Sünderhauf, N., Lange, S., & Protzel, P. (2007). Using the Unscented Kalman Filter in Mono-SLAM with Inverse Depth Parametrization for Autonomous Airship Control. In Proc. of IEEE International Workshop on Safety Security and Rescue Robotics, SSRR 2007, Rome, Italy.

Using the Unscented Kalman Filter in Mono-SLAM with Inverse Depth Parametrization for Autonomous Airship Control

Niko Sünderhauf, Sven Lange and Peter Protzel
Department of Electrical Engineering and Information Technology
Chemnitz University of Technology
Chemnitz, Germany
{niko.suenderhauf, sven.lange, peter.protzel}@etit.tu-chemnitz.de

Abstract — In this paper, we present an approach for aiding control of an autonomous airship by the means of SLAM. We show how the Unscented Kalman Filter can be applied in a SLAM context with monocular vision. The recently published Inverse Depth Parametrization is used for undelayed single-hypothesis landmark initialization and modelling. The novelty of the presented approach lies in the combination of UKF, Inverse Depth Parametrization and bearing-only SLAM and its application for autonomous airship control and UAV control in general.

Keywords: *airship, monocular SLAM, Unscented Kalman Filter.*

I. Introduction

A. The Airship "Fritz"

The airship "Fritz" (see fig. 1 for an image) is a 9 meter long blimp (i.e. it has no rigid skeleton) and is the main part of our research project on rescue robotics. Our goal is to provide an "eye in the sky" to fire brigades, police, emergency response, civil protection and disaster control units. An autonomous or semi-autonomous airship can be used to provide valuable information about the situation during large scale operations with many casualties, or large affected areas like train accidents, flooding, major fire and the like.

Our airship "Fritz" is equipped with a IEEE1394 color camera (Guppy F-046/C from Allied Vision Tech), an infrared camera from FLIR, a Holux GPS receiver and an inertial measurement unit from Crossbow. An Intel Core Duo PC system running Linux does all the high-level work and commands a microcontroller board based on an ATMEGA32 for low-level and hardware related tasks. The airship can be remote controlled if necessary or commanded via wireless connection from the ground.

The very recent development of a novel and now patented [1] propulsion system for "Fritz" is described in [2]. It allows higher agility and maneuverability than with the usual propulsion system where two engines are mounted rigidly on a common joint. Current work comprises autonomous navigation and control, a comparison between stereo and monocular SLAM approaches for UAVs, enhancement of the existing cascaded controller structure, autonomous infrared vision based surveillance and object recognition.

Fig. 1. Our airship Fritz during a flight above the university campus.

In past work [3], [4], a cascaded controller structure was developed for the airship. It consists of three consecutive levels, responsible for control of acceleration, velocity and position. Figure 2 shows a block diagram of the two inner levels. Optical flow was used as the only sensor input in the past.

We now want to replace this fragile approach by a more sophisticated method to provide more sound and meaningful inputs to all three levels of the controller. These inputs shall be generated by the means of short-term SLAM, by combining inertial sensors and GPS with the information provided by a monocular camera. Of course, the methods presented in this paper can not only be applied to airships, but are applicable for other UAV and autonomous vehicle in general.

B. Short-Term SLAM

With "short-term SLAM" we mean SLAM in situations where GPS signals or other sources of global localization are available. Short-term SLAM is used to estimate the vehicle state, especially velocity and angular velocity, but also a more precise vehicle position and orientation in order to aid the

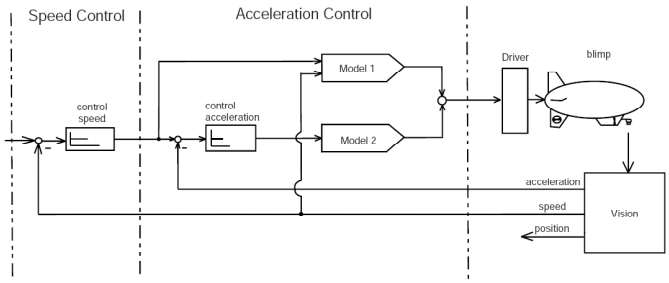

Fig. 2. The inner part of the cascaded control loop running on the airship (Image reproduced from Krause & Protzel [3]). The box named "Vision" on the right is a framework based on optical flow and estimates vehicle acceleration, speed and position. We aim to replace it by a more robust and sound state estimator.

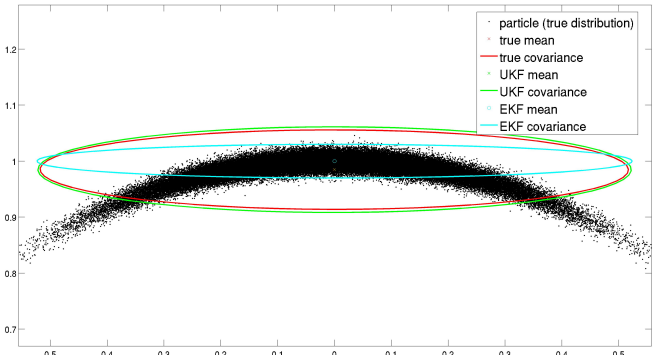

Fig. 3. Comparing UT and EKF-linearization: A single set of polar coordinates (modelled as Gaussian) is converted to cartesian coordinates. The black particles sample the true resulting distribution. The true converted Gaussian is shown in red. The green mean and covariance (from the Unscented Transform) are almost identical with it. The red mean and covariance (from the EKF-like linearization) are both inconsistent with the true Gaussian.

cascaded control loop that runs on the airship [3] [4]. As loop-closing is not necessary, landmarks that are not seen anymore are removed from the filter. This ensures an upper bound to the state vector size.

In contrast to this, if no sources of global localization are available (because GPS signals are disturbed, jammed or not available, like in planetary domains), loop-closing is necessary to prevent localization errors grow without bounds. In this case, however, the approach presented here can not be applied without further adjustments because the number of landmarks and with it the filter update complexity can grow without bounds. Combinations with SLAM techniques that scale better with the number of landmarks in the map may be applicable in this situation. However, these are still open questions for further research.

C. Related Work

The field of monocular SLAM has been very active during the past two years. Among others, Andrew Davison contributed a lot of work to the field [5] [6]. He was also involved in the development of the Inverse Depth Parametrization [7] [8]. The work of Eade & Drummond addresses monocular SLAM with a particle filter approach [9]. Other recent work that performs mono SLAM in an EKF framework are two dissertations from LAAS/Toulouse [10] [11] and the accompanying publications. One of those [12] elaborates a technique of how to combine the advantages of both mono and stereo vision SLAM approaches.

Based on the Unscented Kalman Filter and the original work of Julier & Uhlmann, van der Merve developed a whole new class of filters, coined Sigma Point Kalman Filters, and proved their feasibility for several applications, including control of an autonomous helicopter. See his dissertation [13] and accompanying papers like [14] for more interesting details.

Finally, interesting work has been done by Langelaan about autonomous UAV flight in forests [15] [16]. His work uses the UKF and a monocular camera. The key difference to our approach is the feature initialization, where we use the Inverse Depth Parametrization for undelayed initialization, while he uses either delayed initialization or a rather simple way of pre-estimating the true landmark position by intersecting the landmark's line of sight with the estimated ground plane.

D. Paper Outline

The next sections review the Unscented Kalman Filter, Mono-SLAM techniques and the Inverse Depth Parametrization before illustrating how they can be combined to aid autonomous airship navigation and control.

II. The Unscented Kalman Filter

The Unscented Kalman Filter (UKF) was first published by Julier and Uhlmann [17]. It is a derivative-free recursive linear estimator comparable but superior to the Extended Kalman Filter. The authors showed that the core of the UKF, the Unscented Transform (UT) captures mean and covariance precisely up to the second order, while the EKF captures the mean only up to the first order. The underlying idea is that "it is easier to approximate a Gaussian distribution, than to approximate an arbitrary nonlinear function". How much better the Unscented Transform captures the mean and covariance of a Gaussian distribution that was propagated through a nonlinear function is shown in fig. 3. A simple conversion from polar to cartesian coordinates reveals the flaws of the EKF-like linearization.

By sampling a Gaussian distribution with a fixed number of so called sigma-points, and passing these sigma-points through the desired nonlinear function or transformation, the UT (and with it the UKF) avoids linearization by taking explicit derivatives (Jacobians), which can be very hard in some cases. Once the sigma-points are passed through the nonlinear function, mean and covariance of the resulting transformed distribution can be retrieved from them. The UT-sampling is a *deterministic* sampling, in contrast to techniques like particle filters, that sample *randomly*. This way, the number of samples can be kept small, compared to particle filters: To sample an n-dimensional distribution, $2n+1$ sigma-points are necessary. Further improvements on the UT, like [18] reduce this number to $n+2$.

The reader is referred to the literature for further information about the UKF, UT and its enhancements. Especially how the sigma-points and their weights are calculated, and how mean and covariance are restored from them is not explained any further in this work.

III. Mono-SLAM and the Inverse Depth Parametrization

A. Monocular SLAM

Many algorithms and approaches for vision-based SLAM, localization or visual odometry have been published during the past years. Many of them use stereovision where two cameras are mounted in a known, calibrated configuration. Stereo cameras provide bearing and range measurements, i.e. the 3D-coordinates of observed landmarks are – to a certain extend – known. However, the 3D range of a stereo camera is limited and the immanent uncertainty of a landmark's estimated 3D-coordinates gets larger with its distance from the camera. While this rising uncertainty can be taken into account during the estimation process [19] [20], depending on the camera parameters (base length, focal length, pixel size) the range measurement of a stereo camera bears no more viable information for landmarks from a certain distance. This maximum observable range may be sufficiently large for ground operating robots with acceptable baseline lengths (like 10 cm). Flying robots however, especially airships that operate in altitudes of 100 meter and beyond, need larger baselines. The largest possible baseline length at our airship is (at the moment) 120 cm. After calibration, even the slightest change in the camera orientations leads to massive errors in depth estimation, as stereo vision is very dependend on a good and enduring calibration. Large baselines lead either to more fragile and instable constructions (which are prone to de-calibration) or heavier and massive constructions (meaning higher weight and less payload on an airship). Both is not desirable.

Due to these flaws of stereo approaches, several authors addressed the problem with single or mono camera methods. As a single camera can not provide range measurements, the resulting algorithms are often called bearing-only SLAM or simply monocular SLAM.

Solà [10] gives a vivid motivation for monocular approaches: Using stereo algorithms and having to discard landmarks that are beyond the range of stereo observability is like "walking in dense fog", where you only see the close surrounding, but nothing beyond it. It is intuitively clear that orientation in dense fog is by far more complicated than in clear sight, where landmarks may be too far away to estimate their distance but still provide very usefull information about orientation and course (e.g. the polar star or distinctive landmarks on the horizon). Bearing-only SLAM aims to make use of this information rather than to discard it.

B. Landmark Detection

Landmark detection and matching can be performed with sophisticated and robust features like SURF [21] (we successfully applied SURF features in a FastSLAM context very recently [22]) or with a more "quick and dirty"-like solution with the Harris detector [23] paired with a correlation based matching. Of course, many other feature detector and matching schemes can be applied. However, depending on the desired framerate one may be forced to prefer simple but quick techniques over robust but slow ones.

C. The Inverse Depth Parametrization

With a single observation from a monocular camera, the landmark's position in the world can not be determined. The Inverse Depth Parametrization introduced by Montiel et al. [7] is a convenient way to completely initialize a landmark after it has been observed for the first time and to express its unknown depth. In this section we summarize the idea of Montiel et al. but refer the reader to the original paper for further details and considerations.

A single camera projects points from the 3D world coordinate frame onto a 2D image frame. As one dimension is lost during the projection, the process is not invertible. If a landmark is observed with a single camera, only its bearing is known: From the pixel coordinates $\begin{pmatrix} u & v \end{pmatrix}$, two angles $\begin{pmatrix} \theta & \phi \end{pmatrix}$ can be retrieved and transformed from camera into the world coordinate frame. These two angles, combined with the camera position $\begin{pmatrix} x & y & z \end{pmatrix}$ form a ray in the world. The landmark has to lie somewhere on that ray, however *where* on the ray is not determined.

The landmark can be modeled as

$$\mathbf{p} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + r \cdot \mathbf{m}_{(\theta,\phi)} \tag{1}$$

Here $\mathbf{m}_{(\theta,\phi)}$ is the ray directional vector and r is the unknown depth. As the exact position of the landmark on the ray is not known, r can range from 0 to ∞. Expressing this uncertainty with a single Gaussian hypothesis is not very comfortable. Other approaches use multi-hypothesis formulations were the depth of the landmark is expressed using a series of Gaussians. See for example [10] [11] [24] [5]. Montiel's inverse depth parametrization expresses the unknown depth using its inverse: $\rho = 1/r$. This way, we can write

$$\mathbf{p} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \frac{1}{\rho} \cdot \mathbf{m}_{(\theta,\phi)} \tag{2}$$

For example, to express landmark ranges from ∞ to 1 meter, ρ takes convenient values of $0 \ldots 1$ and can be easily expressed with a single Gaussian distribution with mean 0.5 and std. deviation 0.2.

To summarize the advantages of the Inverse Depth Parametrization: it allows undelayed landmark initialization using a single, Gaussian hypothesis. Montiel et al. [7] also point out that the inverse parametrization bears less linearization errors when used in an EKF SLAM framework. A yet to be answered question is whether this effect also applies to the UKF framework, which itself reduces linearization errors in the first place.

IV. Filter Setup

The filter setup described next is responsible for generating meaningful estimates of the airship's state in the world. The filter results are used as inputs to the two higher controller levels that control linear and angular velocity as well as position and orientation.

A. State Vector - Airship

Besides position and orientation of the system, we need to keep track of its linear and angular velocity. The estimated position and orientation serves as input for the position controller, the speed controller uses the estimated velocities. The vehicle state vector $\mathbf{x}_V$ is therefore defined as

$$\mathbf{x}_V = \begin{pmatrix} x & y & z & \psi & \theta & \phi & \mathbf{v} & \omega & \mathbf{b} \end{pmatrix}^T \tag{3}$$

The angles $\begin{pmatrix} \psi & \theta & \phi \end{pmatrix}$ describe the airships orientation in the world frame. We avoid using quaternions for reasons of simplicity here. Gimbal lock situations should not occur in our scenario as the pitch and roll angles of an airship are going to remain relatively small during normal flight. However, for aerial vehicles that are more agile and can conduct more aggressive flight maneuvers, using the above Euler angles notation can bear severe problems. Quaternions should be preferred in this case. Van der Merwe addresses some arising issues like how to keep the quaternion unit norm constraint during state estimation in his thesis [13]. $\mathbf{v}$ and ω describe linear and rotational velocity in the world frame, $\mathbf{b}$ is a vector containing the IMU biases.

B. State Vector - Landmarks

Each landmark is modeled using the inverse depth parametrization as

$$\mathbf{p}_i = \begin{pmatrix} x_i & y_i & z_i & \theta_i & \phi_i & \rho_i \end{pmatrix}^T \tag{4}$$

The landmarks form the map state vector $\mathbf{x}_M$.

C. Control Input

The control input $\mathbf{u}$ is modeled as

$$\mathbf{u} = \begin{pmatrix} a_x & a_y & a_z & \omega_\psi & \omega_\theta & \omega_\phi \end{pmatrix}^T \tag{5}$$

It contains linear accelerations and rotations as measured by an inertial measurement unit in the vehicle body frame.
The advantage of modelling $\mathbf{u}$ in the way described above and not using the controller's commands (which are linear and rotational accelerations a and $\dot{\omega}$) is that $\mathbf{u}$ automatically captures effects caused by wind and squalls as their influence on the airship is directly measured by the IMU.

D. State Prediction

The predicted updated vehicle state $\mathbf{x}_V^+$ is given by the time-discrete nonlinear function g:

$$\mathbf{x}_V^+ = g(\mathbf{x}_V, \mathbf{u}) = \begin{pmatrix} \begin{pmatrix} x & y & z \end{pmatrix}^T + \mathbf{v}\Delta t \\ \begin{pmatrix} \psi & \theta & \phi \end{pmatrix}^T + \omega \Delta t \\ \begin{pmatrix} v_x & v_y & v_z \end{pmatrix}^T + \mathbf{a}^{\mathcal{W}} \Delta t \\ \omega^{\mathcal{W}} \\ \mathbf{b} \end{pmatrix} \tag{6}$$

Here the upper-index $^{\mathcal{W}}$ denotes that the control input from $\mathbf{u}$ (which was originally given in the vehicle body frame) is transferred into the world coordinate frame. The nonlinearity of g lies in this transformation.
As the landmarks are considered to be stationary, the map state vector $\mathbf{x}_M$ is not changed during the state prediction step of the filter. Notice that it would, however, be possible to track moving landmarks by adding their estimated speed to the state vector.

E. Measurement Equation

Besides the monocular camera a GPS receiver and an inertial measurement unit (measuring linear acceleration and angular velocity) are available. The supplied measurements are:

$$\mathbf{z}_{GPS} = \begin{pmatrix} x & y & z \end{pmatrix}^T \tag{7}$$

$$\mathbf{z}_{IMU} = \begin{pmatrix} a_x & a_y & a_z & \omega_\psi & \omega_\theta & \omega_\phi \end{pmatrix}^T \tag{8}$$

$$\mathbf{z}_{camera} = \begin{pmatrix} u_i & v_i \end{pmatrix}_{1\ldots m}^T \tag{9}$$

The measurements $\mathbf{z}_{IMU}$ of linear acceleration and rotation are already contained in the filter setup by using them in the control vector $\mathbf{u}$. They are therefore not considered a second time in the filter loop and ignored in the measurement equation below. The monocular camera projects the landmarks onto its image plane and returns pixel coordinates $\begin{pmatrix} u_i & v_i \end{pmatrix}$ for each of the m observed landmarks.

$$\mathbf{z} = h(\mathbf{x}) = \begin{pmatrix} \begin{pmatrix} x & y & z \end{pmatrix}^T \\ \begin{pmatrix} u_i \\ v_i \end{pmatrix}_{1\ldots m} \end{pmatrix} \tag{10}$$

F. Combined & Augmented State Vector for SLAM

The complete state-vector $\mathbf{x}$ is formed by stacking the vehicle state vector and the map state vector: $\mathbf{x} = \begin{pmatrix} \mathbf{x}_V & \mathbf{x}_M \end{pmatrix}^T$. With 6 parameters per landmark and 18 parameters for the airship's state, $\mathbf{x}$ has dimension $18 + 6 \cdot m$ where m is the number of tracked landmarks.
In order to correctly account the uncertainty in the control input $\mathbf{u}$ and the measurements $\mathbf{z}$ we have to augment the state vector $\mathbf{x}$ with a vector containing the control input noise, $\mathbf{x}_u$ and the measurement noise $\mathbf{x}_z$. This was already proposed by Julier and Uhlmann in their original paper [17]. The final augmented vector is then formed by

$$\mathbf{x}^a = \begin{pmatrix} \mathbf{x}_V & \mathbf{x}_M & \mathbf{x}_u & \mathbf{x}_z \end{pmatrix}^T \tag{11}$$

Its dimensionality is $18+6\cdot m+6+2\cdot m$, as $\mathbf{u}$ has 6 entries. If the control inputs are not biased, then $\mathbf{x}_u$ and $\mathbf{x}_z$ are initialized with $\mathbf{0}$.
The state covariance matrix Σ is augmented as well and contains the covariances of the vehicle states, the landmarks, the control inputs and the measurements.

V. UKF – Filter Loop

A. Algorithm Description

From the current system state $\mathbf{x}^a$ and the covariance matrix Σ, a new set of sigma-points $\mathcal{X}_i$ is generated. This is done using the scaled unscented transform SUT [25]. The SUT prevents problems with non-positive semidefinite covariance matrices that can occur in the original unscented transform. However, a pretty large number of sigma-points is needed to correctly capture mean and covariance of $\mathbf{x}^a$. Using the SUT, $2n + 1$ sigma-points are generated, where n is the dimensionality of the state vector. The Spherical Simplex Unscented Transform [18] can be used instead to reduce the number of needed sigma-points to $n + 2$.
The sigma-points $\mathcal{X}_i$ are passed through a modified version of the state estimation equation as given in (6). The function g is modified in a way that it not only accounts the vehicle state part of $\mathcal{X}_i$, but the control and measurement noise part in $\mathcal{X}_i$ (arising from $\mathbf{x}_u$ and $\mathbf{x}_z$) as well. This way, the effect of the control and measurement noise is accurately modelled (better than in the standard EKF framework). We yield a set of sigma-points that encode the predicted system state and covariance:

$$\bar{\mathcal{X}}_i = g(\mathcal{X}_i, \mathbf{u}) \tag{12}$$

Given $\bar{\mathcal{X}}_i$ we can predict the measurements the sensors should have provided if $\bar{\mathcal{X}}_i$ was the true system state. This is done using a modified version of the measurement equation (10) which will again provide a set of sigma-points $\bar{\mathcal{Z}}$ that can be decoded to the mean and covariance of a Gaussian distribution describing the predicted measurements. From here, the rest of the algorithm is straight forward and follows the standard UKF scheme. The Kalman gain K is computed and the predicted system state and covariance $\bar{\mu}$ and $\bar{\Sigma}$ are corrected accordingly. The algorithm returns a new estimated system state vector $\mathbf{x}^+$ and its covariance Σ^+.

The entries of $\mathbf{x}^+$ serve as input to the cascaded control loop that controls the airships position, orientation, velocity and acceleration. Details can be found in [3], [2]. and [4].

B. Some Thoughts about Runtime Complexity

Clearly, runtime complexity is directly dependend on the state vector dimension n. Using the Unscented Transform, $2n + 1$ sigma-points are generated. Both state-update and measurement function are evaluated for each sigma-point. The UT further involves a calculation of a matrix square root from the covariance matrix Σ which has runtime complexity $\mathcal{O}(n^3/6)$. A special implementation that avoids the calculation of the matrix square root is described in [26]. The overall complexity remains $\mathcal{O}(n^3)$, but this square root form of the filter is 20% faster than the conventional implementation.

Keeping the state vector small is essential to achieve high framerates. As we already pointed out in the beginning, we perform short-term SLAM, i.e. we do not have to keep a large number of landmarks in the map. Landmarks that are not seen anymore can be discarded, as no loop-closing is necessary due to the global localization information via GPS.

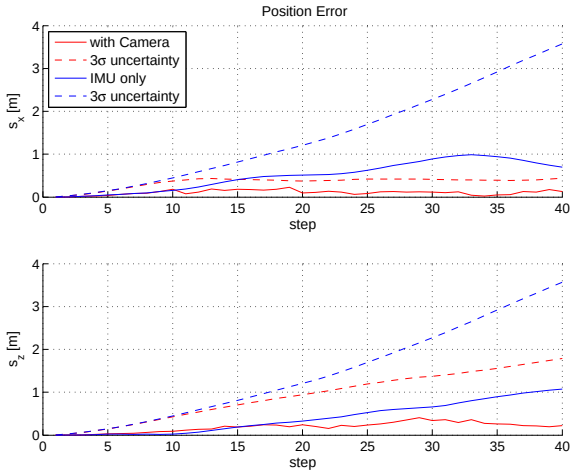

Fig. 4. Absolute position errors of the unfiltered (blue) and filtered (red) estimates of the vehicle position in x and z direction. The lower errors from the filtered estimates clearly prove that UAV navigation can be successfully improved by monocular short-term SLAM. Notice that no GPS measurements were simulated here.

VI. Simulation

We set up a Matlab framework and implemented the proposed algorithm (using its square root form) to show the feasibility of the approach and to identify possible pitfalls.

In the simulation, the camera moves along a curved planar path towards 5 3D landmarks (fig. 6). We did not simulate GPS measurements in order to show the cababilities of the proposed algorithm and to prove that short-term UAV navigation is possible during GPS outages. Figure 4 shows the absolute errors of the predicted vehicle world position using only IMU measurements (blue) and the corrected estimates from the filter (red) along with the respective standard deviations (dashed). The filter bears a clear improvement over the unfiltered estimates. This is especially obvious with the standard deviation of s_x (perpendicular to the landmarks) that grows without bounds for the unfiltered case but converges for the filtered estimates. The improvements of the orientation estimates are smaller but also visible as can be seen in fig. 5.

A major problem that occurred during the simulation runs is that some of the generated sigma-points happen to lie behind the camera. In this case, the measurement function h is undefined. A possible method of avoiding this effect is described in [27]. However, the proposed strategy of adapting the α-parameter in the unscented transformation is not optimal. A better method for constraining the sigma-points to meaningful coordinates should be developed in future research.

VII. Results & Conclusion

In this paper, we illustrated the concept of our approach to aid the autonomous control of airships by the means of SLAM. We motivated the use of monocular short-term SLAM and the Inverse Depth Parametrization and justified the use of the UKF instead of the EKF as state estimator scheme. The shown results from the simulation proved the general feasibility of the

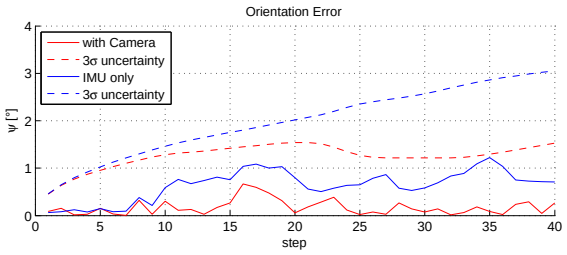

Fig. 5. Absolute orientation errors for the unfiltered (blue) and filtered (red) estimates. Again, the estimates generated by the filter are superior to the unfiltered results, although the differences are smaller than for the position estimates.

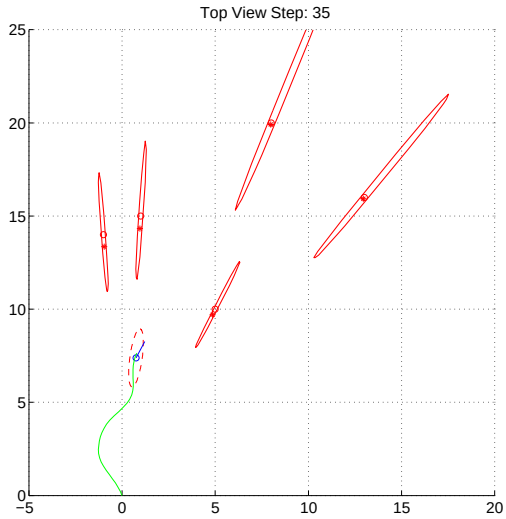

Fig. 6. Top view of the simulated world with 5 landmarks (the real position is marked by the stars, the predicted position by the circles) and the vehicle along with the respective uncertainty ellipses. No GPS measurements were used in the filter.

approach, so practical experiments will be conducted on our airship in future work.

Besides developing a method for avoiding ill-conditioned sigma-points behind the camera and an analysis on the reduction of linearization errors that arise from the use of the UKF compared to the EKF technique in [7], practical experiments with our airship should prove the algorithm's feasibility in reality in upcoming work. Other questions to be answered include the achievable framerates in relation to the number of observed landmarks and how the number of landmarks found by the feature detector (like SURF) can be pruned in a sensible way.

References

[1] Thomas Krause. Luftschiff. Patent DE 10 2007 013 147.1, 2007.

[2] Thomas Krause and Peter Protzel. Verteiltes, dynamisches antriebssystem zur steuerung eines luftschiffes. In *Tagungsband 52. Internationales Wissenschaftliches Kolloquium*, Ilmenau, Germany, September 2007.

[3] Thomas Krause, Pedro Lima, and Peter Protzel. Flugregler für ein autonomes luftschiff. In Levi, Schanz, Lafrenz, and Avrutin, editors, *Tagungsband Autonome Mobile Systeme 2003*, Reihe Informatik aktuell, pages 83–90. Springer-Verlag, November 2003.

[4] Pedro Lima, Luís Custódio, M. Isabel Ribeiro, and José Santos-Victor. The rescue project: Cooperative navigation for rescue robots. In *Proceedings do 1st International Workshop on Advances in Service Robotics, ASER03,*, Bardolino, Italy, September 2003.

[5] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the ninth international Conference on Computer Vision ICCV'03, IEEE Computer Society Press.*, Nice, France, October 2003.

[6] Andrew J. Davison, Ian Reid, Nicholas Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):1052–1067, 2007.

[7] J. Montiel, J. Civera, and A. Davison. Unified inverse depth parametrization for monocular slam. In *Proceedings of Robotics: Science and Systems*, Philadelphia, USA, August 2006.

[8] J. Civera, A. Davison, and J. Montiel. Inverse depth to depth conversion for monocular slam. In *Proceedings of the International Conference on Robotics and Automation, ICRA07*, Rome, Italy, 2007.

[9] Ethan Eade and Tom Drummond. Scalable monocular slam. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 469–476, Washington, DC, USA, 2006. IEEE Computer Society.

[10] Joan Solà. *Towards Visual Localization, Mapping and Moving Objects Tracking by a Mobile Robot: a Geometric and Probabilistic Approach.* PhD thesis, LAAS, Toulouse, 2007.

[11] Thomas Lemaire. *SLAM with Monocular Vision*. PhD thesis, LAAS, Toulouse, 2007.

[12] Joan Solà, André Monin, and Michel Devy. Bicamslam: Two times mono is more than stereo. In *Proc. of IEEE International Conference on Robotics and Automation, ICRA07*, Rome, Italy, April 2007.

[13] Rudolph van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, Oregon Health & Science University, 2004.

[14] R. van der Merwe and E. Wan. Sigma-point kalman filters for integrated navigation. In *Proceedings of the 60th Annual Meeting of The Institute of Navigation (ION)*, Dayton, OH, June 2004.

[15] Jacob Willem Langelaan. *State Estimation for Autonomous Flight in Cluttered Environments*. PhD thesis, Department of Aeronautics and Astronautics, Stanford University, 2006.

[16] Jack Langelaan and Steve Rock. Towards autonomous uav flight in forests. In *Proc. of AIAA Guidance, Navigation and Control Conference*, San Francisco, USA, 2005.

[17] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls, Orlando, FL*, 1997.

[18] The spherical simplex unscented transformation. In *Proceedings of the IEEE American Control Conference*, Denver, USA, June 2003.

[19] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3:239–248, 1987.

[20] C. Olson, L. Matthies, M. Schoppers, and Maimone Maimone. Robust stereo ego-motion for long distance navigation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-00)*, pages 453–458, Los Alamitos, June 13–15 2000. IEEE.

[21] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Proceedings of the ninth European Conference on Computer Vision*, May 2006.

[22] Peer Neubert, Niko Sünderhauf, and Peter Protzel. Fastslam using surf features: An efficient implementation and practical experiences. In *Proceedings of the International Conference on Intelligent and Autonomous Vehicles, IAV07*, Tolouse, France, September 2007.

[23] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference 1988*, pages 147–151, 1988.

[24] Joan Solà, Michel Devy, André Monin, and Thomas Lemaire. Undelayed initialization in bearing only slam. In *IEEE International Conference on Intelligent Robots and Systems*, august 2005.

[25] S. J. Julier and J. K. Uhlmann. The scaled unscented transformation. In *Proceedings of the IEEE American Control Conference*, pages 4555–4559, Anchorage, USA, May 2002.

[26] R. van der Merwe and E. Wan. The square-root unscented kalman filter for state and parameter-estimation. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, USA, May 2001.

[27] Andreas Huster. *Relative Position Sensing by Fusing Monocular Vision and Inertial Rate Sensors*. PhD thesis, Stanford University, July 2003.