Stereo Vision-based Simultaneous Localization and Mapping with Ranging Aid

Young-Hee Lee*, Chen Zhu**
*Institute for Communications and Navigation
Technical University Munich
Munich, Germany
younghee.lee@tum.de, chen.zhu@dlr.de

Gabriele Giorgi**, Christoph Guenther*,**
**Institute of Communications and Navigation
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
gabriele.giorgi@dlr.de, christoph.guenther@dlr.de

Abstract—We propose fusion of stereo visual odometry and ranging for Simultaneous Localization And Mapping (SLAM). Two basic processes of feature- and keyframe-based stereo visual odometry (Tracking and Local mapping) are operated in the algorithm, while saving all local keyframes, map points, and visual constraints in a global map database, as well as the available ranging constraints of the keyframes. At the end of the sequence, state estimates by visual odometry are fused with ranging measurements to mitigate the inherent accumulating errors in the process and achieve global consistency. We formulate a simple graphical representation for the fusion, and perform least squares estimation with the sparse Levenberg-Marquardt algorithm to minimize the summation of the re-projection and distance squared errors over all the defined constraints in the global graph. The proposed algorithm is evaluated both qualitatively and quantitatively on a real stereo image dataset with synthetically generated distance measurements with superimposed Gaussian white noise. The experimental results show that the proposed SLAM algorithm effectively compensates the cumulative bias in visual odometry. Furthermore, the global accuracy of the trajectory estimation is comparable to the one of stereo vision-only SLAM with closing loops.

Index Terms—stereo vision, sensor fusion, ranging sensor, graph-based SLAM

I. Introduction

Stereo visual odometry is a viable tool for exploring unknown and GNSS-denied environments with unmanned ground vehicles (rovers). Unlike monocular vision, it is possible to determine rover poses and build a map database without scale ambiguity with stereo vision. However, incremental errors are unavoidable since stereo visual odometry is a dead-reckoning process in which the current state is propagated from the previous pose determination.

The loop closing technique [3] is generally applied to mitigate accumulating estimation errors in visual odometry and achieve global consistency. In this technique, matches of the feature points observed in the current scene are continuously searched in a global map database (loop detection), and when matches are detected, an update step automatically refines the state estimates in the loop (loop correction). By applying this method, growing estimation errors in visual odometry can be effectively compensated. However, it is challenging to accurately detect and correct loops without introducing significant computational complexity to the system.

Mur-Artal et al. proposed a loop closing method [4] and implemented it in a keyframe- and feature-based SLAM algorithm (ORB-SLAM) [2]. In this method, loop candidates are searched by comparing Bag of Words [7] and ORB [8] only from the previous keyframes which share covisibility information with the current keyframe. When a loop is detected, the poses of the keyframes which have a strong correlation with the current keyframe are corrected first, followed by a global bundle adjustment. Furthermore, the entire process of closing loops is operated on a separated processing thread, thus not interfering a real-time tracking process. However, in order to operate any loop closing technique, a rover has to revisit previously observed places. This is an intrinsic problem of the loop closing technique which results in constraining the mission planning.

We propose a different approach to resolve the drift problem: fusion of keyframe- and feature-based stereo visual odometry and ranging measurements. We incrementally recover rover's poses by tracking the feature observations in the current frame from a local map. In parallel, the local map is refined on another processing thread, while saving all local keyframes, map points, and visual constraints (feature point observations) in a graphically formulated global map database. Additionally, ranging constraints between the keyframes and a static base station are added to the graph. At the end of the sequence, we define the least squares problem with a cost function given by the summation of re-projection and distance squared errors over all the constraints in the global graph with an appropriate relative weighting between vision and range. Furthermore, the Huber kernel [6] is applied to robustify the estimation process. Ranging constraints are absolute distance measurements, independent from the past states, that enable compensating cumulative bias and more accurately estimating the rover's poses.

We evaluate the proposed algorithm on a public stereo vision dataset (The KITTI Vision Benchmark Suite [12]). Distance measurements are synthetically generated using the ground truth positions provided by the dataset with superimposed Gaussian white noise. First, we qualitatively compare the trajectory and the map point estimates of the proposed algorithm to the ones obtained with stereo visual odometry. Then, we quantitatively compare the local accuracy and global

consistency of the proposed algorithm to the ones of stereo visual odometry and stereo vision-only SLAM with loop closures in terms of the Relative Pose Error (RPE) and Absolute Trajectory Error (ATE) [15]. The evaluation results show that the accumulating errors inherent in the dead reckoning visual odometry are efficiently reduced by ranging fusion. Moreover, the performance of the proposed algorithm is comparable to the one of stereo vision-only SLAM with closing loops.

II. PRELIMINARIES

In this section, we introduce the frame definitions and the notations used in the paper. First of all, we define three reference frames: a) the *World frame* (*W-frame*), coinciding with the initial camera frame, b) the *Camera frame* (*C-frame*), attached to the camera body, and c) the *Beacon frame* (*B-frame*) at the static base station. Fig. 1 shows the relative geometry between the three frames. The 3D camera motion is described by an orthonormal rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$. A generic spatial position vector in the 3D Euclidean space is denoted with $\mathbf{p}$. Additional information on the frames linked by the rigid transformation is added in form of subscripts, whereas superscripts are used to denote the time stamps of the camera poses and map point numbers. For example, the camera translation with respect to the *W-frame* at time t is $_{\mathrm{w}}\mathbf{t}_{\mathrm{c}}^{t}$, and the 3D position of the *i-th* map point with respect to the *C-frame* is $_{\mathrm{c}}\mathbf{p}_{\mathrm{m}}^{i}$.
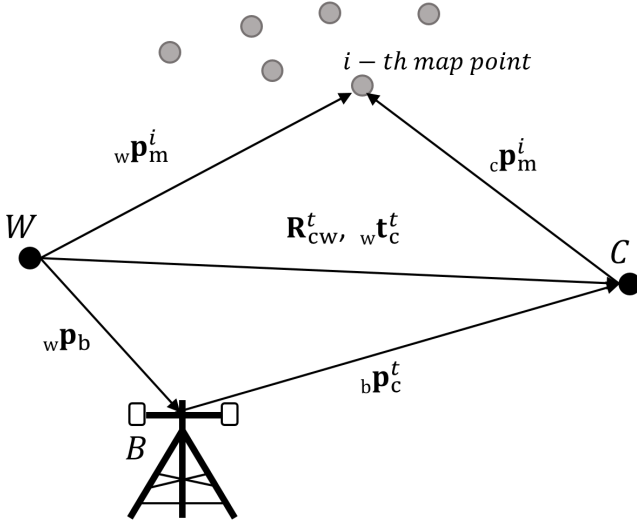

Fig. 1. Definitions of the *World (W)*, the *Camera (C)*, and the *Beacon (B)* frames

III. SYSTEM OVERVIEW

An overview of the proposed SLAM algorithm is given in this section (see Fig. 2). We refer to the state-of-art keyframe- and feature-based algorithm ORB-SLAM [2] for the two basic stereo visual odometry processes: tracking and local map correction. While the poses are incrementally recovered by tracking the features on the current frame in a local map, the map is continuously refined in another thread. All keyframes,

map points, feature observations in the local map, as well as the available ranging measurements of the keyframes are saved in a graphically represented global map database. At the end of the sequence, we use the Levenberg-Marquardt algorithm to compute the keyframe poses and map point positions that minimize the summation of the re-projection and the distance squared errors over all the defined visual and ranging constraints in the graph. We explain the details of the graph-based fusion method in the following section.
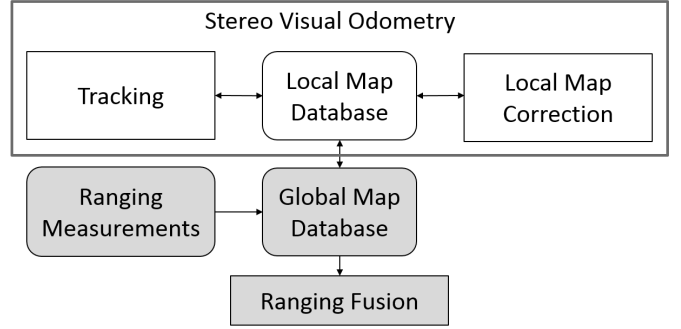

Fig. 2. An overview of the proposed SLAM algorithm

IV. FUSION OF STEREO VISUAL ODOMETRY AND RANGING

In this section, we first present a graphical representation of the global map database in the proposed sensor fusion approach, detailing the state variables and the constraints. Then, the least-squares problem is formulated with a cost function given by the summation over all the constraints in the graph.

A. Graph Representation

We apply a graph-based formulation for fusing stereo visual odometry with ranging measurements as illustrated in Fig. 3. State variables are denoted as $\mathbf{x}$ in circular nodes : those are a) keyframe poses $\mathbf{x}_{\mathrm{c}}$ and b) map point positions $\mathbf{x}_{\mathrm{m}}$ (Table I). The position of the static base station $_{\mathrm{w}}\mathbf{p}_{\mathrm{b}}$ is assumed to be known. Constraints (measurements) are denoted with $\mathbf{z}$ in squares on the edges between nodes: these are a) visual measurements $\mathbf{z}_{\mathrm{v}}$ - feature point observations in keyframes and b) ranging measurements available at keyframes $\mathbf{z}_{\mathrm{r}}$ (Table II). From the graph, we formulate the least squares problem with a multi-objective cost function defined by the summation of the re-projection and distance squared errors over all the edges in the graph as (1) and (2). The state variables and the error functions in (1) and (2) are explained in the following subsections. We use *g2o* [10] to compute the optimal states minimizing the defined cost function (2) with the Levenberg-Marquardt algorithm.

$$\Theta^{*} = \operatorname{argmin} \mathbf{F}(\Theta)$$

$$\text{where } \Theta = \begin{bmatrix} _{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t_0} & \cdots & _{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t_f} & | & _{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{0} & \cdots & _{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{N} \end{bmatrix}^{T}$$
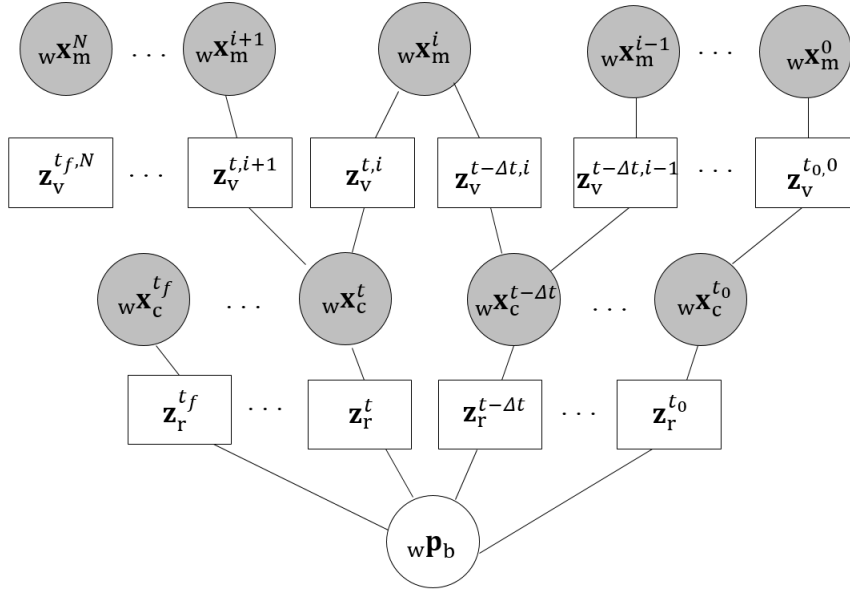
$$(1)$$

Fig. 3. Graphical representation for fusion of stereo visual odometry and ranging measurements

TABLE I
STATE VARIABLES IN THE GRAPH

Variable	Symbol	Domain	Dimension
Camera pose	$_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t}$	$se(3)$	6
Map point position	$_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i}$	$\mathbb{R}^3$	3

TABLE II
CONSTRAINTS IN THE GRAPH

Constraint	Symbol	Domain	Dimension	Error function
Vision	$\mathbf{z}_{\mathrm{v}}^{t,i}$	$\mathbb{R}^3$	3	$\mathbf{e}_{\mathrm{v}}^{t,i}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t},_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i})$ (5)
Range	$\mathbf{z}_{\mathrm{r}}^{t}$	$\mathbb{R}^1$	1	$\mathbf{e}_{\mathrm{r}}^{t}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t})$ (8)

$$\mathbf{F}(\Theta) = \sum_{t,i} \mathbf{e}_{\mathrm{v}}^{t,i}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t},_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i})^T \boldsymbol{\Omega}_{\mathrm{v}}^{t,i} \mathbf{e}_{\mathrm{v}}^{t,i}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t},_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i})$$
$$+ \sum_{t} \mathbf{e}_{\mathrm{r}}^{t}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t})^T \boldsymbol{\Omega}_{\mathrm{r}}^{t} \mathbf{e}_{\mathrm{r}}^{t}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t}) \quad (2)$$

B. State Variables

The camera pose is parameterized in terms of rotation and translation vectors, describing the axis-angle represented orientation and the relative position with respect to the *W-frame*. The location of a map point is defined by a position vector in the 3D Euclidean space with respect to the *W-frame*:

- Camera poses $_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t} = [_{\mathrm{w}}\mathbf{w}_{\mathrm{c}}^{t},_{\mathrm{w}}\mathbf{t}_{\mathrm{c}}^{t}]^T \in se(3)$
- Map point positions $_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i} = _{\mathrm{w}}\mathbf{p}_{\mathrm{m}}^{i} \in \mathbb{R}^3$

C. Visual Constraints and Re-Projection Errors

Feature point observations of a map point are represented via edges connecting the nodes associated to the camera poses and the corresponding map point. In addition to the *x-y* pixel coordinates in the left image, there is an additional

measurement for stereo vision; the *x* coordinate in the right image. Under the assumption of Gaussian white noise, the measurement error can be modeled with a covariance matrix:

- Visual constraint: $\mathbf{z}_{\mathrm{v}}^{t,i} = [u^{t,i}, v^{t,i}, u'^{t,i}]^T \in \mathbb{R}^3$
- Inverse covariance of the measurement: $\boldsymbol{\Omega}_{\mathrm{v}}^{t,i} \in \mathbb{R}^{3\times3}$

The measurement function is defined as

$$\hat{\mathbf{z}}_{\mathrm{v}}^{t,i} = \mathbf{h}_{\mathrm{v}}^{t,i}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t},_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i}) = \begin{pmatrix} f_x \frac{_{\mathrm{c}}p_{\mathrm{m},x}^{i}}{_{\mathrm{c}}p_{\mathrm{m},z}^{i}} + c_x \\ f_y \frac{_{\mathrm{c}}p_{\mathrm{m},y}^{i}}{_{\mathrm{c}}p_{\mathrm{m},z}^{i}} + c_y \\ f_x \frac{_{\mathrm{c}}p_{\mathrm{m},x}^{i}-b}{_{\mathrm{c}}p_{\mathrm{m},z}^{i}} + c_x \end{pmatrix} \quad (3)$$

where $_{\mathrm{c}}\mathbf{p}_{\mathrm{m}}^{i} = [_{\mathrm{c}}p_{\mathrm{m},x}^{i},_{\mathrm{c}}p_{\mathrm{m},y}^{i},_{\mathrm{c}}p_{\mathrm{m},z}^{i}]^T$ is a 3D position vector of the *i-th* map point with respect to the *C-frame*:

$$_{\mathrm{c}}\mathbf{p}_{\mathrm{m}}^{i} = [_{\mathrm{c}}p_{\mathrm{m},x}^{i},_{\mathrm{c}}p_{\mathrm{m},y}^{i},_{\mathrm{c}}p_{\mathrm{m},z}^{i}]^T = f(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^{t},_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^{i})$$
$$= \begin{bmatrix} \mathbf{R}_{\mathrm{cw}}^{t} \mid _{\mathrm{w}}\mathbf{t}_{\mathrm{c}}^{t} \end{bmatrix} \begin{pmatrix} _{\mathrm{w}}\mathbf{P}_{\mathrm{m}}^{i} \\ 1 \end{pmatrix}$$
$$= \mathbf{R}_{\mathrm{cw}}^{t}{}_{\mathrm{w}}\mathbf{p}_{\mathrm{m}}^{i} + _{\mathrm{w}}\mathbf{t}_{\mathrm{c}}^{t} \quad (4)$$

It is assumed that the intrinsic parameters of the left and right cameras are the same, and the center of the right camera is located at $[b, 0, 0]^T$ with respect to the one of the left camera. The focal lengths are denoted with f_x and f_y, and $[c_x, c_y]^T$ is the coordinate of the principal point in an image.

Finally, a re-projection error of the *i-th* map point in the keyframe at t is calculated as

$$\mathbf{e}_{\mathrm{v}}^{t,i}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^t, {}_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^i) = \mathbf{z}_{\mathrm{v}}^{t,i} - \mathbf{h}_{\mathrm{v}}^{t,i}(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^t, {}_{\mathrm{w}}\mathbf{x}_{\mathrm{m}}^i) \tag{5}$$

D. Ranging Constraints and Distance Errors

A distance measurement between a keyframe and a static base station is modeled as a ranging constraint. The measurement error is assumed to be Gaussian white noise:

- Ranging constraint: $\mathbf{z}_{\mathrm{r}}^t = \rho^t \in \mathbb{R}^1$
- Inverse covariance of the measurement: $\mathbf{\Omega}_{\mathrm{r}}^t \in \mathbb{R}^1$

A range measurement is predicted as the norm of a keyframe position vector with respect to the *B-frame*:

$$\hat{\mathbf{z}}_{\mathrm{r}}^t = \mathbf{h}_{\mathrm{r}}^t(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^t) = ||_{\mathrm{b}}\mathbf{p}_{\mathrm{c}}^t|| \tag{6}$$

$$_{\mathrm{b}}\mathbf{p}_c^t = {}_{\mathrm{w}}\mathbf{p}_c^t - {}_{\mathrm{w}}\mathbf{p}_{\mathrm{b}} = -\mathbf{R}_{\mathrm{cw}}^t{}^T{}_{\mathrm{w}}\mathbf{t}_c^t - {}_{\mathrm{w}}\mathbf{p}_{\mathrm{b}} \tag{7}$$

A ranging error of the keyframe at t is computed as the difference between a measurement and a predicted value:

$$\mathbf{e}_{\mathrm{r}}^t(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^t) = \mathbf{z}_t^b - \mathbf{h}_{\mathrm{r}}^t(_{\mathrm{w}}\mathbf{x}_{\mathrm{c}}^t) \tag{8}$$

E. Robust Least Squares

The least squares problem with the cost function (2) is robustified by the Huber kernel [10], which is quadratic for small $|x|$ but linear for large $|x|$:

$$\mathbf{F}_k = \mathbf{e}_k^T \Omega_k \mathbf{e}_k = \rho_H \left(\sqrt{\mathbf{e}_k^T \Omega_k \mathbf{e}_k} \right) \tag{9}$$

$$\rho_H(x) := \begin{cases} x^2 & if \ |x| < r \\ 2r|x| - r^2 & else \end{cases} \tag{10}$$

V. EXPERIMENTS

In this section, we present both qualitative and quantitative experimental results obtained by applying the proposed algorithm on a real stereo image sequence with synthetically generated ranging measurements. We compare the proposed algorithm to stereo visual odometry and stereo vision-only SLAM with loop closures, demonstrating the effectiveness of the proposed approach.

A. Setup

In order to evaluate the performance of the proposed SLAM algorithm, sequence 07 of the KITTI dataset [12] was used. This dataset contains stereo images from a large-scale outdoor environment (191m×209m) over a time interval of 114s (see Fig. 4). The dataset also provides accurate ground truth poses estimated with GPS and Velodyne laser scanner. However, the dataset does not provide distance measurements, hence we derived ranging measurements by using the ground truth positions with the following assumptions:

- A static base station is located at the origin of the *W-frame*
- Ranging measurements are superimposed with Gaussian white noise. Two noise levels are simulated ($\sigma = 0.1, 0.5$)
- Ranging data reception is regular and synchronized with stereo images (a ranging measurement for every 5 stereo image pairs)

Fig. 4. An example image of the KITTI dataset sequence 07 with detected ORB features

B. Experimental Results

We first present the trajectory and map point estimates obtained with the proposed algorithm and stereo visual odometry on Fig. 5. The location of the static base station is marked with a black cross. It is qualitatively shown that the trajectory estimated by the proposed algorithm are closer to the ground truth than the one by stereo visual odometry. Map point estimates cannot be quantitatively assessed, since there is no ground truth available for the map points. However, trajectory and map point estimations are correlated, hence the more precise trajectory estimates, the more accurate map point estimates. To compare the local accuracy, Relative Pose Errors (RPE) [15] are computed with various segment lengths (Fig. 6). It is shown that both translational and rotational errors are greatly reduced with the proposed algorithm. In addition, we calculated the RMS of Absolute Trajectory Errors (ATE) [15] of the trajectory estimates obtained with the proposed approach, stereo visual odometry, and stereo vision-only SLAM to quantify the global estimation accuracy (Table III). The values of the proposed algorithm with both noise levels ($\sigma = 0.1, 0.5$) are lower than the one of visual odometry. Especially, the performance of the proposed SLAM algorithm is comparable to the one of stereo vision-only SLAM with loop closures.

VI. CONCLUSION

We proposed fusion of stereo visual odometry with ranging measurements for reducing the accumulated estimation errors inherent in visual odometry. The proposed algorithm is evaluated both qualitatively and quantitatively by processing a widely used benchmark dataset with synthetically generated ranging measurements. The experimental results show that the estimation drifts in visual odometry are compensated by ranging fusion. Moreover, the performance of the proposed algorithm is comparable with the one of stereo vision-only SLAM with loop closures.
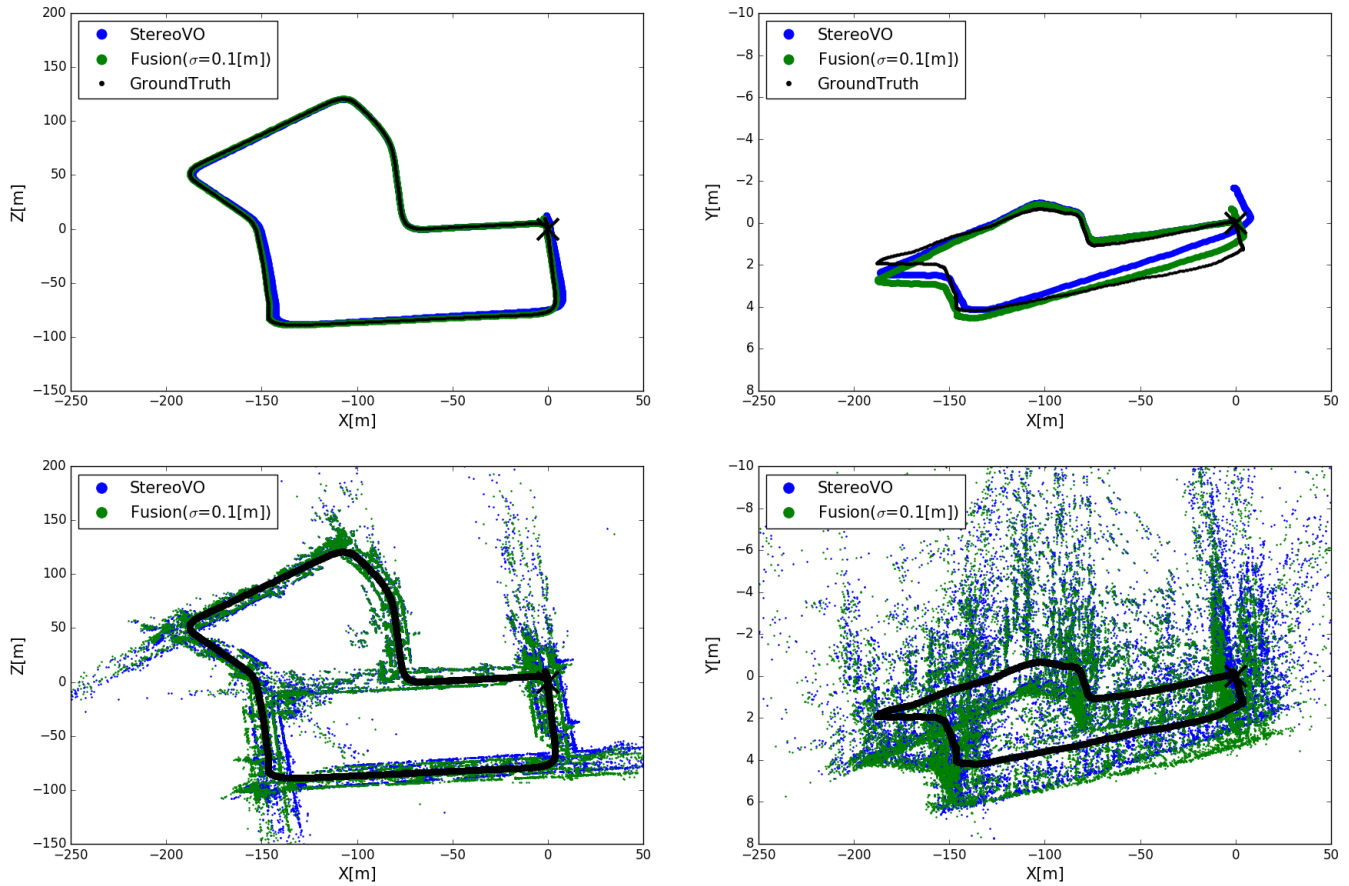
Fig. 5. Trajectories (top) and map points (bottom) of stereo visual odometry without closing a loop (*StereoVO*) and the proposed SLAM algorithm with the Gaussian white noise assumption (*Fusion*($\sigma = 0.1$[m])) on *XZ-* and *XY-Planes* (blue: *StereoVO*; green: *Fusion*($\sigma = 0.1$[m]); black: *GroundTruth*)
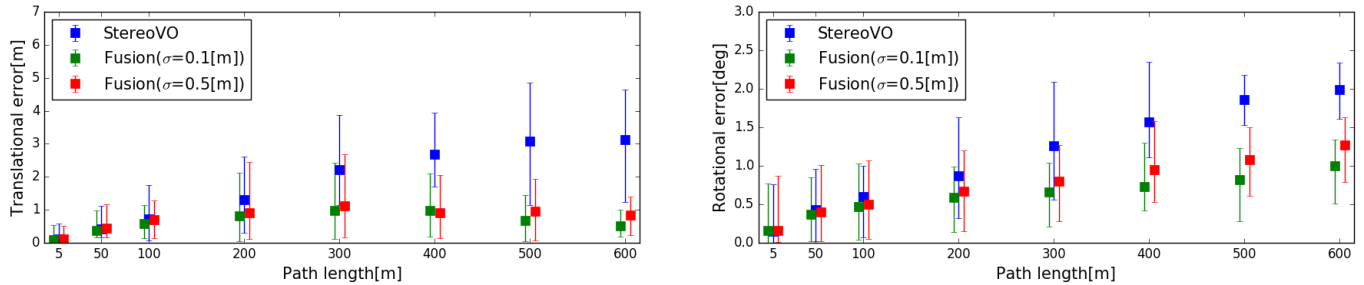

Fig. 6. Relative Pose Error [15] comparison between stereo visual odometry (*StereoVO*) and the proposed SLAM algorithm with superimposed Gaussian white noise (*Fusion*) (blue: *StereoVO*; green: *Fusion*($\sigma = 0.1$[m]); red: *Fusion*($\sigma = 0.5$[m]))

TABLE III
THE ABSOLUTE TRAJECTORY ERRORS [15] BETWEEN THE GROUND TRUTH AND THE ESTIMATES OVER ALL TIME STAMPS (RMSE) [M]

Stereo visual odometry	Stereo vision-only SLAM	SLAM with fusion of stereo visual odometry and ranging	
	(with loop closures)	$\sigma = 0.1m$	$\sigma = 0.5m$
1.43	0.54	0.37	0.52

We applied the fusion algorithm in post-processing, without evaluating the real-time capabilities. As a next step, we are going to implement a real-time version of the method. By operating the fusion steps on a separate processing thread, as proposed in PTAM [1] and ORB-SLAM [2], the real-time tracking will not be hindered by the proposed global optimization step.

A further topic to be addressed is the correct utilization of asynchronous measurements, as the ranging observations and the image frame acquisition may not be triggered by the same clock signal.

ACKNOWLEDGMENT

REFERENCES

[1] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in IEEE and ACM Int. Symp. Mixed and Augmented Reality (ISMAR), 2007.

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Trans. Robot., vol. 31, no. 5, 2015, pp. 11471163.

[3] B. Williams, M. Cummins, J.Neira, P. Newman, I.Reid, and J. Tards, "A comparison of loop closing techniques in monocular SLAM," Robotics and Autonomous Systems, vo. 57, 2009, pp. 1188-1197.

[4] R. Mur-Artal and J. D. Tardos, "Fast relocalisation and loop closing in keyframe-based SLAM," in Proc. IEEE Int. Conf. Robot. Autom., Hong Kong, Jun. 2014, pp. 846853.

[5] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras," arXivpreprint arXiv:1610.06475, 2016.

[6] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 2004.

[7] D. Galvez-Lopez and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," IEEE Trans. Robot., vol. 28, no. 5, Oct. 2012

[8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in Proc. IEEE Int. Conf. Comput. Vision, Barcelona, Spain, Nov. 2011, pp. 25642571.

[9] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" Image Vision Comput., vol. 30, no. 2, 2012, pp. 6577.

[10] R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in IEEE Int. Conf. Robot. and Autom. (ICRA), 2011.

[11] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment a modern synthesis," in Vision Algorithms: Theory and Practice. New York, NY, USA: Springer, 2000, pp. 298372.

[12] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," Int. J. Robot. Res., vol. 32, no. 11, 2013, pp. 12311237.

[13] J. Djugash, S. Singh, G. Kantor, and W. Zhang, "Range-only slam for robots operating cooperatively with sensor networks," Robotics and Automation, ICRA, Proceedings IEEE International Conference on, 2006, pp. 2078-2084.

[14] C. Zhu, G. Giorgi, and C.Guenther, "Scale and 2D Relative Pose Estimation of two Rovers using Monocular Cameras and Range Measurements," ION GNSS+, Proceedings of the International Technical Meeting of The Satellite Division of the Institute of Navigation, 2016, pp. 794-800.

[15] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," Intelligent Robots and Systems (IROS) 2012 IEEE/RSJ International Conference on, 2012, pp. 573-580.

[16] R. Kuemmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner. "On measuring the accuracy of SLAM algorithms," Auton. Robots, 27:387-407, 2009.

[17] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Providence, USA, June 2012.