceleration estimated from the IMU and the velocities estimated from the camera. To do so, we call the IMU accelerations  $\mathbf{a}_t := (a_{t,x}, a_{t,y}, a_{t,\omega})^T$ , where the angular acceleration is approximated by the difference of the current and the previous turning rate  $a_{t,\omega} = \frac{(\omega_t - \omega_{t-1})}{\Delta t}$ .

The approximately linear state transition function

$$\mathbf{v}_t = \mathbf{v}_{t-1} + \Delta t \mathbf{a}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, R_t^v)$$

models the change of the velocity  $\mathbf{v}_{t-1}$  through the acceleration  $\mathbf{a}_t$  in discrete steps. The state transition is disturbed by Gaussian noise.

The visual motion estimation contributes translation  $\Delta \mathbf{p} = (\Delta x, \Delta y)^T$ , orientation change  $\Delta \theta$ , and the transformation covariance  $\Sigma_m$ . The depth image transformation is used as an observation of the velocity  $\mathbf{z}_t^v = \frac{1}{\Delta t}(\Delta x, \Delta y, \Delta \theta)^T$ . The covariance  $Q_t^v = \frac{1}{\Delta t^2} \Sigma_m$  of the velocity observation is obtained by first-order error propagation from the estimated pose deviation covariance  $\Sigma_m$ .

The linear observation model

$$\mathbf{z}_t^v = \mathbf{v}_t + \delta_t, \quad \delta_t \sim \mathcal{N}(0, Q_t^v)$$

explains  $\mathbf{z}_t^v$  as observation of the current velocity, disturbed by Gaussian noise.

A Kalman filter predicts  $(\bar{\mu}_{v,t}, \bar{\Sigma}_{v,t})$ , based on the previous velocity estimate  $(\mu_{v,t-1}, \Sigma_{v,t-1})$ :

$$\begin{aligned} \bar{\mu}_{v,t} &= \mu_{v,t-1} + \mathbf{a}_t, \\ \bar{\Sigma}_{v,t} &= \Sigma_{v,t-1} + R_t^v. \end{aligned}$$

To compute the new velocity estimate  $(\mu_{v,t}, \Sigma_{v,t})$ , the prediction is corrected with the current velocity observation  $\mathbf{z}_t^v$ :

$$\begin{aligned} K &= \bar{\Sigma}_{v,t}(\bar{\Sigma}_{v,t} + Q_t^v)^{-1}, \\ \mu_{v,t} &= \mu_{v,t-1} + K(\mathbf{z}_t^v - \mu_{v,t-1}), \\ \Sigma_{v,t} &= (I - K)\bar{\Sigma}_{v,t-1}. \end{aligned}$$

## V. ORTHOGONAL WALL CORRECTION

As both the IMU and the depth measurements lack any absolute reference in the 3D pose space, the poses obtained by integrating the estimated velocities are subject to drift. The drift in the robot orientation can be eliminated by exploiting the typical structure of indoor environments: Straight walls that are parallel or perpendicular to each other and hence are compatible with the main orientation of the building.

Our approach is to estimate both the main orientation of the walls in the current depth measurement and the main orientation of the walls in the map that is being built. This allows for continuous correction of estimated poses towards an alignment between both main orientations. Such an orthogonal wall correction greatly reduces orientation uncertainty and hence essentially reduces trajectory estimation to the two Cartesian coordinates.

### A. Estimating the Main Orientation of the Building

The main orientation of the building coordinate system is represented as angle  $\psi \in [0, \frac{\pi}{2})$ . This means that the orientations  $\psi_w$  of most wall segments should be parallel or orthogonal to  $\psi$ :  $\Psi := \{\psi - \frac{\pi}{2}, \psi, \psi + \frac{\pi}{2}, \psi + \pi\}$ .

To estimate the main orientation of a map that represents the building, we treat the orientations  $\psi_{w_{1:W}}$  of the  $W$  wall segments  $w_{1:W}$  in the map as observations of the main orientation of the building. We assume that the observations are disturbed by zero-mean Gaussian noise with variance  $\sigma_{\psi,k}^2$ , which is derived from the variance of the wall-segment defining landmarks.

The main orientation is estimated from the observations  $z_{1:W}^\psi := \psi_{w_{1:W}}$  through recursive histogram filtering [2]. Mean and variance of the main orientation are then estimated through a mean shift procedure [16]. Fig. 3 illustrates an

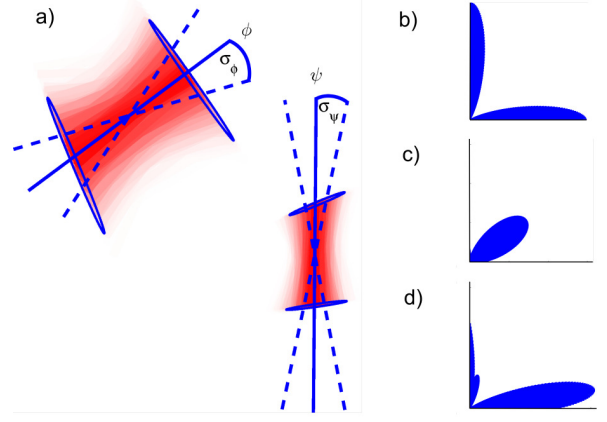


Fig. 3. Estimating the main orientation of wall segments. a) two wall segments with mean orientations  $\psi$  and  $\phi$  and standard deviations  $\sigma_\psi$  and  $\sigma_\phi$ , respectively. b,c) orientation histograms for  $\psi$  and  $\phi$ , respectively. d) orientation histogram after integrating both measurements.

example histogram for the main orientation of two wall segments.

The histogram is discretized into  $B$  bins of size  $\frac{\pi}{2B}$ . The value  $p_b$  of a bin  $b$  represents the probability that the main orientation corresponds to the angle  $\psi_b = \frac{b}{B} \cdot \frac{\pi}{2}$ . After the integration of the  $k$ -th observation, the histogram contains an estimate for the main orientation that is based on  $w_1$  to  $w_k$ , i.e.  $p_{b,k} := p(\psi = \psi_b | z_{1:k}^\psi)$  for all  $b \in \{0, \dots, B-1\}$ . The estimation of the main orientation histogram is done recursively over index  $k$  of the wall segments. For a wall segment  $w_k$ , the orientation distance to all bins  $\Delta\psi_{b,k} = \text{cyc}(\psi_b - \psi_{w_k}, -\frac{\pi}{4}, \frac{\pi}{4})$  is mapped to the interval  $[-\frac{\pi}{4}, \frac{\pi}{4})$ . The weight  $w_{b,k} = p(\psi_{w_k} | \psi = \psi_b)$  of a bin results from the probability of  $\psi_{w_k}$  under the expectation that  $\psi_b$  is the main orientation:

$$w_{b,k} = \frac{1}{\sqrt{2\pi}\sigma_{\psi,k}} \exp\left(-\frac{\Delta\psi_{b,k}^2}{2\sigma_{\psi,k}^2}\right).$$

The bin probabilities  $p_{b,k-1}$  are multiplied with the weight  $w_{b,k}$  and normalized:

$$\forall b \in \{0, \dots, B-1\} : p_{b,k} = \frac{w_{b,k} p_{b,k-1}}{\sum_b w_{b,k} p_{b,k-1}}.$$

After integrating all wall segments into the histogram, we estimate the main orientation using a mean shift procedure. A window size smaller than  $\frac{\pi}{4}$  prevents mutual extinction of direction vectors. The iterative improvement of the estimate is terminated when the change of the estimated angle falls below a threshold or a maximum number of iterations is exceeded. After the estimate of the mean orientation  $\psi$ , we compute its standard deviation  $\sigma_\psi$  in the window around  $\psi$  from the histogram.

### B. Estimating the Main Orientation of Depth Measurements

We estimate the main orientation of the wall segments in the current depth image as follows. In our application, the depth image contains 1024 measurements, which are transformed into Cartesian coordinates  $Q_t$  (see Sec. IV-B).

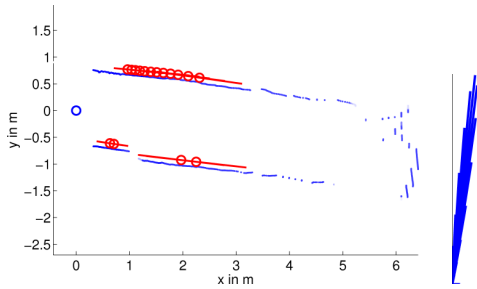


Fig. 4. Left: Extracted lines (red) from depth measurement (blue) at sampled window positions (red circles). Right: Scan orientation histogram.

A window  $W$  of length  $K = 128$  is placed at equidistant angular positions within the current depth image such that adjacent windows overlap significantly. The points  $Q_W := \{\mathbf{q}_k = (x_k, y_k)^T\}_{k=1}^K$  within the window are approximated with a line  $\gamma_W : x \cos(\alpha_W) + y \sin(\alpha_W) - r_W = 0$  in Hessian parametrization [17]. Each depth measurement is weighted with its extraction confidence.

The line fit is only computed if the average weight  $\overline{w_W}$  is above a threshold. Furthermore, successive points within  $W$  must have a small Euclidean distance. Two more conditions must be met for using the fitted line to estimate the main orientation of the depth image: The ratio between the mean squared approximation error and the line length must be below a threshold, in order to ignore fits to non-straight depth windows, and the estimated length of the line must be above a threshold (set to 0.5m in our system), in order to obtain a confident estimate of the line orientation.

In an analogous way to the estimation of the building's main orientation from the wall segments, described in Sec. V-A, we construct an orientation histogram of the line segments and apply mean shift to estimate the main orientation of the depth image  $\phi$  and its variance  $\sigma_\phi^2$ .

### C. Orientation Correction

The estimation of both, the main orientations of the building and the depth measurement, can be used to constrain possible poses. A pose estimate  $(\hat{s}_t, R_t)$  obtained, e.g., by integrating the fused velocities from Sec. IV-C, is corrected by treating the estimated main orientation in the depth image as observation of the building's main orientation  $(\psi, \sigma_\psi^2)$ :

$$z_t^\psi = h(s_t, \psi) + \delta_t = \psi - \theta_t + \delta_t, \quad \delta_t \sim \mathcal{N}(0, Q_t^\psi)$$

The pose estimate is corrected by a Kalman Filter update with innovation  $\Delta\phi := \text{cyc}(\phi - h(\hat{s}_t, \psi), -\frac{\pi}{2}, \frac{\pi}{2})$  and measurement covariance  $Q_t^\psi := \sigma_\phi^2 + \sigma_\psi^2$ , which takes the uncertainty in the estimate of the building's main orientation into account. Fig. 5 illustrates the procedure with an example. Please note that when this orientation correction is done continuously, the individual corrections will be much smaller than shown in the figure.

## VI. EXPERIMENTAL RESULTS

We evaluate our approach with data captured in the test environment illustrated in Fig. 6. The robot trajectory is

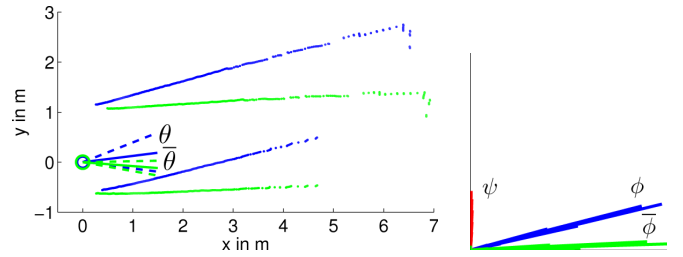


Fig. 5. Pose correction by registration of the main orientation of the depth images with the building's main orientations  $\psi$ . In the initial pose (blue) that has orientation  $\theta$ , the depth image (blue) has the global orientation  $\phi$ . After correcting the pose to the new orientation  $\bar{\theta}$  (green), the global orientation  $\bar{\phi}$  of the depth image (green) is more compatible with the building's main orientation. Dashed lines indicate the  $2\sigma$  ranges of the orientation estimates.

also shown in the figure. It is approximately 76m long and contains two loops: A larger loop around the core of the building and a smaller loop in the lower right corridor.

We applied the motion integration described above, with and without orthogonal wall correction. To correct the pose orientation, we used the main orientation estimated from a map obtained from running our visual SLAM system on the same data. Fig. 7 depicts typical mean estimated trajectories of both methods. One can clearly see that, without orientation correction, the robot orientation drifts to an extent that makes closing of loops difficult for a visual SLAM system. This effect can be prevented by the orientation correction.

We also estimated the uncertainty of the pose distribution by sampling poses from the motion model and calculating the entropy of the normal distribution fitted to the pose samples in each time step. Fig. 7 shows how the uncertainty in the orientation estimate is nearly constant when applying orthogonal wall correction. As a consequence, also the uncertainty in the Cartesian position estimate and in the 3D pose estimate grows slower than without orthogonal wall correction.

When using the orthogonal wall correction within our visual SLAM system [18], its performance improves significantly. Here, we continuously estimated the building's main orientation from the map built so far. Our system uses particles to represent the distribution over possible robot trajectories and maps [5]. The maps are described

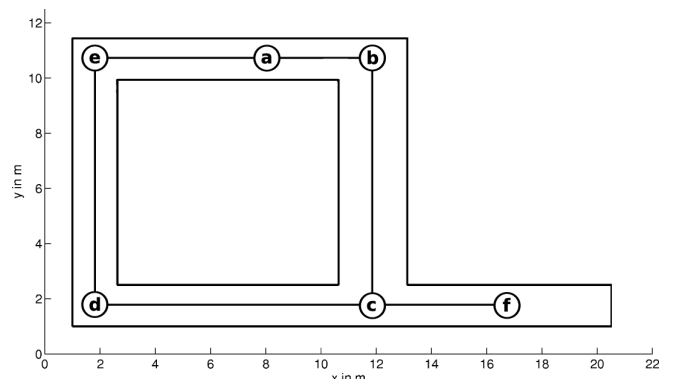


Fig. 6. Outline of the test environment. The robot starts in a and moves along the waypoints a→b→c→d→e→a→b→c→f→c→b.



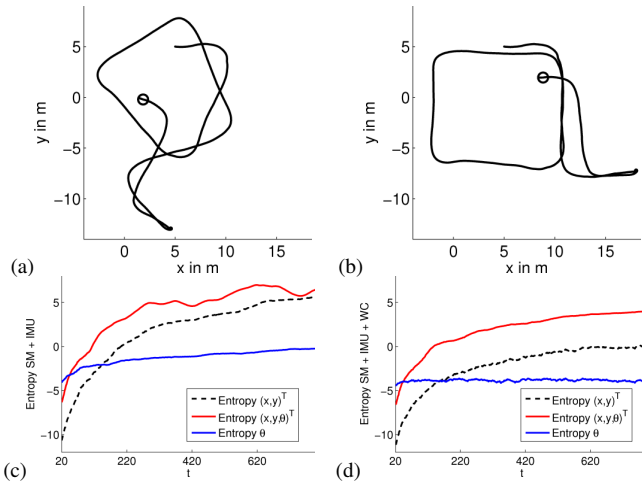


Fig. 7. Mean estimated trajectories (top) and uncertainty of estimated poses (bottom) (a)/(c) without, (b)/(d) with orthogonal wall correction.

by point landmarks (blobs and edges), connected by walls. With only 50 particles, both loops in the trajectory in Fig. 6 can be closed in nine out of ten runs when the orthogonal wall correction is applied. In contrast, only one of ten runs succeed without the proposed method. Fig. 8 shows a typical map built and the estimated robot trajectory. It can be observed that the correct trajectory could be recovered and an abstract map of our test environment has been obtained.

## VII. CONCLUSIONS

In this paper, we proposed a method to exploit the typical structure of indoor environments, walls that are either parallel or orthogonal to each other, to prevent orientation drift, a frequent problem of odometry-based motion models. We detect walls in monocular depth measurements to correct odometry, obtained from matching successive images and from inertial measurements, such that the observed walls are aligned with the main orientation estimated from the map that is being built. Our experiments show that such an orthogonal

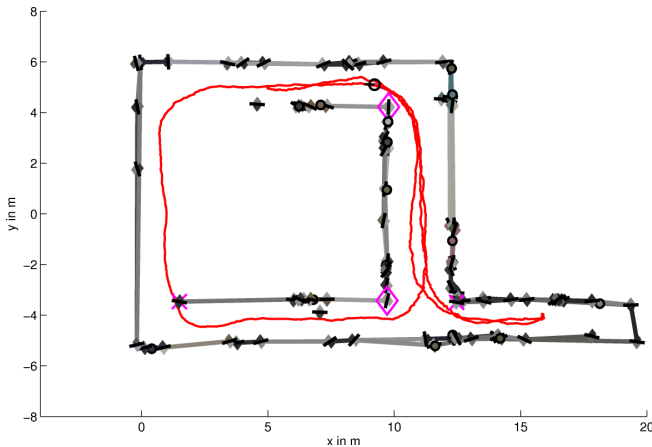


Fig. 8. Example map with trajectory constructed with our visual SLAM system, when applying orthogonal wall correction. Please note that the large loop goes on the left side around an open space that is secured by a banister, which does not fulfill our criteria for proper walls.

wall correction prevents orientation drift and significantly reduces uncertainty in the orientation estimates.

The idea of orthogonal wall correction is not restricted to visual odometry and visual depth measurements. It can be applied directly to depth measurements from a laser-range-scanner and to odometry obtained from wheel encoders. Also, the assumption that the environment contains mostly walls of the main orientation is not imperative. If a significant number of walls deviate from this main orientation, this can be detected easily and these walls can be excluded from the orientation estimation. The reduction of orientation uncertainty essentially reduces trajectory estimation to a 2D problem. This leads to better pose estimates and makes the closing of loops in SLAM easier. As our experiments showed, this can make the difference between reliable mapping and failure.

## REFERENCES

- [1] S. Thrun, "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millennium*. Morgan Kaufmann, 2003.
- [2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
- [3] P. Maybeck, "The Kalman filter: An introduction to concepts," in *Autonomous Robot Vehicles*, I. J. Cox and G. T. Wilfong, Eds. Springer, 1990, pp. 194–204.
- [4] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "Tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [5] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *Proc. of Int. Conf. on Artificial Intelligence (IJCAI)*, 2003, p. 1151–1156.
- [6] C. Schröter, H.-J. Böhme, and H.-M. Groß, "Extraction of orientation from floor structure for odometry correction in mobile robotics," in *Pattern Recognition*, ser. LNCS 2781. Springer, 2003, pp. 410–417.
- [7] P. Jensfelt, H. Christensen, and G. Zunino, "Integrated systems for mapping and localization," in *Proc. of ICRA SLAM Workshop*, 2002.
- [8] P. Newman, J. Leonard, J. D. Tardbs, and J. Neira, "Explore and return: Experimental validation of real-time concurrent mapping and localization," in *Proc. of IEEE ICRA*, 2002.
- [9] V. Nguyen, A. Harati, N. Tomatis, A. Martinelli, and R. Siegwart, "OrthoSLAM: A step toward lightweight indoor autonomous navigation," in *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006.
- [10] A. Harati and R. Siegwart, "Orthogonal 3D-SLAM for indoor environments using right angle corners," in *Proc. of 3rd European Conf. on Mobile Robots (ECMR)*, 2007.
- [11] P. Kohlhepp, G. Bretthauer, M. Walther, and R. Dillmann, "Using orthogonal surface directions for autonomous 3D-exploration of indoor environments," in *Proc. of IEEE Int. Conf. on IROS*, 2006.
- [12] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene labeling by relaxation operations," in *IEEE Trans. on Systems, Man and Cybernetics*, vol. 6, 1976, pp. 420–433.
- [13] D. C. Brown, "Close-range camera calibration," *Photogrammetric Engineering*, vol. 37, pp. 855–866, 1971.
- [14] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [15] A. Censi, "An accurate closed-form estimate of ICP's covariance," in *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2007.
- [16] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [17] K. O. Arras, "An introduction to error propagation: Derivation, meaning and examples of  $cy = fx \text{ } cx \text{ } fx'$ ," Autonomous Systems Lab, Swiss Federal Institute of Technology Lausanne, Tech. Rep. EPFL-ASL-TR-98-01 R3, 1998.
- [18] J. Stückler, "Visuelle Orientierung in Innenräumen mit einer Kamera (Visual indoor orientation with a camera, in German)," Diploma thesis, Computer Science Institute, University of Freiburg, Germany, 2007.