A NOVEL STATE PARAMETRIZATION FOR STEREO-SLAM

Arne Petersen, Reinhard Koch
Multimedia Information Processing, Christian-Albrechts-University, Hermann-Rodewald-Str. 3, D-24118 Kiel, Germany
{*apetersen, rk*}*@mip.informatik.uni-kiel.de*

Keywords: SLAM:SfM:Stereo:Linear Estimation

Abstract: This paper proposes a novel parametrization for probabilistic stereo SLAM algorithms. It is optimized to fulfill the assumption of Gaussian probability distributions for system errors. Moreover it makes full use of the contraints induced by stereo vision and provides a close to linear observation model. Therefore the position and orientation are estimated incremetally. The parametrization of landmarks is chosen as the landmarks projection in the master camera and its disparity to the projection in the slave camera. This way a minimal parametrization is given, that is predestinated for linear probabilistic estimators.

1 INTRODUCTION

The task of visual **S**imultaneous **L**ocalization **A**nd **M**apping (SLAM) is to keep track of the position and orientation, pose for short, and the environment map of a vision system. The environment map mostly consists of so called landmarks that represent parts of the map that can be observed using cameras. Models for the observations of the environment and for most SLAM methods the systems motion are used to estimate this parameters.

After the first SLAM systems have been developed in the mid 1980's, see (Brooks, 1985) and (Crowley, 1989), algorithms for SLAM have been studied intensely, e.g. see (Leonard and Durrant-Whyte, 1992), (Dissanayake et al., 2001), (Paz et al., 2008), (İmre and Berger, 2009). This is due to its applicability to autonomous navigation and automatic environment field mapping. These tasks become increasingly important as automatization using autonomous vehicles is applied to more and more areas. This includes airborne and ground (Lemaire et al., 2007) as well as underwater (Hildebrandt and Kirchner, 2010) vehicles.

Many solutions to the SLAM problem use the **K**alman **F**ilter (KF) introduced by Kalman in 1960 (Kalman, 1960). This estimator is used due to its low computational costs despite its drawback in estimating non linear processes. Thus, the choice of state parametrization is crucial to keep the filter state and reliability (covariance matrix) estimation consistent. It has to be chosen in a way that allows for close to linear system models (prediction and observation) and meets the Gaussian noise assumption made by the KF. Therefore a new parametrization for the parameters to be observed by a stereo vision system is introduced, that uses the stereo correspondences for the respective landmarks and a local navigation frame. The projection of the landmark into the left (master) stereo camera and its disparity to the right (slave) camera represent its position in 3D space. By this means the landmarks are described relative to the actual system pose. For this representation the assumption of zero mean Gaussian noise for its errors is met. The model proposed in this article enables linear least squares estimators like KFs to determine consistent state and reliability information.

Even though stereo camera systems are limited by the stereo baseline in direct depth measurements, their use is helpful for many reasons. The limitation due to the baseline can be reduced by increasing image resolution and using models relating stereo images over time (see sec. 3.2). Moreover monocular reconstructions have the drawback of a global scale ambiguity, see (Hartley and Zisserman, 2003). This ambiguity also causes a drift in the scale due to error accumulation. Since the reconstruction using stereo is metric, sensors like accelerometers can be incorporated without estimating the reconstruction scale.

Beside synthetic test experiments are made using a dataset provided by the Rawseeds Project (www.rawseeds.org, see (Ceriani et al., 2009) and (Bonarini et al., 2006)). It delivers the sensor data (stereo camera images and IMU data) from a wheel driven robot system navigating in a plane. Moreover ground truth information is available for the whole

covered trajectory. This way the pose estimation along with its consistency to the estimated covariance can be verified. As will be shown, the proposed method is able to reliably estimate both entities consistently even for large datasets.

Following this introduction previous work in the field visual SLAM is presented and some prerequisites are given. After the description of the used models an comparison with state of the art methods is done and some experiments are discussed.

2 PREVIOUS WORK

Various kinds of estimators have been designed to solve the task of SLAM. These range from direct estimation of quadrifocal tensors (Hildebrandt and Kirchner, 2010) and particle filters (İmre and Berger, 2009) to extended Kalman Filters (Schleicher et al., 2007). Nevertheless the most commonly used estimators are based on the famous KF. This is due to the fact that SLAM usually requires real time performance and the simple integration of multiple sensors, allowing for compensating the errors induced by the different sensors types. The ability to not only estimate the system state but also information on the estimations reliability is another reason for its popularity.

As stated the KF is error-prone when non linear models are used. Because of this the use of more advanced versions of the KF have been studied. The advantages and drawbacks of these filters have been analyzed in (Lefebvre et al., 2004). Their applications to SLAM include beside others the unscented Kalman Filter (Sünderhauf et al., 2007) as well as iterated sigma point Kalman Filters (Song et al., 2011). Such approaches try to overcome the drawbacks of non linear models by using estimators more robust in such cases. Thus the non linearities themselves are not eliminated but more complex algorithms are used. This results in increased computational costs and thus less applicability to real time demands.

Apart from the used estimators the solutions to the SLAM problem differ in the representation of landmarks. For visual SLAM the so called **I**nverse **D**epth (**ID**) representation is used mostly, e.g. see (Civera et al., 2008),(Sünderhauf et al., 2007). There a landmark is represented by the position of the camera from where the point was observed first, the direction to the 3D point and its inverse distance to the camera center. A drawback of this parametrization is, that is uses a 6 dimensional representation. Since a landmark (point in 3D-space) only has 3 degrees of freedom (DOF), it is over parametrized. Thus, ambiguities in the estimation can arise, since points initialized from the same camera end up with different camera positions. (Civera et al., 2008) have shown, that in ID the error propagation for depth is linear under certain assumptions. Nevertheless the models include inverse tangent and normalization functions resulting in non linearities not analyzed by Civera et al..

Since the ID parametrization for landmarks suffers from inconsistent covariance estimation for short distances, (Paz et al., 2008) discuss a combination of representations. For their stereo SLAM system they partition landmarks in far away and nearby landmarks. For far points ID is used, and the 3D **E**uclidean **S**pace (**ES**) for nearby points. This improves the estimation of landmarks and the system pose, but as can be seen in their evaluation, the parametrization still has problems covering the true error distribution for reconstructed 3D points.

Due to the above mentioned benefits provided by stereo camera systems, many stereo SLAM systems have been developed, see (Paz et al., 2008), (Hildebrandt and Kirchner, 2010), (Schleicher et al., 2007). The latter propose a system solely based on wide angle stereo cameras and a two stage algorithm. They create small local maps and detect close loops using SIFT-fingerprints. In (Hildebrandt and Kirchner, 2010) a system is discussed, that uses an IMU aided estimation of the quadrifocal tensor.

In (Solá et al., 2007) the authors claim, that using two independent cameras outperforms stereo rigs. In fact, they propose a combination of monocular and stereo vision resulting in improved estimation. Since this is a combination of both models, this approach can be used to fuse most monocular and stereo models. In (Herath et al., 2006) a representation similar to the one presented here, has been proposed for observation models. In contrast to our work they do not use this for modeling landmarks in the SLAM system. It was shown, that such observation models undergo a Gaussian noise assumption.

Our Contribution: The main contribution of this paper is to introduce a novel parametrization for landmarks optimized for stereo SLAM algorithms using image **P**oints and **D**isparities (**PD**). This parametrization allows a minimal representation of a 3D landmark, that is, 3 parameters are used for the 3 DOF. On the one hand this parametrization outperforms the one in Euclidean 3-space by matching the Gaussian noise distribution much better. On the other hand compared to inverse depth it reduces the number of variables to be estimated from 6 to 3 per landmark. Moreover the resulting observation model is kept simple and the landmarks can be initialized accurately in position and uncertainty.

Because of the reduced parameter size and the simple observation model compared to inverse depth the computational costs are decreased for SLAM systems. Thanks to the new model the equations used for estimation can be accurately linearized using their analytical derivatives. This again reduces the computational effort. As will be shown, the proposed parametrization is capable of improving the estimated poses covariance matrix. Thus, in contrast to ID parametrization the computed global position and orientation are consistent with the estimated variances. Hence, the proposed parametrization outperforms the ID by improving consistency and computational effort.

3 PREREQUISITES

In this section the notation used throughout this paper is introduced and some prerequisites needed for the proposed algorithms are given.

3.1 Iterated Kalman Filters

The description of the iterated extended KF (IEKF) used in this paper is based on the EKF in (Welch and Bishop, 2006). Since a detailed description of the KF is out of the scope of this paper we refer to them. For simplicity the notation introduced there will be used to augment the proposed extended KF by iterative linearization. Since the prediction is done as in the EKF only the update step is described here.

An additional index is introduced to the Jacobian H_k of the used measurement models, denoted as h. $H_k^{(\nu)}$ refers to the ν^{th} iteration step and the respective state estimation $p_k^{(\nu)}$, that is:

$$H_k^{(\nu)} = \left.\frac{\partial h(p)}{\partial p}\right|_{p=p_k^{(\nu)}} \tag{1}$$

Let l_k the observation for the k^{th} time step and P_{l_k} its covariance matrix. Using the state prior as iteration start $p_k^{(0)}$ and the respective covariance matrix P_k^- the ν^{th} iteration is defined by:

$$p_k^{(\nu)} = p_k^{(0)} + K_k^{(\nu-1)} \cdot z_k^{(\nu-1)} \tag{2}$$

$$K_k^{(\nu)} = P_k^- H_k^{(\nu)} \left(H_k^{(\nu)} P_k^- H_k^{(\nu)^T} + P_{l_k} \right)^{-1} \tag{3}$$

$$z_k^{(\nu)} = l_k - h\left(p_k^{(\nu)}\right) - H_k^{(\nu)} \cdot \left(p_k^{(0)} - p_k^{(\nu)} \right) \tag{4}$$

The iteration process is continued until ν reaches a certain maximum or $\left\| p_k^{(\nu)} - p_k^{(\nu-1)} \right\|_2$ falls beneath a given threshold. The covariance matrix P_k^+ for the finally estimated state is afterwards computed by:

$$P_k^+ = \left(\mathbf{I} - K_k^{(\nu)} H_k^{(\nu)} \right) P_k^- \tag{5}$$

This equation can be derived directly using the derivation for the Gauß-Markov-Models and the Kalman Filter in (McGlone et al., 2004), chapter 2.2.4. Using this update equations the error incorporated by non linear observation models can be minimized. However, this applies only to models where all non linearities of the system are observed at the same time. In this case the IEKF outperforms even the unscented or sigma point Kalman Filters (see (Lefebvre et al., 2004) for details).

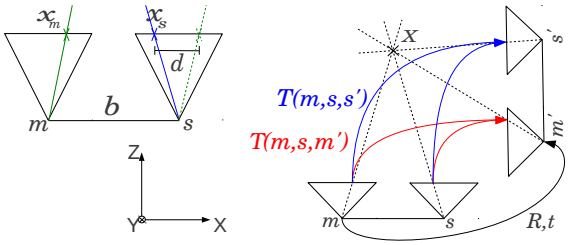

Figure 1: Hardware setup: master and slave camera (m,s) and transformed system (m',s'). TFT T and transformation R,t

3.2 Stereo Vision and Trifocal Tensor

The stereo camera system is made up of 2 cameras, the master camera m and the slave camera s, differing only in a translation $t_{ms} = \begin{pmatrix} b & 0 & 0 \end{pmatrix}^T$ with baseline b. Projecting a 3D point X into both cameras, 2 points in the respective image planes result x_m and x_s. Using $x(i)$ to refer to the i^{th} component of x the disparity is $d = x_m(1) - x_s(1)$. Since $b > 0$ and all observed points are in front of the cameras, $d > 0$ holds for perfect correspondences x_m and x_s. In projective geometry $d = 0$ holds for points at infinity. These entities and the used coordinate system are visualized in figure 1.

Let P_1, P_2 and P_3 the projections for 3 different cameras. Projecting a 3D-point X into the image planes of the 3 cameras results in corresponding 2D-points x_1, x_2 and x_3. The **Tri**-**F**ocal **T**ensor (TFT, see (Hartley and Zisserman, 2003)) T can be used to map 2 corresponding points to the third in the respective camera, that is:

$$T\left(P_i, P_j, P_k, x_i, x_j\right) = x_k \tag{6}$$

for all permutations of i, j, k, see figure 1. This is especially useful when dealing with stereo camera systems. It can be used for estimation of transformations between stereo rigs without triangulation of X. When

the system m,s is transformed to m',s' the two TFTs can be computed. Since the transformations $m \rightarrow s$ and $m' \rightarrow s'$ are known, only one transformation constrained by both TFTs has to be estimated.

4 SLAM USING PD

Since the task of SLAM integrates map building, the landmarks have to be modeled as parameters for estimation. As stated in section 2 mainly two representations of 3-space points are used. The most simple uses the 3 parameters of Euclidean 3-space (ES) and the other one is ID, having 6 DOF. The former uses a minimal set of parameters but hardly fulfills the assumption of Gaussian noise. The latter is over parametrized but has proven to fit a normal distribution for its depth errors under certain assumptions. A drawback for ID is the non linearity caused by using spherical coordinates for ray parametrization.

4.1 PD Parametrization

The landmark parametrization proposed in this paper uses a minimal set of parameters, relates the image measurements to the system state nearly linear and undergoes a Gaussian error distribution.

The state to be estimated includes the systems pose as well as its landmarks. Therefore a representation for landmarks and poses has to be used, well suited to the needs of linear estimators.

The PD (point-disparity) representation proposed in this article codes a position in 3-space as its projection to the stereo system. Since the stereo calibration is assumed to be known, the projection to the master camera x_m and its 1-dimensional displacement d to the projection to the slave camera (see figure 1) is sufficient. This results in a representation

$$\mathcal{X}_k = \begin{pmatrix} x_{m,k} \\ d_k \end{pmatrix} \tag{7}$$

using 3 DOF, which is minimal for a point in 3-space. Here the index k is needed since the representation is relative to the actual pose and thus, dependent on the time step k.

The parametrization used for estimation of the pose π_k consists of the 3-space position increment t_k of the master camera and the orientation increment $\Delta\phi_k$ in Euler angles relative to the last estimated pose. Thus, the pose describes the transformation between the past and actual pose in the global coordinate frame. For the ES and ID estimation the pose is given relative to the global coordinate system defined by the initial camera pose. Thus, for N landmarks $\mathcal{X}^1,\ldots,\mathcal{X}^N$ in the respective representation ES, ID and PD the parameter vector to be estimated is:

$$p_k = \begin{pmatrix} \pi_k \\ \mathcal{X}_k^1 \\ \vdots \\ \mathcal{X}_k^N \end{pmatrix} \tag{8}$$

For the representation in ES both parts are independently represented in the same global reference system. By this, the observation model relates the observations to the landmarks and the actual global pose directly. For the ID the poses of past time-steps are used to represent landmarks. Thus, the observations relate to landmarks, the actual and the past poses. The PD representation relates the observations to the difference of the actual and the past pose as well as the landmarks. Comparing this relations, the estimation using ES and PD representations includes parameters of actual poses only, leaving the estimation steps independent over time. For the ID representation actual and past poses are estimated. Moreover, in PD each estimation step relates only the actual and last pose. That is, it is incremental and correlations in time (i.e. drift) don't contradict the unbiasedness of KF models.

4.2 Observation Model

In the following the time index k is omitted for better readability. For observing the PD landmarks, stereo correspondences are used. That is a landmark $\mathcal{X}^i$ is mapped to 2 image points y_m^i in the master camera and y_s^i in the slave camera respectively. The landmark describes the respective stereo correspondence relative to the previously estimated pose. Using the trifocal tensor, y_m^i and y_s^i can be predicted using the initial estimate p^0, possibly including a guess for the systems movement as in KF prediction. This is implemented in the measurement models h_m and h_s. Thus, the observations and the measurement model are:

$$l = \begin{pmatrix} y_m^i \\ y_s^i \end{pmatrix}_i \quad \text{and} \quad h(p) = \begin{pmatrix} h_m\left(\pi, \mathcal{X}^i\right) \\ h_s\left(\pi, \mathcal{X}^i\right) \end{pmatrix}_i \tag{9}$$

Let R the rotation matrix for the differential orientation $\Delta\phi$. Further let R_g be the rotation matrix corresponding to the global orientation (integrated over all preceding R) for the initial parameters guess p^0. Using this notation the landmarks projection can be predicted in homogeneous coordinates via the TFT, see (Hartley and Zisserman, 2003):

$$\mathcal{Y}_m^i = R^T \left(b \begin{pmatrix} x_m^i \\ 1 \end{pmatrix} - d^i R_g^T t \right) \tag{10}$$

for the master camera. The disparity is predicted using $b \cdot d_i \cdot \left(\mathcal{Y}_m^i(3)\right)^{-1}$. Thus, the prediction for the cor-

respondence i ends up with:

$$h_m(p) = \frac{1}{\mathcal{Y}^i_m(3)} \begin{pmatrix} \mathcal{Y}^i_m(1) \\ \mathcal{Y}^i_m(2) \end{pmatrix} \tag{11}$$

and

$$h_s(p) = \frac{1}{\mathcal{Y}^i_m(3)} \begin{pmatrix} \mathcal{Y}^i_m(1) - b \cdot d_i \\ \mathcal{Y}^i_m(2) \end{pmatrix} \tag{12}$$

Note that this model and the landmark parametrization allow disparities of 0. This way, the proposed models can be easily extended to include points at infinity to improve orientation estimation. The simplicity of this observation model enables the determination of its analytical derivatives. Thus, a precise linearization can be done and there is no need for a time consuming numerical computation of the Jacobians required for linear estimation. Compared to the models used for projecting points represented in ID (see (Civera et al., 2008)) this is an advancement. For them the computation of the view ray $b\left(x^i_m(1)\ x^i_m(2)\ 1\right)^T$ is replaced by a conversion from spherical to Euclidean coordinates.

4.3 Initialization and Estimation

When a landmark is observed first as y_m and y_s, it has to be incorporated into the system state. For the PD representation this can be achieved by:

$$x_m(1) = y_m(1) \quad d = y_m(1) - y_s(1)$$

$$x_m(2) = \frac{1}{2}\left(y_m(2) + y_s(2)\right)$$

In contrast to the initialization of ID-landmarks (conversion from Euclidean to spherical coordinates and depth determination by vector norm) this is strictly linear, so that the assumed measurement noise can be directly propagated to the landmarks covariance. As for most pose estimators the systems initial position and orientation are used as the world coordinate frame.

To estimate the pose for time k the IEKF from section 3.1 is applied using the given observation model. After the actual pose has been estimated, the pose and landmarks have to be transferred to the new camera body frame. Therefore the landmarks are mapped to the new camera view using the same model as for the observation. That is, the projections to the new master camera and the respective disparities are predicted. Afterwards the position and orientation increments are added to the global pose and set to 0. According to this, the parameters covariance matrix is computed using linear error propagation.

The global poses are computed by integrating the position and orientation increments estimated over time. This also applies to the estimated covariances. This way the estimated entities correspond to a random walk process. By this, the standard deviation bounds the global poses even though a drift is incorporated. Another advantage of this incremental estimation is, that loop closes, applied to sub tracks of the estimated trajectory, directly effect new poses. Building the integral over the state sequence also updates subsequent poses. Thus, partial adjustments improve the complete trajectory estimation.

The global landmarks are computed using their local representation and the global poses. For each view a 3-space point is created and these are averaged over all views. Since their representation is as seen from the respective pose, this can be done incrementally. This way the information on this landmark from all views is exploited. For advanced landmark modeling particle casts can be performed to create empirical probability distributions for all landmarks and views. Therefore the estimated covariance matrices for the landmarks and poses can be used to generate the respective particle cloud.

5 PERFORMANCE ANALYSIS

In this section analysis and comparison of the proposed and state of the art methods are carried out. Therefore the landmarks parametrizations for ES, ID and PD and the estimations qualities are compared.

For the evaluation of landmark initialization two different tests have been performed. At first the landmarks represented in ES, ID and PD respectively are generated using a particle cast from stereo correspondences. Therefore, the ground truth correspondences are disturbed by Gaussian noise. The resulting parameter clouds represent the parametrizations empirical probability distributions, when initialized using stereo features. Thereby the assumption of Gaussian noise for the respective representation can be validated when used for stereo vision systems. In figure 2 the accordant distributions are visualized. As can be seen, the ID and PD undergo a Gaussian error distribution. In contrast, ES obviously violates this assumption for large distances. For close landmarks all three parametrizations fit the Gaussian noise quite well.

For the second test of initialization an ID and a PD landmark are generated from a perfect stereo correspondence using the given initialization methods. Afterwards noise with covariance C_{2D} is added to the stereo correspondences and a 3-space cloud is generated via particle cast. Following this, covariances C_{ID} and C_{PD} are computed from C_{2D} using the Jacobians of the initialization functions. Normal dis-

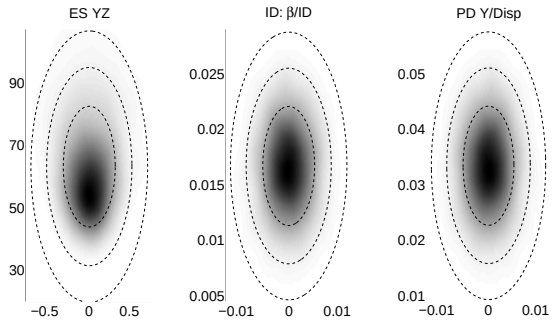

Figure 2: Empirical probability distributions and propagated covariance ellipses in parameter space. Units, ES: baselines ID: radiants,fractions of baselines PD: normalized pixels

tributed noise is added to the ground truth landmarks in ID/PD domain according to C_{ID} and C_{PD}. Finally the 3-space, ID-space and PD-space clouds are reprojected into the image and their distributions are compared. As in the former evaluation the distribution generated by stereo correspondences is used as ground truth for validation. As can be seen in figure 3, the deviation between the true error distribution and the one for ID is much higher compared to the one of PD, which is negligible. Moreover, it increases strongly with decreasing point-camera distance. A similar effect was observed by (Paz et al., 2008) for the reconstruction of 3-space points from ID. This is due to the violation of the assumption made in (Civera et al., 2008), that the view angle between the first observations (i.e. initial and second) of the ID feature have to be small. From this it can be concluded, that the measurement model for PD yields a more consistent update prediction, i.e. in case of close features.

A drawback of the ID parametrization is that the poses position is used as the landmarks reference position. This way a correlation equaling 1 is introduced (see figure 4, components 1-3 and 7-9 respectively). Because of this the systems covariance matrix is singular after initialization. This remains when no prediction noise is added for subsequent estimations. Moreover the covariance will be closer to singular the less noise is added. On the one hand, this leaves the ID unapplicable for general least squares estimators and according statistical analysis, as are described in (Petersen and Koch, 2010). On the other hand, the matrix might loose positive semidefiniteness (required for covariance matrices) due to numerical issues. Because the PD landmarks initialization is uncorrelated with the system pose, no $1:1$ correlations occur.

To verify the consistency and quality of pose esti-

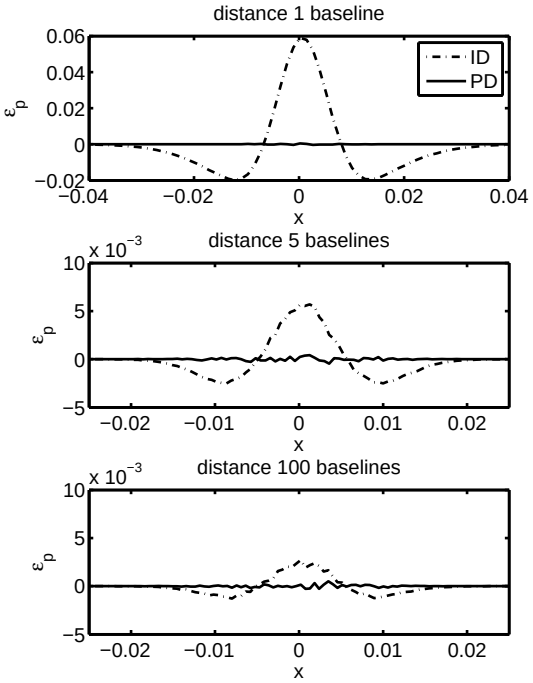

Figure 3: Difference between empirical and ground truth probability distributions for reprojection with different point-camera distances. X-axis marks normalized pixel coordinates and y-axis the error of probability ε_p

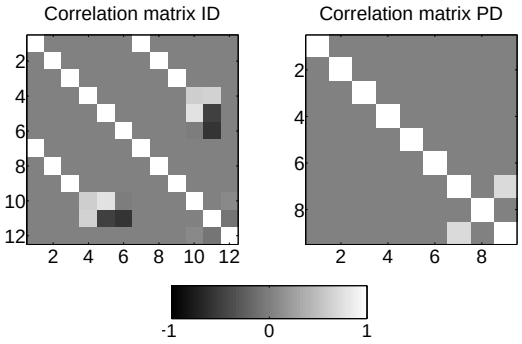

Figure 4: Initial ID/PD correlation matrices for pose and a single landmark. Left ID: component 1-6 pose, 7-12 landmark (7-9 pos., 10-12 ray/idepth). Right PD: 1-6 pose, 7-9 landmark (7,8 master pixel, 9 disparity)

mation, different trajectories have been estimated using synthetical data. Therefore position and orientation increments are generated and summed up to global poses. For every trajectory step an estimation was performed. For the pose prediction the ground truth increments were used disturbed with Gaussian noise according to the assumed prediction uncertainty

		position					orientation				
		RMSE	mean	1σ [%]	2σ [%]	3σ [%]	RMSE	mean	1σ [%]	2σ [%]	3σ [%]
ES	x	1.49	1.18	4.8	8.5	12	1.6	1.26	6.25	11.5	16.6
	y	1.28	1.02	5.7	10.6	14.8	1.64	1.29	6.17	11.4	16.4
	z	1.54	1.23	5.1	9.1	12.7	1.6	1.24	6.93	12.6	18
ID	x	0.88	0.71	37.2	65.3	82.9	0.9	0.72	23.6	43.6	58.8
	y	0.9	0.72	37.2	65.5	82	0.93	0.74	22.1	41.4	56.8
	z	0.94	0.76	39.2	66.2	82.3	0.93	0.75	19.3	36.2	49.7
PD	x	1.13	0.9	62.4	88.9	97	1.2	0.94	83.5	99.3	∼100
	y	1.06	0.84	65.2	91.2	98.1	1.25	0.99	82.1	99.3	∼100
	z	1.27	1.02	58.7	86.8	96.1	1.2	0.94	85.6	99.5	∼100

Table 1: RMS, absolute error (baselines/deg.) and $[1,2,3]\,\sigma$ inliers in %. Averaged over 100 randomly generated trajectories with 1000 estimations each. Avg. landmarks 35, translation 3.5 baselines, orientation change 14.5 degrees per time step. Stdev prediction noise 0.7 baselines, 3 degrees. Stdev observation noise 0.005 normalized pixels ($\sim$ 3 px for 500^2 px images)

(prediction of P_k^- in equation 5). When initializing landmarks from stereo correspondences disparities/depths smaller zero might occur (points behind camera). This can be circumvented by omiting points with disparity smaller three times the observations pixel noise. This is done to keep the analysis comparable for all systems. For error evaluation, the results have been averaged over 100 test runs.

As can be seen in table 1 the root mean square (RMSE) and mean absolute error for PD is slightly increased compared to ID. In contrast, the ES parametrization obviously suffers from imprecise landmark estimation for long term applications. Another advantage of ID and PD estimation compared to ES is, that they are much less biased. This was concluded from the fact, that estimating a trajectory parallel to a planar point cloud results in a strong bias along the direction to the cloud for ES. The same results apply for the RMSE and mean error of landmarks after the final estimation. As stated in section 4.3, the PD landmarks estimation is thought to be averaged over time using particle casts. In this case the error can be reduced by 66% compared to ID at the cost of increased computational effort.

In table 1 the inliers for the 1,2 and 3 σ-confidence interval are listed. An inlier is defined by being inside a given confidence interval. That is, for the 1σ, 2σ and 3σ confidence area the expected percentage of inliers is 69%, 95% and >99% respectively. As can be seen, the ES and ID estimation are inconsistent. In contrast to this, the PD estimation stays consistent for the orientation even for long term runs, in this case 1000 estimations. The PDs position variance is improved considerably but still shows small inconsistencies of a few percent. This is due to the time correlations of position increments being much higher compared to orientation. To compensate this, the consistency method for joint covariances discussed in (Uhlmann, 2003) can be applied by multiplying the cumulative covariance sum for positions with a factor of 2. This improves the result as shown in table 2. Note that this method is applied to the incremental sum of global positions covariances only and does not effect the estimation process in any way.

1σ [%]		2σ [%]		3σ [%]	
woJC	wJC	woJC	wJC	woJC	wJC
62.4	77.8	88.9	96.3	97	99.2
65.2	77.6	91.2	95.7	98.1	98.6
58.7	78.4	86.8	96.4	96.1	99.2

Table 2: $[1,2,3]\,\sigma$ inliers in % for PD position estimation, with (wJC) and without (woJC) joint covariance consistency.

In figure 5 the estimation errors and the estimated uncertainty regions for an exemplary trajectory are visualized. The uncertainty regions are defined by the 95% confidence interval. As can be seen the error for the estimated position using PD bounds the true trajectory error. In contrast the ID becomes inconsistent with increasing time. The ES estimation is highly inconsistent, even after a few estimations.

Using an average of 35 landmarks the average runtime is 53.8$[ms]$, 87.5$[ms]$ and 69.8$[ms]$ per estimation for ES, ID and PD respectively. The tests were carried out on a desktop pc (Intel(R) Core(TM) i7 @ 2.67GHz) using a interpreter simulation software (MathWorks MatLab(R)) without exploiting sparse matrix structures. That is, the computational effort for all estimators can be reduced significantly by using efficient programming and sparse matrices.

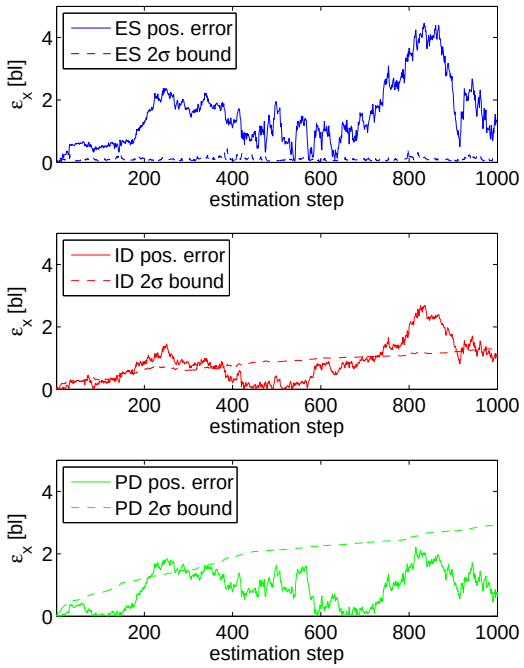

Figure 5: Estimation error ε_x for x-position in baselines and 2σ bounds for ES, ID and PD respectively.

	$\|\varepsilon_x\|$	$\|\varepsilon_z\|$	$\|\varepsilon_h\|$
estimated std.	0.56	0.75	1.65
empirical std.	0.52	0.48	0.74
mean abs. error	0.79	0.81	1.26

Table 3: First and second row: means of standard deviations for trajectory error in x- and z-position ε_x,ε_z in [m] and heading ε_h in [deg], estimated using our method and computed empirically from trajectory error. Third row: mean of absolute errors

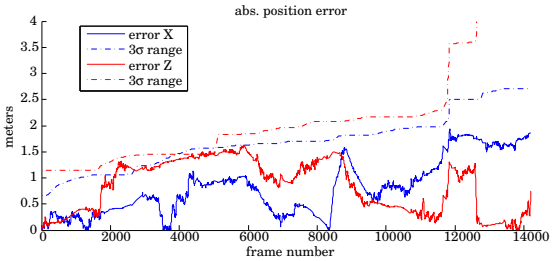

Figure 6: Absolute position error and 3σ range without adjustment

6 Experiments

In this section some experiments performed on the dataset provided by the Rawseeds Project (see sec. 1) are presented. The complete dataset includes 26000 images. For this evaluation only a subset is used containing 14200 images and a traveled distance of 376 meters. Ground truth information is available only for the movement in the X/Z-plane and the systems heading. Thus, only a 3 DOF pose is estimated here. In addition to the pose parameters the systems velocity is included in the estimation to model system motion. Data provided by a low cost IMU is used for predicting the velocity and orientation.

In contrast to section 5 only points with disparities ≤ 0 are rejected. If such disparities occur during state updates, the respective point is excluded from the state and the iteration process is continued. This has proven to be a feasible exception handling and during all tests the estimators convergence was not effected. Moreover, because the uncertainty of disparities rapidly decreases due to camera translation and stereo constraints this has to be applied very seldom (about 30 times for 14200 estimation steps).

For the global poses standard deviations the square root of summed variances of the incremental poses are used (see sec. 4.3). The absolute trajectory errors and their estimated standard deviations are visualized in figures 6 and 7 respectively. Table 3 lists the standard deviations and means of these errors. In addition the tables first row shows the means of the standard deviations estimated by the KF.

As can be seen the position and orientation estimates are consistent with the estimated covariance. Moreover the estimation is informative, that is, the standard deviations are not estimated as too large. The trajectory starts with a heading of 0 (z-axis direction) and no movement for some frames. Thus, the uncertainty of the velocity only effects the positions z-component so its variance is higher than in x-direction. The same effect can be observed at im-

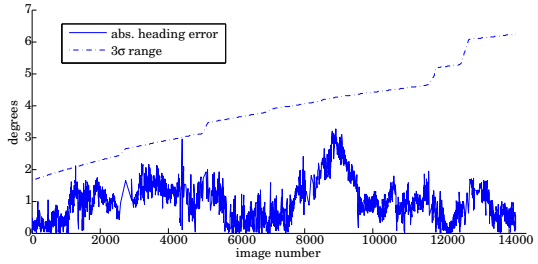

Figure 7: Absolute heading error and 3σ range without adjustment

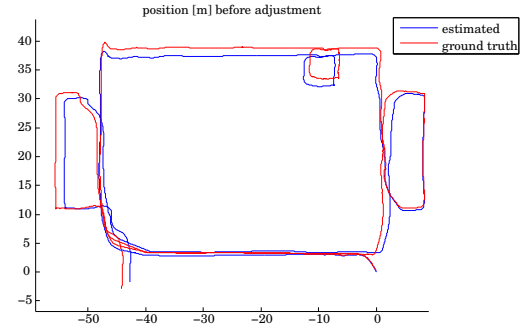

Figure 8: Estimated trajectory in [m] without adjustment

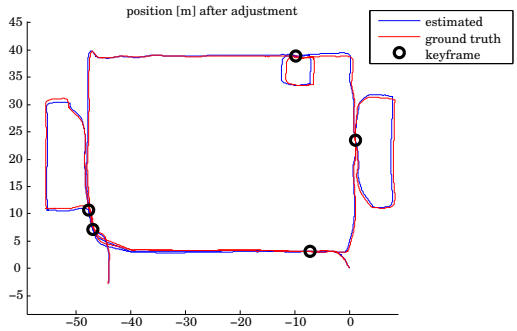

Figure 9: Estimated trajectory in [m] after adjustment. Black circles mark loop close positions.

age index 12000. Here the system stops in front of a white wall, thus, no visual features are present for some frames.

The global trajectory estimate is visualized in figure 8. The trajectory estimate shows the reliability of pose estimation even without adjustment. This proves, that an automatic detection of closed loops could be applied by matching key frames for close-by positions estimates. In figure 9 an adjusted estimation is shown, using hand selected key points and a constraint to poses only. That is, no bundle adjustment was performed, but the poses increments inside a loop are constrained to sum up to 0. Because the velocity in the systems body frame was constraint to be consistent with the position increment, the velocity and heading were also improved.

For the used dataset the system performs at about 6-7 fps using a desktop PC with an average of 16.5 landmarks. Since the software is programmed using a rapid prototyping framework including visualization of landmarks and without optimizations this can be improved. A great deal of time (about 30%) is spent with the generation of KLT-features because a simple CPU implementation is used. Moreover taking advantage of the sparse Jacobian matrices would speed up the system. Keeping this in mind we are confident that a frame rate of 20 to 30 fps can be achieved.

The system models are applicable to full 6 DOF pose estimation also, as introduced in chapter 4. Experiments on the same dataset apparently yield good results. Especially the height (Y-axis) was estimated reliably. The final deviation for the complete trajectory was about 2 to 3 meters (close to 1% of trajectory length), assuming start and end point have the same height. This is notable, since IMUs are known to be highly unstable predicting the height (parallel to gravity acceleration). Since ground truth information is given for 3 DOF only, these results are not discussed further here.

7 CONCLUSIONS

We have introduced a novel parametrization for stereo SLAM systems. Because of optimal exploitation of stereo constraints the observation models are nearly linear and the parametrization is minimal. Moreover it meets the assumption of Gaussian noise, such that it is predestinated for application in linear estimators like kalman filters.

It was proven in synthetical tests and real-world experiments, that the estimation is precise and consistent even for long term estimation. The synthetical tests showed that for stereo systems the PD outperforms the ID in terms of computational effort and covariance estimation, i.e. the consistency is preserved. Consistent variance estimation is a major advancement for navigation, since it improves long term stability, reliability information and the quality of global adjustments. Moreover, due to the reduction in computational costs for PD, much more landmarks can be used to improve the estimations accuracy. The experiments using the RawSeed data showed the applicability of the proposed methods to actual SLAM problems.

For future work it is planed to introduce points at infinity as was done for ID. These are characterized by a disparity equal to 0. Thus, they can be modeled using the same observation model (see equation 10) and parametrization, by omitting the disparity component (fixed to 0). This way the points, rejected due to possibly negative disparities, can be exploited to improve orientation estimation. Another task is to generalize the modeling to monocular SLAM systems, for example by using epipolar lines for offset (disparity) representation.

REFERENCES

Bonarini, A., Burgard, W., Fontana, G., Matteucci, M., Sorrenti, D. G., and Tardos, J. D. (2006). Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In *proceedings of IROS'06 Workshop on Benchmarks in Robotics Research*.

Brooks, R. (1985). Visual map making for a mobile robot. In *Proceedings IEEE International Conference on Robotics and Automation 1985*, volume 2, pages 824 – 829.

Ceriani, S., Fontana, G., Giusti, A., Marzorati, D., Matteucci, M., Migliore, D., Rizzi, D., Sorrenti, D. G., and Taddei, P. (2009). Rawseeds ground truth collection systems for indoor self-localization and mapping. *Autonomous Robots*, 27(4):353–371.

Civera, J., Davison, A. J., and Montiel, J. M. M. (2008). Inverse depth parametrization for monocular slam. *Autonomous Robots*, 24(5):932–945.

Crowley, J. L. (1989). World modeling and position estimation for a mobile robot using ultrasonic ranging. In *IEEE Conference on Robotics and Automation*, pages 1574–1579.

Dissanayake, M. W. M. G., Newman, P., Clark, S., Durrant-Whyte, H. F., and Csorba, M. (2001). A solution to the simultaneous localization and map building (slam) problem. *Transactions on Robotics and Automation*, 17(3):229–241.

Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry*. Cambirdge University Press, 2 edition.

Herath, D. C., Kodagoda, K. R. S., and Dissanayake, G. (2006). Modeling errors in small baseline stereo for slam. Technical report, ARC Centre of Excellence in Autonomous Systems (CAS).

Hildebrandt, M. and Kirchner, F. (2010). Imu-aided stereo visual odometry for ground-tracking auv applications. Technical report, Underwater Robotics Department, DFKI RIC Bremen.

İmre, E. and Berger, M.-O. (2009). A 3-component inverse depth parameterization for particle filter slam. In *Lecture Notes in Computer Science*, volume 5748/2009, pages 1–10. Springer.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82.

Lefebvre, T., Bruyninckx, H., and De Schutter, J. (2004). Kalman filters for non-linear systems: a comparison of performance. *International Journal of Control*, 77(7).

Lemaire, T., Berger, C., Jung, I.-K., and Lacroix, S. (2007). Vision-based slam: Stereo and monocular approaches. *International Journal of Computer Vision*, 73(3):343–364.

Leonard, J. and Durrant-Whyte, H. (1992). *Directed Sonar Sensing for Mobile Robot Navigation*. Springer.

McGlone, J. C., Förstner, W., and Wrobel, B. (2004). *Manual of Photogrammetry*, chapter 2 Mathematical Concepts in Photogrammetry. asprs.

Paz, L. M., Piniés, P., Tardós, J. D., and Neira, J. (2008). Large-scale 6-dof slam with stereo-in-hand. *Transactions on Robotics*, 24(5).

Petersen, A. and Koch, R. (2010). Statistical analysis of kalman filters by conversion to gauss-helmert models with applications to process-noise estimation. In *Proceedings of ICPR2010*, Istanbul, Turkey.

Schleicher, D., Bergasa, L. M., Barea, R., López, E., Ocana, M., Nuevo, J., and Fernández, P. (2007). Real-time stereo visual slam in large-scale environments based on sift fingerprints. Technical report, Department of Electronics, University of Alcalá.

Solá, J., Monin, A., and Devy, M. (2007). Bicamslam: Two times mono is more than stereo. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 4795–4800, Rome, Italy.

Song, Y., Song, Y., and Li, Q. (2011). Robust iterated sigma point fastslam algorithm for mobile robot simultaneous localization and mapping. *Chinese Journal of Mechanical Engineering*, 24.

Sünderhauf, N., Lange, S., and Protzel, P. (2007). Using the unscented kalman filter in mono-slam with inverse depth parametrization for autonomous airship control. In *Proceedings of IEEE International Workshop on Safety Security and Rescue Robotics*.

Uhlmann, J. K. (2003). Covariance consistency methods for fault-tolerant distributed data fusion. *Information Fusion*, 4(3):201 – 215.

Welch, G. and Bishop, G. (2006). An introduction to the kalman filter. Technical report, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.