Proceedings of the 2000 IEEE/RSJ
International Conference on
Intelligent Robots and Systems

Multisensor On-the-Fly Localization Using Laser and Vision

Kai O. Arras, Nicola Tomatis, Roland Siegwart

Autonomous Systems Lab
Swiss Federal Institute of Technology Lausanne (EPFL)
CH–1015 Lausanne
{kai-oliver.arras, nicola.tomatis, roland.siegwart}@epfl.ch

Abstract

In this paper a multisensor setup for localization consisting of a 360 degree laser range finder and a monocular vision system is presented. Its practicability under conditions of continuous localization during motion in real-time (referred to as on-the-fly localization) is investigated in large-scale experiments. The features in use are infinite horizontal lines for the laser and vertical lines for the camera providing an extremely compact environment representation. They are extracted using physically well-grounded models for all sensors and passed to the Kalman filter for fusion and position estimation. Very high localization precision is obtained in general. The vision information has been found to further increase this precision, particular in the orientation, already with a moderate number of matched features. The results were obtained with a fully autonomous system where extensive tests with an overall length of more than 1.4 km and 9,500 localization cycles have been conducted. Furthermore, general aspects of multisensor on-the-fly localization are discussed.

1. Introduction

Localization in unmodified environments belongs to the basic skills of a mobile robot. In many potential service applications of mobile systems, the vehicle is operating in structured or semi structured surroundings. This property can be exploited by modelling these structures as geometric primitives and using them as reliably recognizable features for navigation. As it will be shown in this work, this approach leads to very compact environment descriptions which allow for precise navigation with the limited computational resources fully autonomous systems typically provide. Furthermore, due to the extraction step, which is essentially an abstraction from the type and amount of raw data, information from sensors of any kind can directly be included and managed in the same way, leading to versatile and easily extensible environment models.

In this paper we take advantage of this property by simultaneously employing geometric features from different sensors with complementary properties. We consider localization by means of infinite lines extracted from 1D range data of a 360° laser scanner and vertical edges extracted from images of an embarked CCD camera. An extended Kalman filter (EKF) is used for fusion and position estimation.

Navigation in a step-by-step manner where localization is performed only at standstill is unsatisfactory for many reasons: The vehicle advances slowly, it has not a continuous movement which is important for certain applications like cleaning tasks. The position update rate is low with respect to the distance travelled making the matching problem difficult for any localization method, and it is aesthetically suboptimal. Continuous localization during motion in real-time – henceforth referred to as *on-the-fly localization* – is therefore desirable but contains difficulties which are present but only hidden at low speed or step-by-step navigation. This includes resolution and uncertainties of time stamps the system can provide for sensory inputs. They impose bounds on localization precision and feature matching rates whose influence is to be studied when a localization method shall prove its relevance for real-world applications.

Kalman filter localization with line segments from range data has been done early [6][7]. Vertical edges in combination with an EKF have been employed in [4] and [8]. The combination of these features is used in [9] and [10]. In [9], a laser sensor providing range and intensity images within a 60° opening angle was utilized, and in a recent work [10], the absolute localization accuracy of laser, monocular and trinocular vision was examined. Similar precision has been found for the three cases.

In contrast to these contributions this paper reports extensive experiments where the practicability of this multisensor setup, the above mentioned features and an EKF is examined under application-like conditions. We consider the improvement with respect to precision when the vision information is added to the range information by means of the uncertainty bounds of the *a posteriori* position estimates. For this, throughout of this work it was attempted to employ physically well grounded uncertainty models for odometry, laser range finder and vision system.

2. Sensor Modelling

Odometry: Non-systematic odometry errors occur in two spaces: the joint space and the Cartesian space. With a differential drive kinematics the joint space is two-dimensional and includes the left and right wheel. Effects of wheel slippage, uneven ground and limited encoder resolution appear in this space. In [5] a physically well-grounded model for this kind of errors is presented starting from the uncertain input $u(k+1) = [\Delta d_L, \Delta d_R]^T$ with $\Delta d_L, \Delta d_R$ as the distances travelled by each wheel, and the diagonal input covariance matrix

$$U(k+1) = \begin{bmatrix} k_L|\Delta d_L| & 0 \\ 0 & k_R|\Delta d_R| \end{bmatrix} \quad (1)$$

which relies on the assumption of proportionally growing variances per $\Delta d_L, \Delta d_R$ travelled. The odometry model for the first and second moment of the state vector $x = (x, y, \theta)^T$ is then

$$x(k+1|k) = f(x(k|k), u(k+1)) \quad (2)$$

$$P(k+1|k) = \nabla f_x P(k|k) \nabla f_x^T + \nabla f_u U(k+1) \nabla f_u^T \quad (3)$$

where $f(.)$ uses a piecewise linear approximation, $P(k|k)$ is the state covariance matrix of the last step and $\nabla f_{x,u}$ is the Jacobian of $f(.)$ with respect to the uncertain vectors $x(k|k)$ and $u(k+1)$. k_L and k_R are constants with unit meter.

The Cartesian space is spanned by x encoding position and orientation of the vehicle. Effects of external forces (mainly collisions) occur in this space. Non-systematic Cartesian errors can be additionally modelled in eq. (3) by a 3×3 covariance matrix $Q(k+1)$ being a function of the robot displacement $\Delta x, \Delta\theta$ in the robot frame. Such a model has been used in [4]. In any case it is difficult to identify these models, i.e. to obtain rigorous values for k_L, k_R and the coefficients in $Q(k+1)$ which are valid for a range of floor types. In this work we used only the model of [5].

Laser Range Finder: The laser range finder which was used in the experiments is the Acuity AccuRange4000LIR. The rotation frequency of the mirror is 2.78 Hz, yielding a 1° angular resolution with its maximal sampling frequency in calibrated mode of 1 kHz. It delivers range ρ and intensity i as analogue signals. The latter is the signal strength of the reflected beam and predominantly affects range variance. In order to have a good physically based uncertainty model of range variability accounting not only for the distance to the target but also for its surface properties, a relationship $\sigma_\rho = f(i)$ is sought. Identification experiments with a Kodak gray scale patch performed in [2] yielded a simple relationship describable by two parameters: i_{min} allows to reject too uncertain range readings with $i < i_{min}$ and for measurement with $i > i_{min}$ a constant value for range variance $\sigma^2_{\rho_{const}}$ could have been found.

Camera: The vision system consists in a Pulnix TM-9701 full-frame, EIA (640 x 480), grayscale camera with an 90° objective and a Bt848 based frame grabber which delivers the images directly to the main CPU memory. There is no dedicated hardware for image processing.

The camera system is calibrated by combining method [11] with spatial knowledge from a test field. This provides a coherent set of extrinsic, intrinsic and distortion parameters. Since the visual features are vertical lines, only horizontal calibration is needed, yielding the simplified model of eq. (4) for parameter fitting

$$C \cdot \frac{y_r - \beta(x_r - O_x)}{\beta y_r + (x_r - O_x)} = S[x_c + x_c(k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8)] \quad (4)$$

(x_r, y_r, z_r) is the position of a point in the robot frame, $x_c = x - H_x$, $y_c = y - H_y$ and $r^2 = x_c^2 + y_c^2$, where the coordinates (x, y) refer to the distorted location of the point in the uncorrected image. Focal length C, scale factor S and image center (H_x, H_y) are instrinsic parameters, β and O_x are extrinsic parameters defining the robot to sensor transformation and k_1, k_2, k_3, k_4 are the parameters of radial distortion.

Uncertainties from the test field geometry and those caused by noise in the camera and acquisition electronics are propagated through the camera calibration procedure onto the level of camera parameters, yielding a 10×10 parameter covariance matrix.

3. Feature Representation and Extraction

Laser Range Finder: The algorithm for line extraction has been described in [1]. The method delivers lines and segments with their first order covariance estimate using polar coordinates. The line model is

$$\rho\cos(\varphi - \alpha) - r = 0 \quad (5)$$

where (ρ, φ) is the raw measurement and (α, r) the model parameters. α is the angle of the perpendicular to the line, r its length. The method differs from the widely used recursive split-and-merge technique which is also applied in [6] and [10] in the segmentation criterion: Instead of using a line specific decision on a single point, it decides on a model independent criterion on a group of points. Multiple segments which lie on the same physical object are merged for particular precise re-estimates of the line position. This is realized by an clustering algorithm with a Mahalanobis distance matrix. It merges segments until their distance in the (α, r)-model space is greater than a threshold from a χ^2-distribution. Figure 1 shows an extraction example where six lines have been found.

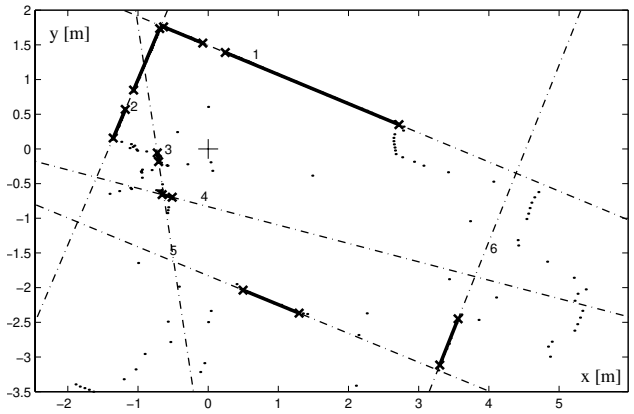

Figure 1: *A scan of the Acuity sensor and the extraction result. Eight segments on six lines have been found. Two closely situated objects produced evidence for the two 'outlier' segments. Thus, the local map contains six* (α, r)*-pairs which are passed to the EKF matching step.*

Camera: Vertical lines are extracted in four steps:

- Vertical edge enhancement: Specialized Sobel filter approximating the horizontal image gradient.
- Non-maxima suppression with dynamic thresholding: The most relevant edge pixels (maximal gradient) are extracted and thinned by using a standard method.
- Edge image calibration: The horizontal position of each edge pixel is corrected yielding a new position $\bar{x}$ with

$$\bar{x} = S[x_c + x_c(k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8)] \qquad (6)$$

  resulting from the camera model.
- Line fitting: Columns with a predefined number of edge pixels are labelled as vertical edges. Line fitting reduces to a one-dimensional problem. The resulting angle is $\varphi = \mathrm{atan}(x/C)$, where C is the focal length and x the weighted mean of the position of the pixels in the extracted line.

Uncertainty from the camera electronics is modelled on the level of the uncalibrated edge image. Together with the uncertainty of the calibration parameters it is propagated through calibration and line fit, yielding the first two moments $(\varphi, \sigma_\varphi^2)$ of the vertical edges.

Map: The a priori map contains 117 infinite lines and 172 vertical edges for the 50×30 m portion of the institute building shown in fig. 2. This is an environment model of extreme compactness with a memory requirement of about 30 bytes$/m^2$.

4. Multisensor EKF On-The-Fly Localization

Under the assumption of independent errors, the estimation framework of a Kalman filter can be extended with information from additional sensors in a straight-forward way. Since this paper does not depart from the usual use of extended Kalman filtering and first-order error propagation, most mathematical details are omitted. Please refer e.g. to [3] for a profound treatment and [7] for its use in the context of mobile robot localization. Only aspects which are particular are presented.

Matching: We assume independent errors between the sensors and between the features. Thus the observation covariance matrix $R(k+1)$ is blockwise diagonal and we have the freedom to integrate matched pairings in a manner which is advantageous for filter convergence:

The laser observations are integrated first since they typically exhibit far better mutual discriminance making their matching less error-prone, followed by the vertical edges from the camera where often ambiguous matching situations occur. Starting from the same idea, each pairing is integrated according to its quality in an iterative procedure for each sensor: *(i)* matching of the current best pairing, *(ii)* estimation and *(iii)* re-prediction of features not associated so far. This procedure has also been used in [9] and [10] where similar observations concerning feature discriminance have been made.

The quality of a pairing of prediction $\hat{z}_{l,v}^{[j]}$ and observation $z_{l,v}^{[i]}$ is different for both sensors:

- For the line segments the quality criterion of a pairing is *smallest observational uncertainty* – not smallest Mahalanobis distance like in [9] and [10]. This renders the matching robust against small spurious and uncertain segments which have small Mahalanobis distances (see fig. 1). The 'current best' pairing $(z_l^{[i]}, \hat{z}_l^{[j]})$ is therefore that of observation $z_l^{[i]}$ with $trace(R_l^{[i]}) = min_i$ which satisfies the validation test

$$(z_l^{[i]} - \hat{z}_l^{[j]})\, S_{ij}^{-1}\, (z_l^{[i]} - \hat{z}_l^{[j]})^T \le \chi_{\alpha,n}^2 \qquad (7)$$

  where S_{ij} is the innovation covariance matrix of the pairing and $\chi_{\alpha,n}^2$ a number taken from a χ^2 distribution with $n = 2$ degrees of freedom. α is the level on which the hypothesis of pairing correctness is rejected.
- The criterion for vertical edges is uniqueness. Predictions $\hat{z}_v^{[j]}$ with a single observation $z_v^{[i]}$ in their validation gate are preferred and integrated according to their smallest Mahalanobis distance provided that they satisfy eq. (7) with $n = 1$ (subscript l become v). When there is no unique pairing anymore, candidates with multiple observations in the validation region or observations in multiple validation regions are accepted and chosen according to the smallest Mahalanobis distance.

Time Stamps: The main difference from the viewpoint of multisensor localization between step-by-step and on-the-fly navigation is that temporal relations of sensor observations, predictions and estimations of all sensors involved have to be maintained and related to the present. This is

done by assigning time stamps to observations and recording odometry into a temporary buffer. When sensor A performs its data acquisition, the data receive a time stamp T_A and, after feature extraction is completed, the corresponding state prediction is read out from the odometry buffer. When the position estimate arrives from the Kalman filter, it is valid at time stamp T_A. Based on the odometry model, a means is then needed to relate this old position estimate to the current position of time t. This is done by forward simulation of eq. (2) and eq. (3) from T_A to t.

For a multisensor system, care has to be taken that prediction and estimation results of one sensor are not overwritten by those of another sensor. This would be the case if each sensor would have its own EKF running independently from the others with its own cycle time, yielding temporally nested updates. Nested updates are unfavourable since a slow outer update cycle (e.g. vision) overwrites the estimation results of faster running inner cycles (e.g. laser). A sequential scheme of EKFs for each sensor is therefore required with the constraint that the estimates get integrated in the succession of their respective observation.

Scan Compensation: The vehicle movement imposes a distortion on the raw data of the laser scans. This deformity depends on the ratio robot speed to mirror rotation velocity which in our case is non-negligible. It is important to note that scans have to be compensated on the raw data level and not on the feature level. Since in the latter case the extraction method would operate with an artifically modified features evidence.

We compensate for the vehicle displacement during a scan by transforming each range reading ${}^{S'}P$ acquired in the sensor frame S' into the non-stationary robot frame R' and then into the world frame W. Followed by a re-transform into the stationary robot frame R and finally into the desired reference frame of the scan S. For a compensation on-the-fly, S must be the sensor frame at the start position of a new scan. By reading out odometry each time when a new range reading arrives, it gets immediately transformed by the expression

$$ {}^{S}P = {}^{R}_{S}T^{-1}\, {}^{W}_{R}T^{-1}\, {}^{W}_{R'}T\, {}^{R'}_{S'}T\, {}^{S'}P, \qquad (8) $$

where ${}^{R}_{S}T = {}^{R'}_{S'}T$. The 4×4 matrices T are homogeneous transforms casting the rotation and translation of the general transform into a single matrix. ${}^{R}_{S}T$ is the sensor-to-robot frame transform and ${}^{W}_{R}T$ the world-to-robot transform given by the actual robot pose vector x. The compensated scan receives the time stamp of S, that is, the time when the scan has been started recording.

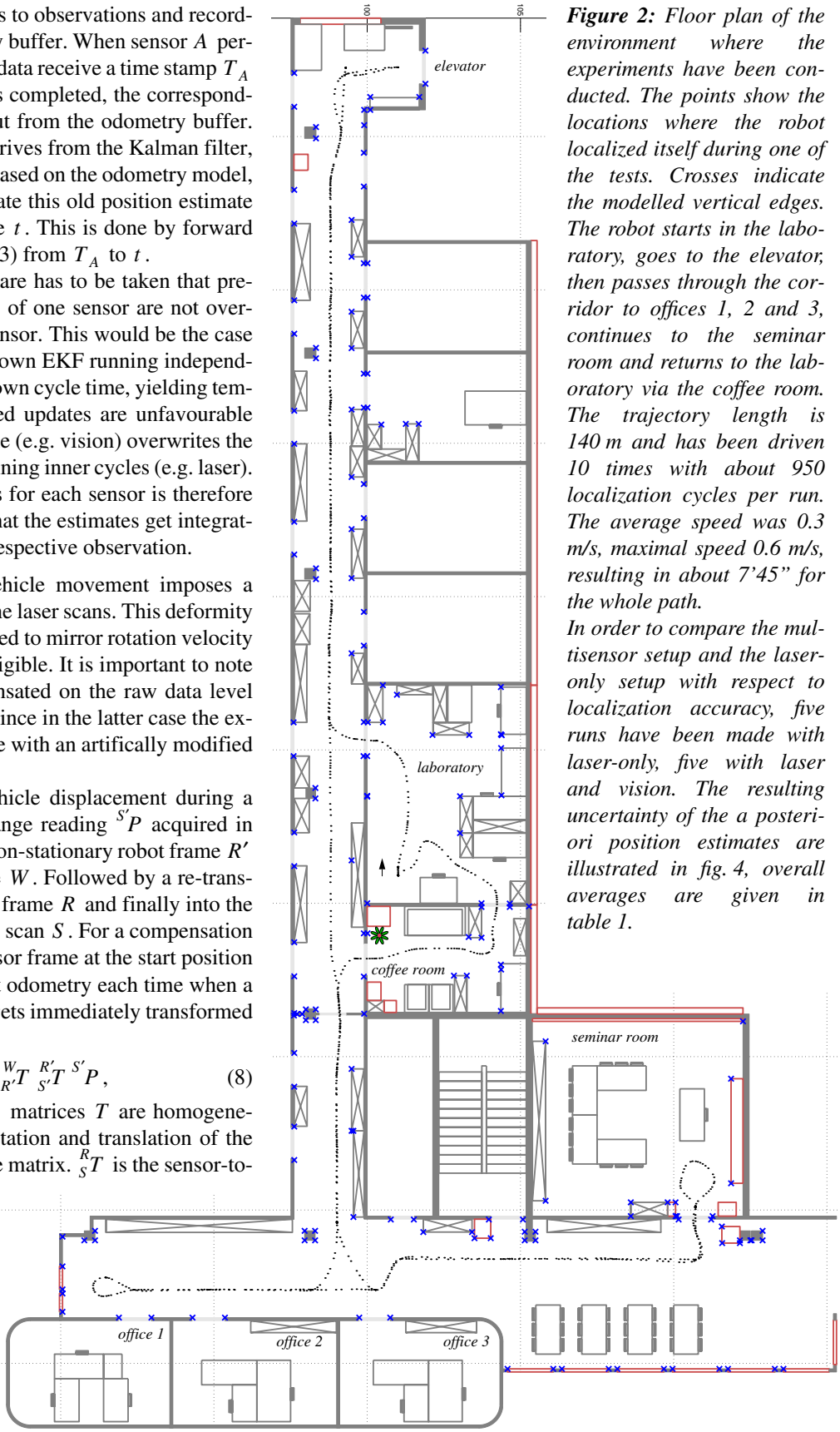

Figure 2: *Floor plan of the environment where the experiments have been conducted. The points show the locations where the robot localized itself during one of the tests. Crosses indicate the modelled vertical edges. The robot starts in the laboratory, goes to the elevator, then passes through the corridor to offices 1, 2 and 3, continues to the seminar room and returns to the laboratory via the coffee room. The trajectory length is 140 m and has been driven 10 times with about 950 localization cycles per run. The average speed was 0.3 m/s, maximal speed 0.6 m/s, resulting in about 7'45" for the whole path.*
In order to compare the multisensor setup and the laser-only setup with respect to localization accuracy, five runs have been made with laser-only, five with laser and vision. The resulting uncertainty of the a posteriori position estimates are illustrated in fig. 4, overall averages are given in table 1.

5. Implementation and Experiments

5.1 The Robot

Figure 3: *Pygmalion, the robot which was used in the experiments. It is a VME based system carrying currently a PowerPC card at 300 MHz. Besides wheel encoders and bumpers, the sensory system includes a 360° laser range finder and a gray-level CCD camera discussed in the second chapter. During the experiments it ran in a fully autonomous mode.*

Our experimental platform is the robot *Pygmalion* which has been built in our lab (fig. 3). Its design principles are oriented towards an application as service or personal robot. Long-term autonomy, safety, extensibility, and friendly appearance were the main objectives for design. With its dimensions of about 45x45x70 cm and its weight of 55 kg it is of moderate size and danger opposed to many robots in its performance class.

5.2 Experimental Results

The experiments have been conducted in the environment illustrated in figure 2. It shows the floor plan of a 50×30 m portion of the institute building. In the laser-only mode and in the multisensor mode the trajectory has been driven five times. The overall trajectory length is 1,4 km with 9,500 localization cycles. Care has been taken that both experiments had the same localization cycle time by limiting the implementation to 2 Hz resulting in about 950 cycles on the 140 m test trajectory. The average speed was 0.3 m/s, maximal speed 0.6 m/s. The robot was driven by its position controller for non-holonomic configurations. No obstacle avoidance was active.

The resulting 2σ-uncertainty bounds of the a posteriori position estimates are shown in figure 4. For both cases they generally reflect a very high localization accuracy in all three state variables. Subcentimeter precision is approached. Table 1 shows the overall means of error bounds $2\bar{\sigma}_x$, $2\bar{\sigma}_y$, $2\bar{\sigma}_\theta$, number of matches per localization cycle $\bar{n}_l$, $\bar{n}_v$, and execution times $\overline{t_{exe}}$. The vision information contributes equally to a reduction of uncertainty in x and y (–20%), but particularly in the orientation (–40%). This although the average number of matched vertical edges is moderate. A cycle time stands for one localization iteration under full CPU load and sensor data acquisition.

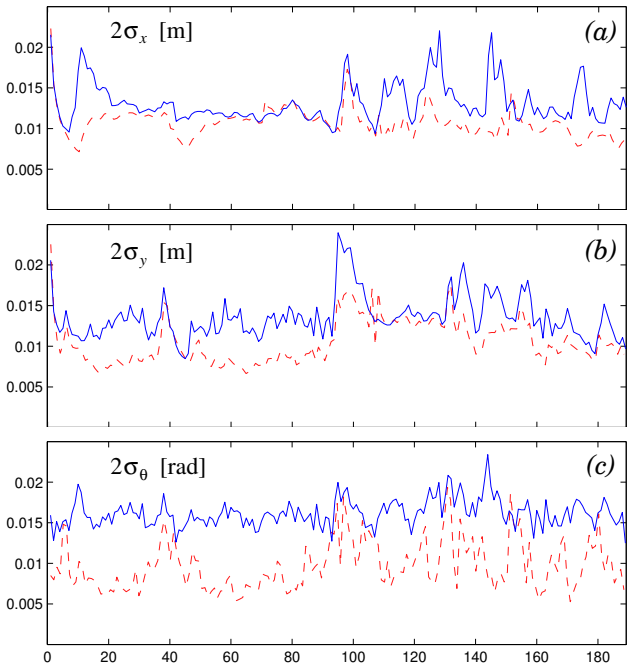

Figure 4: *Averaged 2σ–error bounds of global x (a), y (b) and θ (c) a posteriori uncertainty during the test trajectory (showing only each 5th step). In each mode, five runs have been made. Solid lines: laser range finder only, dashed lines: laser and vision. In some cases the uncertainty in the multisensor mode is greater than for the single-sensor setup. This is possible since the values are averaged over five runs containing noise on the level of matched features.*

	laser	laser and vision
$2\bar{\sigma}_x$	1.31 cm	1.07 cm
$2\bar{\sigma}_y$	1.35 cm	1.05 cm
$2\bar{\sigma}_\theta$	0.92°	0.56°
$\bar{n}_l / \bar{n}_v$	2.73 / –	2.66 / 2.00
$\overline{t_{exe}}$	64 ms	411 ms

Table 1: *Overall mean values of the error bounds, the number of matched line segments n_l and matched vertical edges n_v, and the average localization cycle time $\overline{t_{exe}}$ under full CPU load.*

5.3 Discussion

Even carefully derived uncertainty bounds do not necessarily permit inference about the sought first moments, since the estimation error could be arbitrarily big without being noticed (estimator inconsistency). We argue that the simple fact that the robot always succeeded in returning to its start

point is compelling evidence for the correctness of these bounds. In fact, they are even conservative estimates since the true bounds could be better. Otherwise the robot would have gone lost due to a lack of matches caused by first moments drifted away from the true values. Ground truth information like in [10] would be preferable but is impractical and expensive to obtain for experiments of this kind and extent. Positioning accuracy of the vehicle in the endpoint has been determined and further confirms the values in Table 1.

Matching vertical edges is, due to their lack of depth information and their frequent appearance in compact groups, particularly error-prone. For example, door frames commonly have multiple vertical borders which, dependent on the illumination conditions, produce evidence for several closely situated vertical edges. In the matching stage, they might be confronted with a large validation region, position bias from odometry or time stamp uncertainty making the predicted model edge difficult to identify. In such an ambiguous matching situations, incorrect pairings are likely to occur and, in fact, have been occasionally produced in the multisensor experiments. But their effect remains weak since these groups are typically very compact.

However, this lack of discriminance in the presence of time stamp uncertainty is the main cause of reproducible failure of vision-only navigation. With the frame grabber in use, it is difficult to identify the precise (i.e. down to a few ms) instant when the image is taken. Also odometry quantization (in our case 5 ms), furthermore bounding time stamp accuracy, became noticeable particularly during fast turns. (the camera of Pygmalion is not mounted on a turret which maintains a constant orientation). Modeling time stamp imprecision would yield larger validation gates around the predictions. But this does not solve the problem if matching situations are already found to be ambiguous.

6. Conclusions and Outlook

In this paper a multisensor setup for localization consisting of a 360° laser range finder and a monocular vision system is presented. It combines infinite horizontal lines from the laser and vertical edges from the camera. Its practicability under conditions of on-the-fly localization is investigated in large-scale experiments. Very high localization precision is achieved with an extremely compact environment description provided by the employed features. The vision information has been found to further increase this precision, particular in the orientation, already with a moderate number of matched edges. By having performed extensive tests with a fully autonomous system on an overall length of more than 1.4 km and 9,500 localization cycles we demonstrated the relevance of the localization setup for real-world applications.

Vision-only navigation failed in our experiments. This is due to the modest mutual discriminance of vertical edges, making them difficult to match, in combination with non-negligible time stamp uncertainties. This motivates the use of constraint-based matching schemes with this type of feature particularly for vision-only navigation. Future work will focus on such matching techniques by introducing unary or binary constraints. Besides, more complex vision features shall be employed for semantically richer environment models.

Acknowledgements
The authors would like to thank Andrea Terribilini for valuable discussions on camera calibration and Iwan Märki for his help with the a priori map.

References

[1] Arras K.O., Siegwart R.Y., "Feature Extraction and Scene Interpretation for Map-Based Navigation and Map Building", Proc. of SPIE, Mobile Robotics XII, Vol. 3210, p. 42-53, 1997.

[2] Arras K.O., Tomatis N., "Improving Robustness and Precision in Mobile Robot Localization by Using Laser Range Finding and Monocular Vision", Proc. of the 3rd European Workshop on Advanced Mobile Robots (Eurobot 99), Zurich, Switzerland, 1999.

[3] Bar-Shalom Y., Fortmann T.E., *Tracking and Data Association*, Mathematics in Science and Engineering, Vol. 179, Academic Press Inc., 1988.

[4] Chenavier F., Crowley J.L., “Position Estimation for a Mobile Robot Using Vision and Odometry”, Proc. of the 1992 IEEE Int. Conf. on Robotics and Automation, Nice, 1992.

[5] Chong K.S., Kleeman L., “Accurate Odometry and Error Modelling for a Mobile Robot”, 1997 IEEE Int. Conf. on Robotics and Automation, NM, USA, 1997.

[6] Crowley J.L., "World Modeling and Position Estimation for a Mobile Robot Using Ultrasonic Ranging," Proc. of the 1989 IEEE Int. Conf. on Robotics and Automation, Scottsdale, AZ, 1989.

[7] Leonard J.J., Durrant-Whyte H.F., *Directed Sonar Sensing for Mobile Robot Navigation*, Kluwer Academic Publishers, 1992.

[8] Muñoz A.J., Gonzales J., “Two-Dimensional Landmark-based Position Estimation from a Single Image”, Proc. of the 1998 IEEE Int. Conf. on Robotics and Automation, Leuven, Belgium, 1998.

[9] Neira J., Tardos J.D., Horn J., Schmidt G., "Fusing Range and Intensity Images for Mobile Robot Localization," IEEE Trans. on Robotics and Automation, 15(1):76-84, 1999.

[10] Pérez J.A., Castellanos J.A., Montiel J.M.M., Neira J. and Tardós J.D., "Continuous Mobile Robot Localization: Vision vs. Laser," Proc. of the 1999 IEEE Int. Conf. on Robotics and Automation, Detroit, 1999.

[11] Prescott B., McLean G.F., “Line-Based Correction of Radial Lens Distortion”, Graphical Models and Image Processing, 59(1), 1997, p. 39-47.