2D Positioning of Ground Vehicles using Stereo Vision and a Single Ranging Link

Chen Zhu[1], Gabriele Giorgi[1], Young-Hee Lee[2], and Christoph Günther[1,2]

[1] *Institute of Communications and Navigation,*
German Aerospace Center (DLR), Oberpfaffenhofen, Germany
Email: {chen.zhu, gabriele.giorgi, christoph.guenther}@dlr.de

[2] *Institute for Communications and Navigation, Department of Electrical and Computer Engineering,*
Technische Universität München, Munich, Germany
Email: younghee.lee@tum.de

BIOGRAPHY

Mr. Chen Zhu is a full-time researcher at the Institute of Communication and Navigation, German Aerospace Center (DLR) and he is finishing his Ph.D. at Technische Universität München at the Institute for Communications and Navigation. He received his B.Sc. in Automation Engineering from Tsinghua University, in Beijing, China in 2009, and his M.Sc. in Communications Engineering in 2011 from Technische Universität München, in Munich, Germany. His research interests include visual navigation, multi-sensor fusion, robotic swarm navigation, and he is focusing on the integrity of the techniques.

Dr. Gabriele Giorgi is a full-time researcher at the Institute of Communication and Navigation, German Aerospace Center (DLR). He obtained a Ph.D. following his work on Global Navigation Satellite System (GNSS) for aerospace applications from the Delft Institute of Earth Observation and Space Systems (DEOS), Delft University of Technology, in Delft, The Netherlands. He holds a M.Sc. degree in space engineering from the University of Rome "La Sapienza" and a B.Sc. degree in aerospace engineering from the same university. His main research focuses on satellite navigation, visual navigation and multi-sensor fusion.

Ms. Young-Hee Lee is a researcher at the Institute for Communications and Navigation, Technische Universität München, in Munich, Germany.

Prof. Christoph Günther studied theoretical physics at the Swiss Federal Institute of Technology in Zurich, Switzerland. He received his diploma in 1979 and completed his Ph.D. in 1984. He worked on research in cryptography, coding, communication, and information theory with Asea Brown Boveri, Ascom and Ericsson. Since 2003, he is the Director of the Institute of Communication and Navigation at the German Aerospace Center (DLR). The institute employs around 120 scientists. Since 2004, Günther is additionally holding the Chair of Communication and Navigation at Technische Universität München (TUM). The focus of his research work is on navigation. At TUM, he and his team are developing algorithms for achieving a high accuracy. At DLR, the focus is on achieving high levels of integrity.

ABSTRACT

In this work we propose a positioning method for ground vehicles in planar motion, based on sensor fusion of stereo cameras and sparse ranging measurements obtained from a wireless network. The proposed method is an alternative localization solution when Global Navigation Satellite System (GNSS) is unavailable, with notably low requirements on infrastructures. It does not require a database of landmarks and it works in single-link scenarios, i.e., at most one station reachable at

any time. In theory, to estimate two dimensional position without ambiguity, at least three ranging anchors are required. However, in GNSS-denied environments, it is often difficult to achieve simultaneous connectivity to three wireless stations. We propose to apply visual odometry technique to estimate relative motion of the vehicle using stereo cameras, and fuse the vision system with a single ranging link. The sensor fusion method can resolve absolute position unambiguously if the vehicle sequentially connects to two stations with known coordinates. Furthermore, the accuracy of the estimated trajectory is improved by fusing both ranging and visual measurements.

INTRODUCTION

In GNSS-denied environments, various onboard sensors are used for the relative navigation of dynamic vehicles, such as inertial measurement units (IMUs) and vision systems. Visual odometry (VO) techniques based on stereo cameras have shown promising performance in the trajectory estimation of dynamic vehicles using visual cues. The state-of-the-art visual odometry methods, such as [1–4], are able to keep the accumulated position error as low as one percent of the travelled distance, using real images collected on urban streets from the public benchmark datasets [5]. With the aid of other sensors such as IMUs and laser scanners, the drift in visual odometry can be further reduced [6–8]. However, the technique is only capable of positioning the rover relative to the starting location. There are also landmark-based visual navigation approaches that provides absolute localization capability, such as the work [9–11]. The map-based methods, however, requires an up-to-date database of landmarks with known coordinates. Additionally, the mismatch error between the images and the database is difficult to monitor, especially when there are similar or repetitive patterns in the scenes.

Reliable absolute positioning in absence of GNSS often relies on signals-of-opportunity (SOP) networks consisting of stations or anchors with known positions, e.g., cellular networks, WiFi, ultra-wideband (UWB) tags, etc. One state-of-the-art method of positioning using cellular network as SOP is [12]. In general, by estimating the round-trip delay, range measurements can be obtained from the wireless radio signals. In order to obtain an unambiguous position of a vehicle in planar motion, at least three anchor points with known location are required for trilateration. For ranging based two dimensional (2D) positioning problems, by exploiting the range from a single station, the possible position of the rover distributes uniformly on a circle centered at the station, i.e., the red-circle in the example in Fig. 1. If a second station is in connection at the same time, the localization solution is constrained to two ambiguous points, marked as blue stars in the figure. The localization ambiguity can be solved by adding a third anchor for ranging-only positioning. However, in practice, most signals of opportunity networks do not provide ubiquitous threefold coverage, since they are not designed for pure navigation purposes. Fig. 2 is an example of a wireless network providing only single-station coverage with some dead zone between stations. It is hence valuable to investigate the absolute positioning capability in such challenging but realistic scenarios, in which only one ranging link from the base station network is available at any time.

In this work, we propose a 2D positioning approach based on the sensor fusion of a stereo camera rig and single link ranging measurements. By processing the visual cues and the ranging measurements from multiple epochs, the absolute position in the global reference frame can be estimated if the vehicle sequentially connects to two stations with known coordinates.
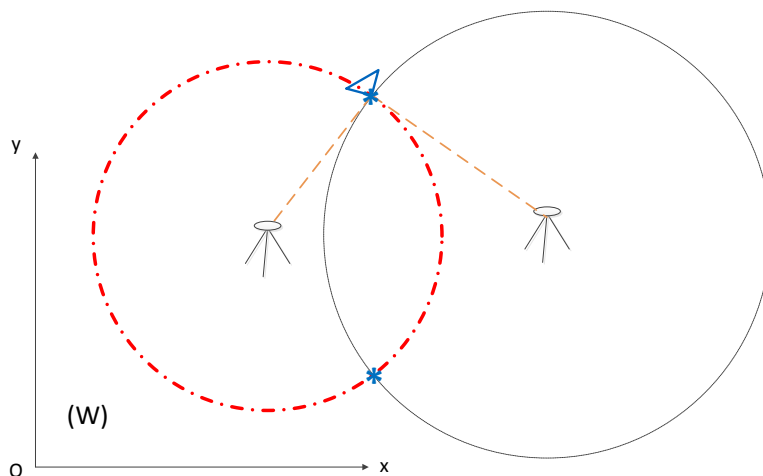
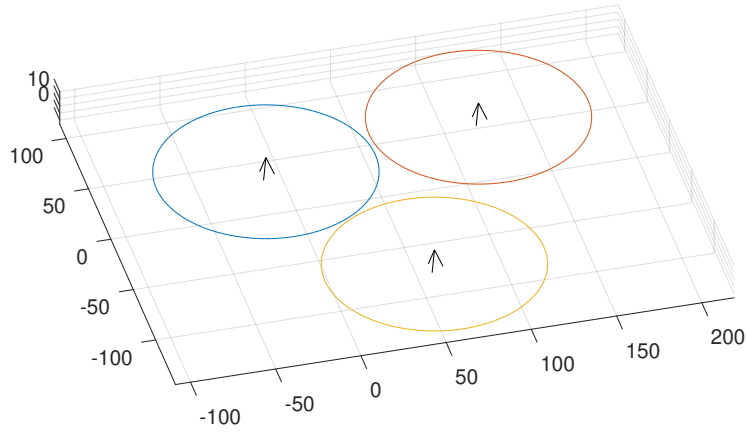

Figure 1: *2D positioning geometry*

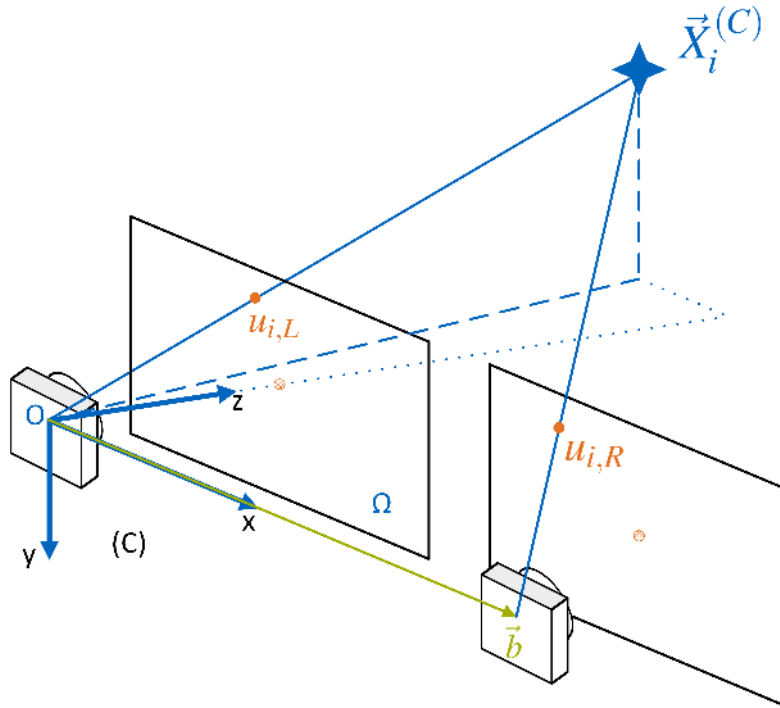Figure 2: *A wireless network with single-station coverage*

Figure 3: *Projection model for a stereo camera rig*

Although two stations are required to resolve the whole geometry, it is not necessary in our method to connect to both stations simultaneously. It can tolerate short hand-over periods with radio black-outs, which enables positioning of ground vehicles using networks even with non-overlapping coverage. The proposed method has significantly lower requirements on the infrastructures than other methods. Furthermore, we couple the visual and ranging measurements to improve the accuracy of the vehicle trajectory estimation.

The organisation of the paper is as follows: first, we briefly introduce the stereo-camera-based visual odometry technique. Subsequently, the sensor fusion algorithm is proposed. We begin with the estimation of the anchor position in navigation frame using vision and a single ranging link, followed by proposing a method to improve the numerical convergence probability. Then, the method to estimate the vehicle position in the global reference frame is introduced. In the following section, the simulation results of the proposed positioning method are provided for evaluation, and conclusions are drawn from the analyses.

Figure 4: *Projection of a point in the navigation frame*

SYSTEM MODEL AND INTRODUCTION TO STEREO VISUAL ODOMETRY

Fig. 3 shows the projection model for the chosen stereo setup. The origin of the camera frame (C) is defined at the projection center of the left camera. $\Omega \subset \mathbb{R}^2$ is the image plane. Applying the pinhole model, the perspective projection can be formulated as

$$\tilde{u}_{iL} = d_i[u_{iL}, 1]^T = K_L \vec{X}_i^{(C)}, \tag{1}$$

where $d_i = X_{i,z}^{(C)}$ is the depth of the point, and K_L is the camera intrinsic matrix. $u_{iL} \in \mathbb{R}^2$ denotes the Cartesian coordinates of the point's two-dimensional (2D) location in the image, and $\tilde{u}_{iL} \in \mathbb{P}^2$ is the corresponding homogeneous coordinates in the extended Euclidean space. Assuming the image planes of both cameras in the stereo rig are well aligned, the position of the right camera in the camera frame is $\vec{b}^{(C)} = [l, 0, 0]^T$. The projection of the same point on the right camera is

$$\tilde{u}_{iR} = d_i[u_{iR}, 1]^T = K_R(\vec{X}_i^{(C)} - \vec{b}^{(C)}). \tag{2}$$

Using the matched visual features at both image planes, the depth d_i can be retrieved and the three-dimensional (3D) location of the point can be estimated as $\hat{X}_i^{(C)}$.

We define a navigation frame (N) as a fixed coordinate frame with its origin at the starting location of the rover. The navigation frame of each rover is related to the world reference frame by a specific transformation dependent on the initial position and attitude of the vehicles. The projection of a point in the navigation frame is shown in Fig. 4. For a stereo rig mounted on a vehicle constrained to be moving in a plane, the pose can be parameterized by three parameters $\xi_{[k]}^{(N)} = [c_{[k],x}^{(N)}, c_{[k],y}^{(N)}, \phi_{[k]}^{(N)}]^T \in \mathfrak{se}(2)$ as

$$\vec{c}_{[k]}^{(N)} = \begin{bmatrix} c_{[k],x}^{(N)} \\ c_{[k],y}^{(N)} \\ 0 \end{bmatrix}, R_{(N \to k)} = \begin{bmatrix} \cos(\phi_{[k]}^{(N)}) & -\sin(\phi_{[k]}^{(N)}) & 0 \\ 0 & 0 & -1 \\ \sin(\phi_{[k]}^{(N)}) & \cos(\phi_{[k]}^{(N)}) & 0 \end{bmatrix}. \tag{3}$$

The planar position is $\vec{\beta}_k^{(N)} = [c_{[k],x}^{(N)}, c_{[k],y}^{(N)}]^T$. The feature location measurements of the point i at time k can be modeled as

$$\mu_{ik} = u_{ik} + n_{u,ik} \in \mathbb{R}^4, \tag{4}$$

with $u_{ik} = [u_{i[k],L}; u_{i[k],R}] \in \mathbb{R}^4, E\{n_{u,ik}\} = \mathbf{0}, E\{n_{u,ik}n_{u,ik}^T\} = \Sigma_{u,ik}$. $E\{\cdot\}$ denotes the expected value function.

Using feature detectors, several feature points can be matched between the stereo images and tracked over frames for a period of time. Using N_k tracked features with 3D coordinates $\{\hat{X}_i^{(N)} : i = 1,...,N_k\}$ triangulated at time k, the pose of the vehicle at time $k+1$ can be estimated by minimizing the reprojection error

$$\hat{\xi}_{[k+1]}^{(N)} = \arg\min_{\xi_{[k+1]}^N} \sum_{i=1}^{N_k} \left\| \mu_{i,k+1} - \pi(\hat{X}_i^{(N)}, \xi_{[k+1]}^{(N)}) \right\|_{\Sigma_{u,ik+1}^{-1}}^2, \tag{5}$$

where $\pi(\cdot,\cdot) : \mathbb{R}^3 \times \mathfrak{se}(2) \to \mathbb{R}^4$ is the projection function, and $\|\cdot\|_{\Sigma^{-1}}$ denotes the Mahalanobis distance in the metric given by the covariance matrix Σ.

Using the estimated pose, the position of the newly detected features in frame $k+1$ can be updated. As a result, the motion tracking can be continued as long as sufficient features can be tracked across consecutive frames. In addition, state-of-the-art visual odometry algorithms usually apply a batch optimization for both 3D map point position and the vehicle poses over a short period of time, so that the drift of the estimated trajectory can be mitigated. By applying the stereo visual odometry technique, the rover obtains a set of egomotion estimates $\{\hat{\xi}_{[k]}^{(N)}\}$ expressed in its own navigation frame, which consists of position estimates $\{\hat{\beta}_{[k]}^{(N)}\}$ and attitude (heading) estimation $\{\hat{\phi}_{[k]}^{(N)}\}$.

ANCHOR POSITION DETECTION IN NAVIGATION FRAME USING VISION AND RANGING FUSION

In many cases, a wireless network can only provide a single-link coverage, i.e., the rover is connected to only one station at a time. In one-link situations, the rover cannot position itself by just using the ranging measurements obtained from the radio signal. However, for a dynamic rover which is capable of estimating its egomotion using visual odometry, we show in this section that the anchor position in the rover's navigation frame (N) can be estimated given sufficient motion dynamics. It should be mentioned that even if the coordinates of the anchor is known in both global frame and navigation frame, the position of the rover in the global reference frame cannot be determined uniquely. The problem and a possible solution are discussed in detail in the next section.

Fig. 5 shows the model of the 2D anchor detection problem in the navigation frame of the rover.

At time instant k (sampled at k-th keyframe of the vision system), the range measurement between the rover and the base station is

$$\rho_{[k]} = \|\vec{x}_b^{(N)} - \vec{\beta}_{[k]}^{(N)}\| + \eta_{[k]}, \tag{6}$$

where $\vec{x}_b^{(N)}$ is the base station location in the navigation frame, and $\eta_{[k]}$ is the ranging noise with covariance $E\{\eta_{[k]}\eta_{[k]}^T\} = \sigma_k^2$. The position of the rover in the navigation frame can be obtained from egomotion estimation using visual odometry. With sufficient dynamics of the rover, the station position estimation problem is formulated as in Eqn. (7) by stacking the measurements over time:

$$\hat{x}_b^{(N)} = \arg\min_{\vec{x}_b^{(N)}} \|\rho - F_b(\vec{x}_b^{(N)})\|_{\Sigma^{-1}}, \tag{7}$$

where the k-entry of $F_b(\vec{x}_b^{(N)})$ is the range function $\|\vec{x}_b^{(N)} - \hat{\beta}_{[k]}^{(N)}\|$ and Σ^{-1} is the measurement covariance matrix. Obviously the cost function is nonlinear, so the convergence of the numerical algorithm to the global minimum is highly dependent on the initialization of the estimated parameters. The initial values can be obtained by a coarse estimation using direct trilateration methods such as the approach from Thomas and Ros [13]. However, depending on the measurement error distribution, the coarse estimates can be significantly inaccurate in practice, which results in unsolved convergence problem of the nonlinear optimization.

As a solution, we parameterize the station location using polar coordinates instead of Cartesian coordinates, because the range measurements provide good initialization for one of the parameters. By using the parameterization $\zeta_b = [r_{[1]}, \varphi]^T$, the problem can be reformulated as

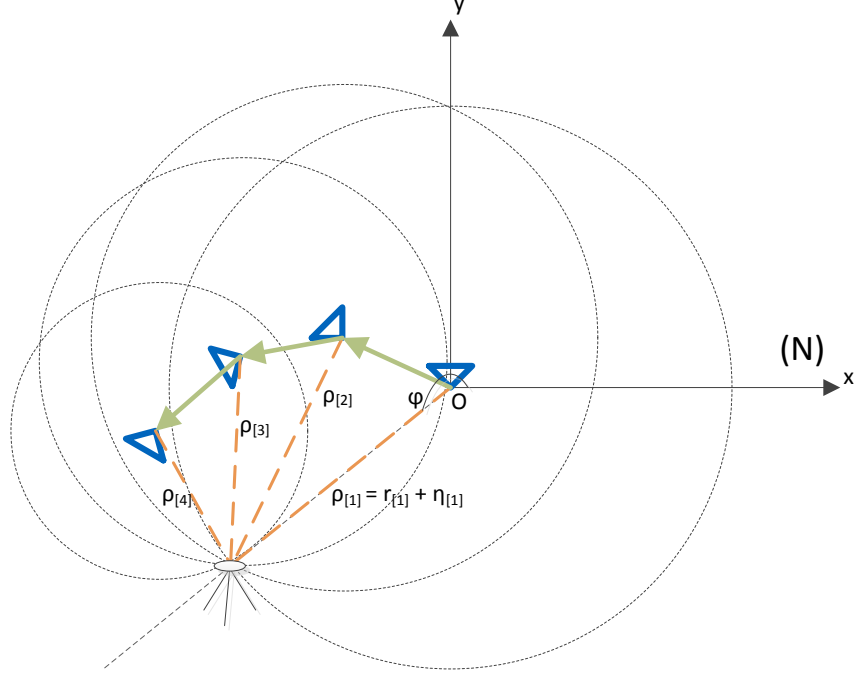$$\hat{\zeta}_b = \arg\min_{\zeta_b} \|\rho - F(\zeta_b)\|_{\Sigma^{-1}}, \tag{8}$$

Figure 5: *2D anchor detection using ranging measurements*

where $F(\zeta_b) = [F_1, ..., F_K]^T$ with the k-th entry $F_k = \|[r_{[1]}\cos(\varphi), r_{[1]}\sin(\varphi)]^T - \hat{\beta}_{[k]}^{(N)}\|$. The optimization can be obtained iteratively by solving a linearized problem as

$$\hat{\zeta}_b = \arg\min_{\zeta_b} \|\rho - J_F(\zeta_b)\zeta_b\|_{\Sigma^{-1}}^2 \tag{9}$$

$$\hat{\zeta}_{b,i+1} = \hat{\zeta}_{b,i} + \left(J_F^T|_{\hat{\zeta}_{b,i}}\Sigma^{-1}J_F|_{\hat{\zeta}_{b,i}}\right)^{-1}J_F^T|_{\hat{\zeta}_{b,i}}\Sigma^{-1}\left(\rho - F(\hat{\zeta}_{b,i})\right) \tag{10}$$

where $J_F|_{\hat{\zeta}_{b,i}}$ is the Jacobian matrix of function $F(\cdot)$ linearized at $\hat{\zeta}_{b,i}$.

The initialization of the azimuth angle φ can be done based on multiple assumptions $\{j = 1, ..., N_j | \varphi^j\}$, and the optimization in Eqn. (8) can be resolved independently for each initialization assumption. With $\hat{\zeta}_b^j$ as the solution of the j-th initialization, the global minimum can be selected from the local minima by

$$\hat{\zeta}_b = \min(\{j = 1, ..., N_j | \hat{\zeta}_b^j\}). \tag{11}$$

As a result, the base station (anchor) location in the rover's navigation frame can be obtained from

$$\hat{x}_b^{(N)} = [\hat{r}_1\cos(\hat{\varphi}), \hat{r}_1\sin(\hat{\varphi})]^T. \tag{12}$$

Ideally only three keyframes are required to solve the problem as long as the egomotion is non-colinear. Nevertheless, the uncertainty can be rather high due to the presence of ranging noise. As a result, using all available measurements over a period of time to solve the least-squares problem can significantly improve the accuracy.

2D POSITIONING IN GLOBAL REFERENCE FRAME

As discussed in the previous section, the base station location can be estimated in the rover's navigation frame using ranging measurements from the single station and the visual odometry outputs. However, in the global frame (W), there is an unobservable angle which results in ambiguous rover position. Fig. 6 shows the ambiguity in an intuitive example. As
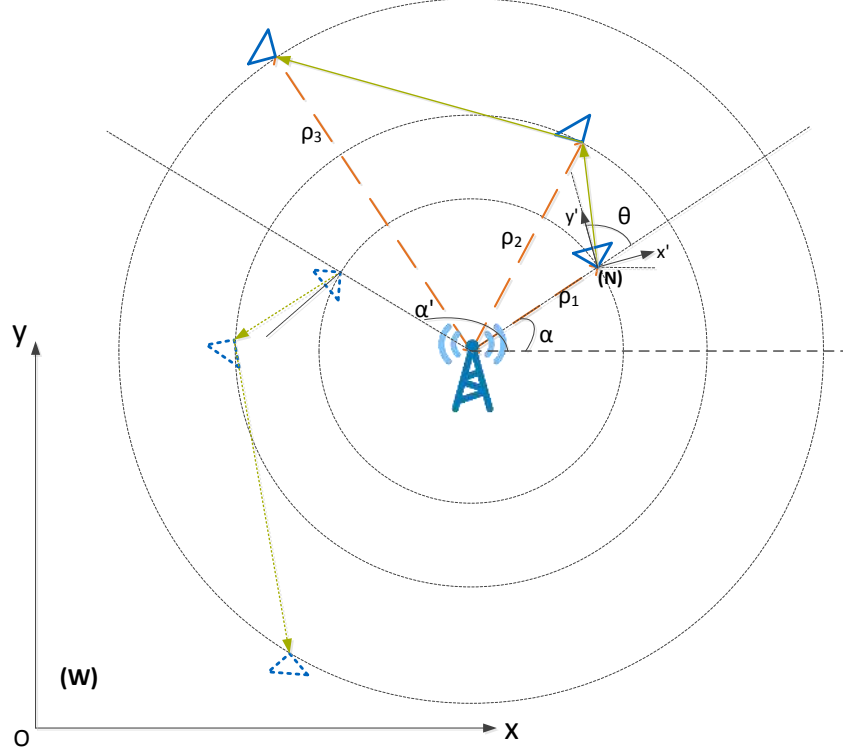
Figure 6: *Polar ambiguity in positioning based on single-station ranging and vision fusion.*

shown in the figure, if the whole trajectory of the rover is rotated from the true positions with polar angle α to the trajectory in dashed line with α', all the measurements from both cameras and radio receiver are invariant. The problem is discussed in detail in our earlier work [14]. By using range measurements and a monocular camera, the heading angle θ can be estimated by the method proposed in [14]. The heading angle θ is an equivalent parameterization of the beacon direction angle φ in the navigation frame. The transformation between the two representations in Fig. 6 and Fig. 5 follows the relation

$$\varphi = \frac{\pi}{2} - \theta + \pi = \frac{3\pi}{2} - \theta. \tag{13}$$

The geometric ambiguity caused by the unobservable polar angle can be resolved after the rover is connected to a second station. In a radio network with non-overlapping coverage that no more than one station is in the covered range at any place, e.g., the trajectory in Fig. 7, the connection to the two stations are never simultaneous. Nevertheless, since the motion in the navigation frame can be estimated using visual odometry, the ambiguity can be resolved even if the two stations are connected sequentially, which can also tolerate short hand-over periods with radio black-outs.

By using the anchor position detection method in , the position of the base stations in the rover's navigation frame (N) can be estimated. We denote the two stations' location in the global global frame as $\vec{x}_{b_1}^{(W)}$ and $\vec{x}_{b_2}^{(W)}$ respectively, and the estimated position in the navigation frame are $\hat{x}_{b_1}^{(N)}$ and $\hat{x}_{b_2}^{(N)}$. With the anchor position estimation, the positioning problem is transferred to find the optimal transformation between the two reference frames, which is the well-known rigid body estimation problem [15]. In the planar motion case, the problem is simplified to the following least-squares estimation:

$$\{\hat{t}_{(N\to W)}, \hat{R}_{(N\to W)}\} = \arg\min_{t,R} \sum_{i=1}^{2} \|R\hat{x}_{b_i}^{(N)} + t - \vec{x}_{b_i}^{(W)}\|^2, \tag{14}$$

where $t \in \mathbb{R}^2, R \in \mathbf{SO}(2)$.

The problem can be solved by the state-of-the-art algorithms mentioned in [15] using singular value decomposition (SVD).
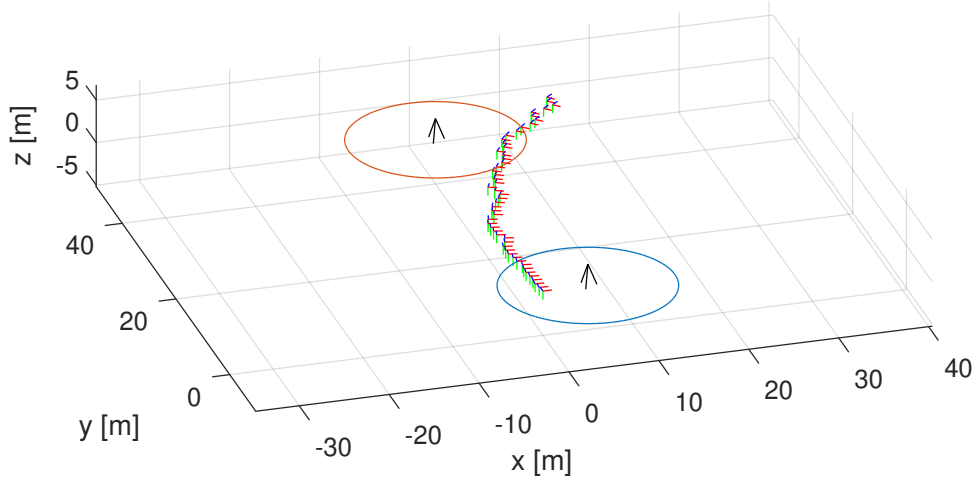
Figure 7: *A trajectory instance in a network with non-overlapping coverage.*

By knowing the coordinate transformation between the global frame and the navigation frame, the rover trajectory can be transformed to estimates in the global frame without ambiguity as:

$$\hat{\beta}_{[k]}^{(W)} = \hat{R}_{(N \to W)} \hat{\beta}_{[k]}^{(N)} + \hat{t}_{(N \to W)}. \tag{15}$$

In addition, the absolute attitude can be estimated as:

$$\hat{R}_{(k \to W)} = \hat{R}_{(N \to W)} \hat{R}_{(k \to N)}, \tag{16}$$

where $\hat{R}_{(k \to N)}$ can be obtained from visual odometry using the method introduced in .

The obtained coordinates $\hat{\beta}_{[k]}^{(W)}$ in the global frame are coarse position estimates which are transformed from the visual odometry outputs. In order to improve the positioning accuracy, the coarse estimates can be used as initial values for a non-linear optimization which fuses both visual and ranging measurements as:

$$\hat{\beta}^{(W)} = \arg\min_{\beta^{(W)}} \sum_{i=1}^{N_b} \left\| \rho_i - \|\vec{x}_{b_i}^{(W)} - \beta^{(W)}\|_2 \right\|_{\Sigma_{\rho_i}^{-1}}^2 + \sum_{k=2}^{N_k} \|l_{[k]} - S_k \beta^{(W)}\|_{\Sigma_{l_{[k]}}^{-1}}^2, \tag{17}$$

where $\beta^{(W)} = [\beta_{[1]}^{(W)}, \beta_{[2]}^{(W)}, ..., \beta_{[N_k]}^{(W)}]^T$ is the position vector in global frame, $\rho_i = [\rho_{i,[1]}, \rho_{i,[2]}, ..., \rho_{i,[N_k]}]^T$ the ranging measurements from i-th station, $l_{[k]} = \hat{\beta}_{[k]}^{(N)} - \hat{\beta}_{[k-1]}^{(N)}$ the translation estimate from visual odometry at time k, and

$$S_k = \begin{bmatrix} \vec{0} & -1 & 0 & 1 & 0 & \vec{0} \\ \vec{0} & 0 & -1 & 0 & 1 & \vec{0} \end{bmatrix}^T$$

is a sparse selection matrix with non-zero entries only at $2k-3$ to $2k$-th columns. The non-linear optimization can be solved iteratively using Levenburg-Marquart algorithm [16], which is similar as the process in Eqn. (9) and (10). Due to the sparsity of the measurement space, the solution can be computed efficiently using iSAM2 algorithm [17].

Since in this work we focus on positioning instead of mapping, the fusion in Eqn. (17) is based on loose coupling of the visual and ranging sensors to reduce complexity. If a global map of the feature points is tracked and maintained during the visual odometry estimation (it becomes visual simultaneous localization and mapping technique), the accuracy can be further improved using tightly coupled fusion as used in [18] and [19]. Nevertheless, there is a trade-off between accuracy and computational complexity.
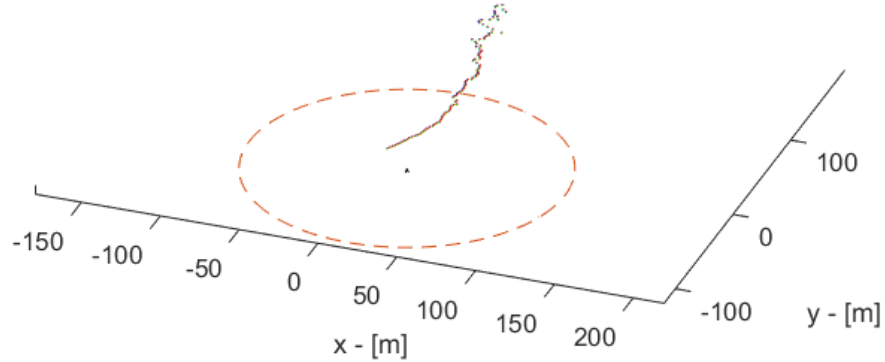
Figure 8: *Instance of test trajectory for anchor detection method.*
Table 1: *Estimation error of the station position in the navigation frame.*

	Convergence rate	RMSE($\hat{x}_{b,x}^{(N)}$) - [m]	RMSE($\hat{x}_{b,y}^{(N)}$) - [m]
polar parameterization	94%	0.7188	0.2258
Cartesian parameterization	0%	44.6757	16.1089

SIMULATION RESULTS

500 trajectory instances are generated with independent random walk noise for testing our anchor detection method. An instance of the trajectories is shown in Fig. 8. In the simulation, the base station locates at $(20, -20, 0)$ in the navigation frame. The ranging measurement noise has standard deviation of 1 [m] and the error of the visual odometry is 1 percent of the translation vector. Both parameters are chosen according to empirical values from real camera and UWB sensor hardware measurements. The root-mean-square-error of the estimated station position is shown in Table 1.

To evaluate the performance of the sensor fusion method, we test the algorithm on various trajectories and different geometry between the base stations and the rover. Take the instance in Fig. 7 as a typical scenario. We set two stations without overlapping coverage. The coverage radius of a base station is assumed to be 10 meters in radius. The stereo rigs' intrinsic parameters and sensor model are provided by a real camera. Noise is added on both the visual measurements and the range measurements. Fig. 9 shows an instance of the noisy feature points distribution on both images of the stereo rig in the simulation.

In this scenario the rover is sequentially connected to two different stations, and there is a radio dead zone between the covered areas of the two stations. Neither the radio-only method nor the vision-only method is capable of locating the rover in global frame in such cases. By applying the anchor position detection method, the coordinates of the two stations in the rover's navigation frame (N) can be estimated respectively. Following the proposed sensor fusion method, the trajectory of the rover in global frame (W) can be estimated with the limited infrastructure in the scenario. The result of 500 runs of the simulation exploiting independent random noise is shown in Fig. 10. The ranging noise is assumed to be 0.5 percent of the true range, which is set according to empirical values of our ultra-wide-band (UWB) based two-way ranging hardware. The standard deviation of the visual odometry error is 1 percent of the translation vector. In the plot, the yellow curve illustrates the root-mean-squared-error (RMSE) of the rover position in global frame over time. The positioning error is
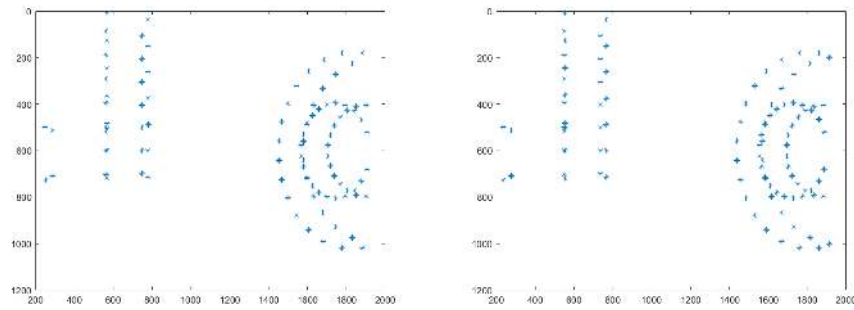
Figure 9: *An instance of the stereo feature locations in the simulation.*
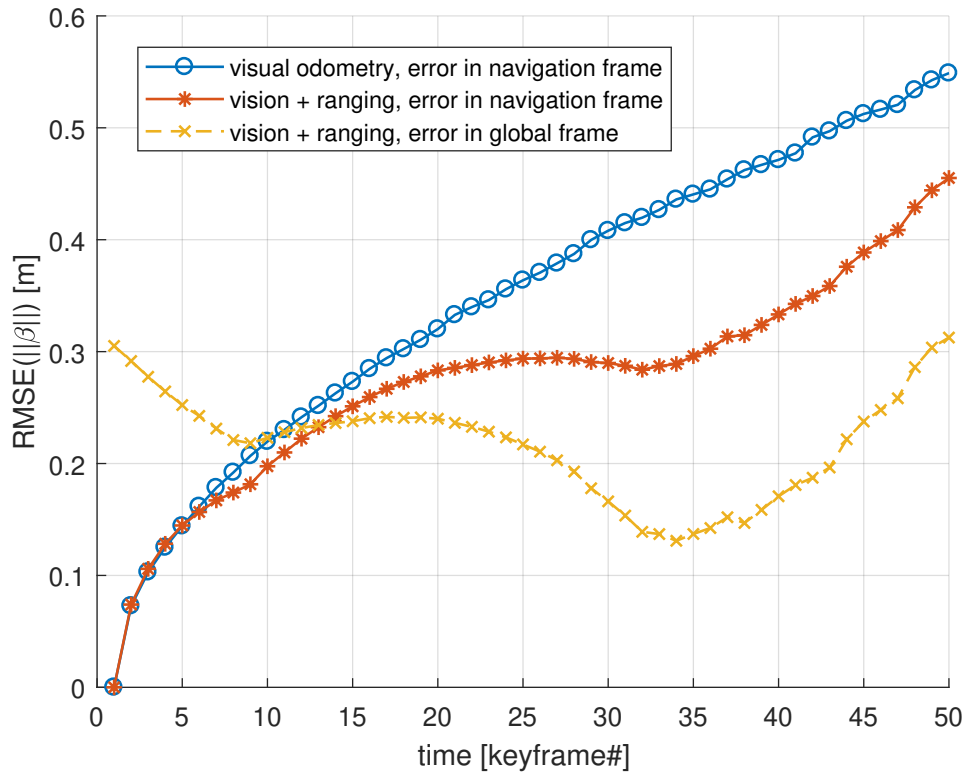

Figure 10: *Global position estimation for trajectory 1.*

between 0.1 [m] to 0.3 [m]. The visual odometry approach, which is denoted using the blue curve in the plot, can only estimate the position in navigation frame, but cannot deliver a positioning solution in global frame. It can be seen that the estimation error accumulates over time, since it can only estimate the relative motion. In order to compare with the vision-only approach, the position error of the sensor fusion method in navigation frame is shown as the red curve. By applying the ranging measurements from the anchor stations, the drift is significantly mitigated. Referring to the geometry in Fig. 7, the sensor fusion method has advantages in accuracy over the vision-only approach when the rover is in the connectivity range of the base stations.

CONCLUSION

In GNSS-denied environments, the positioning of a dynamic rover can rely on a wireless network with ranging capability and an onboard stereo camera rig. We propose a sensor fusion based positioning method, which only requires to connect to one single station at any time. By exploiting the pose estimation from the visual odometry technique, the trajectory of the moving rover can be estimated without ambiguity by connecting to different base stations sequentially. Compared with pure radio based positioning methods, the proposed technique has much less requirements on infrastructures and is more feasible in practice.

REFERENCES

[1] I. Cvišić, J. Ćesić, I. Marković, and I. Petrović, "Soft-slam: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles," *Journal of Field Robotics*, vol. 35, no. 4, pp. 578–595. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21762

[2] R. Wang, M. Schwrer, and D. Cremers, "Stereo dso: Large-scale direct sparse visual odometry with stereo cameras," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 3923–3931.

[3] R. Mur-Artal and J. D. Tards, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, Oct 2017.

[4] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: Semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, April 2017.

[5] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[6] J. Zhang and S. Singh, "Laservisualinertial odometry and mapping with high robustness and low drift," *Journal of Field Robotics*, vol. 35, no. 8, pp. 1242–1264. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21809

[7] J. Zhang, M. Kaess, and S. Singh, "A real-time method for depth enhanced visual odometry," *Autonomous Robots*, vol. 41, no. 1, pp. 31–43, Jan 2017. [Online]. Available: https://doi.org/10.1007/s10514-015-9525-1

[8] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, Feb 2017.

[9] M. Salarian, A. Manavella, and R. Ansari, "Accurate localization in dense urban area using google street view images," in *2015 SAI Intelligent Systems Conference (IntelliSys)*, Nov 2015, pp. 485–490.

[10] G. Floros, B. van der Zander, and B. Leibe, "Openstreetslam: Global vehicle localization using openstreetmaps," in *2013 IEEE International Conference on Robotics and Automation*, May 2013, pp. 1054–1059.

[11] S. Ramalingam, S. Bouaziz, P. Sturm, and M. Brand, "Skyline2gps: Localization in urban canyons using omni-skylines," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 3816–3823.

[12] K. Shamaei, J. Khalife, and Z. M. Kassas, "Exploiting lte signals for navigation: Theory to implementation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2173–2189, April 2018.

[13] F. Thomas and L. Ros, "Revisiting trilateration for robot localization," *IEEE Transactions on Robotics*, vol. 21, no. 1, pp. 93–101, Feb 2005.

[14] C. Zhu, G. Giorgi, and C. Günther, "Planar pose estimation using a camera and single-station ranging measurements," in *Proceedings ION GNSS+ 2017, Portland, Oregon*, 2017.

[15] D. Eggert, A. Lorusso, and R. Fisher, "Estimating 3-d rigid body transformations: a comparison of four major algorithms," *Machine Vision and Applications*, vol. 9, no. 5, pp. 272–290, Mar 1997. [Online]. Available: https://doi.org/10.1007/s001380050048

[16] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[17] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.

[18] C. Zhu, G. Giorgi, Y. Lee, and C. Gnther, "Enhancing accuracy in visual slam by tightly coupling sparse ranging measurements between two rovers," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, April 2018, pp. 440–446.

[19] Y. Lee, C. Zhu, G. Giorgi, and C. Guenther, "Stereo vision-based simultaneous localization and mapping with ranging aid," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, April 2018, pp. 404–409.