Absolute localisation in confined spaces using deep geometric features

Brogaard, Rune Y.; Ravn, Ole; Boukas, Evangelos

Published in:
Electronics Letters

Link to article, DOI:
10.1049/ell2.12199

Publication date:
2021

Document Version
Peer reviewed version

Absolute localisation in confined spaces using deep geometric features

Rune Y. Brogaard,✉ Ole Ravn, and Evangelos Boukas

Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark

✉**Correspondence**

Rune Y. Brogaard, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark.

Email: rybro@elektro.dtu.dk

When operating in dark and confined spaces, the capacity of the robots to localise in an absolute reference frame is of utmost importance. This letter presents an absolute localisation system, using deep 3D landmarks, for known confined spaces. The system estimates the robot's relative localisation using visual inertial odometry. Local deep 3D landmarks are extracted from the robot's view. Similar 3D landmarks are, also, extracted from the prior map which are then registered with the local landmarks to provide absolute localisation via an extended Kalman filter. To the best of knowledge, deep 3D feature registration has not been used before for absolute localisation. The proposed localisation system is tested within a representative application area—i.e. a structured, confined space—and the results indicate greater accuracy and lower processing time when compared to mainstream 3D registration approaches.

Introduction: Localising within industrial dark and confined spaces can be a challenging task in robot applications. Structured areas, such as industrial environments, often contain featureless surfaces hindering visual tracking or recognition. Visual localisation systems are known to produce at best 1% error [1, 2] which is also unbounded without some prior knowledge or re-observation. Repetitive patterns — of both visual and structural nature—can make loop closure, and in extension absolute localisation, prone to errors. Furthermore, occlusions can diminish the accuracy of visual inertial odometry (VIO) systems, while low light conditions limit the number of reliable visual landmarks. Additionally, environmental attributes, such as dust, can introduce noise to sensors and, in consequence to the estimation.

Using structural features from confined spaces for absolute localisation has been previously investigated in [3], however the landmarks were custom designed to represent use case-specific structural attributes. Another approach investigated by the authors in [4], is to fuse positions from a pre-installed ultra wide band(UWB) satellite system with a VIO pose estimate. Even though this method could be suitable for warehouses, it is not possible to equip all the relevant GNSS-denied confined spaces with UWB systems. Another application area for GNSS-denied absolute localisation is the robotics space exploration [5, 6]. A different approach could be the referencing of online data to offline 3D models (stemming either from CAD or from previous stereo reconstruction [7, 8]).

Absolute localisation using 3D geometric features has become an interesting research subject due to the rise of better feature detection [9, 10] and robust registration approaches [11–12]. The aforementioned, novel registration approaches have been applied to the problem of object pose estimation, rather than robot localisation. To the best of our knowledge, this is the first work employing deep 3D feature registration for absolute localisation in industrial confined spaces.

In this letter we propose an absolute localisation system for confined spaces using deep 3D geometric features extracted from point clouds. Extracting geometric features from both the preexisting map and the current field of view of the robot, we find the relative pose between the two point clouds to obtain an absolute pose in the map frame. Due to the computational cost of the point cloud registration, we employ VIO—which provides high frequency relative localisation—and couple it with the absolute pose at a lower frequency, via an extended Kalman filter.

Using state of the art deep geometric feature descriptors, on both the robot's field of view and the already available CAD models of confined spaces, we can extract better point correspondences between the clouds. Combining these correspondences with a fast and outlier-robust registration algorithm results in better registration performance compared to previous methods. Furthermore, fusing a fast robust VIO relative localisation system with the registration, gives rise to a reliable absolute localisation system even in highly repetitive structural environments.

The rest of the letter will describe the system architecture, the experiments conducted in a real-world confined area, and the performance of the system with regards to position accuracy and computational time.

System Description: The proposed localisation system, Figure 1, consists of two modules, the first of which is pose estimation based on 3D point cloud registration. The second module contains the filter, which rejects 3D registrations with sudden motion changes, and fuses the 3D registered pose estimations with VIO estimates to increase the update rate of the resulting absolute pose estimation.

To extract features in both the map and sensor point cloud, we employ a 3D deep convolutional neural network-based algorithm, i.e. 3DSmoothNet [9], which comprises a two-part pipeline. The first part is a voxelised smoothed density value (SDV) representation, which is computed per interest point and aligned to the local reference frame (LRF) to achieve rotation invariance. The second part is the Siamese deep learning architecture which receives as input the normalised 3D SDV voxel grid. The result of the 3DSmoothNet pipeline is a feature descriptor for the given interest points in the point cloud. The interest points are selected randomly—approximately 50% of all points in the map and 25% of the sensor point cloud. By randomly selecting points in both clouds, we minimise the computational load while maintaining a high probability of acquiring correspondences. The feature descriptors of the interest points are organised in k-d trees to allow the efficient correspondences search between the map and sensor point cloud.

The noisy corresponding points are fed into a robust registration algorithm, i.e. TEASER++ [11] which calculates the transformation between the map and sensor cloud. In an ideal case, where no outliers exist among the correspondences, the registration can be defined as a nonlinear least squares solution, as follows:

$$\min_{s>0,\mathbf{R}\in SO(3),t\in\mathbb{R}^3} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \|\mathbf{b}_i - s\mathbf{R}\mathbf{a}_i - t\|^2 \tag{1}$$

where the minimisation is performed over the scale s, the rotation $\mathbf{R}$, and the translation t. Due to the metric environment of the use case, a scale of 1 can be assumed and, therefore, no additional computational resources will be spent on estimating the scale. The 3D correspondence points are notated as $\mathbf{a}$ and $\mathbf{b}$, where (a_i, b_i) represent a correspondence pair between the map and sensor point cloud features and N represents the total number of corresponding points. Due to the noisy measurements of the point cloud reconstruction, a Gaussian noise with isotropic covariance described by σ^2 is included. However, for most real world cases, correspondences with zero outliers cannot be safely assumed [11] and therefore the robust registration can be performed using a truncated least squares function 2, as follows:

$$\min_{s>0,\mathbf{R}\in SO(3),t\in\mathbb{R}^3} \sum_{i=1}^{N} \min\left(\frac{1}{\beta_i^2} \|\mathbf{b}_i - s\mathbf{R}\mathbf{a}_i - t\|^2, \bar{c}^2\right) \tag{2}$$

Equation (2) describes a least squares solution of measurements with small residuals ($\leq \bar{c}^2$), where β_i is a given bound of the noise. This noise bound is set as the maximum error expected from an inlier or 3σ. Large residuals ($> \bar{c}^2$) are discarded. In our experimental setup, $\bar{c}$ is set to 1.

To simplify the solution, the rotation and translation are decoupled in TEASER++, as expressed by Equations (3) and (4) respectively.

$$\hat{\mathbf{R}} = \arg\min_{\mathbf{R}\in SO(3)} \sum_{k=1}^{K} \min\left(\frac{\left\|\bar{\mathbf{b}}_k - \hat{s}\mathbf{R}\bar{\mathbf{a}}_k\right\|}{\delta_k^2}, \bar{c}^2\right) \tag{3}$$

Firstly, to estimate $\hat{\mathbf{R}}$, we want to minimise the distance between corresponding points expressed as $\bar{\mathbf{b}}_k$ and $\bar{\mathbf{a}}_k$, with a bounded noise δ_k. Using the estimated rotation, the translation can then be determined by

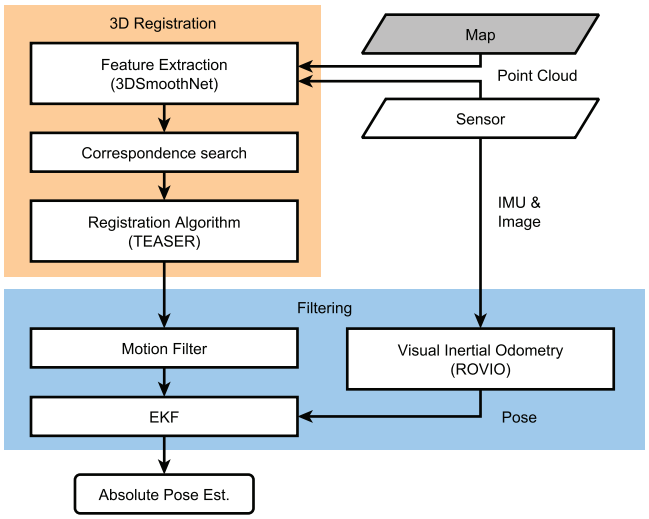

Fig. 1 *The pipeline of the proposed absolute localisation system. The main parts of the system are the feature extraction (using 3DSmoothNet), robust 3D point registration (using TEASER++), and the extended Kalman filter which fuses the pose estimate from the registration algorithm with a visual-inertial pose*

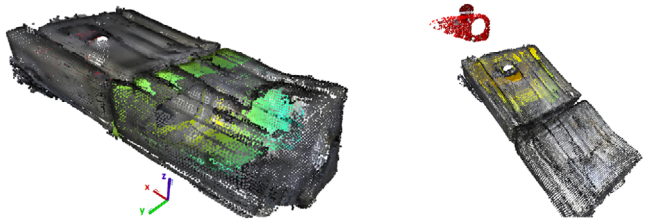

Fig. 2 *Illustration of a correct(left) and wrong(right) match with the point clouds used in the experiments, where the map is shown in grey and the sensor points in colour. The geometric similarity of repeated structured environments can cause errors when matching without prior location information. These wrong matches are avoided using simple filtering techniques*

Equation (4). The translation is acquired component-wise, i.e. the entries t_1, t_2, t_3 of $\mathbf{t}$ are computed independently. Large residuals ($> \bar{c}^2$) are, as in the previous equations, discarded:

$$\hat{t}_j = \arg\min_{t_j} \sum_{i=1}^{N} \min\left(\frac{\left(t_j - [\mathbf{b}_i - \hat{s}\hat{\mathbf{R}}\mathbf{a}_i]_j\right)^2}{\beta_i^2}, \bar{c}^2 \right) \tag{4}$$

An analytic derivation of the aforementioned formulation can be found in [11]. Via the registration, we obtain the pose of the robot in the map. The registration assumes no knowledge about previous poses and a simple motion filter and an extended Kalman filter are introduced to improve the pose estimate. To obtain a better estimate, the pose from the 3D registration is evaluated by a motion filter, which rejects movements with velocity above 0.3 m/s. The filtered pose is then used in an extended Kalman filter along with the VIO pose. The result of the Kalman filter is an absolute pose estimate in the map.

Experimental Setup: The proposed localisation system has been tested within a representative structured and confined space, specifically a mock-up model of a topside water ballast tank (WBT) onboard a marine cargo vessel. The structure of these confined spaces are often known. The localisation system can use this information in the form of a point cloud extracted from the CAD drawings, which we consider as the map. The environment, or the map, the unmanned aerial vehicle (UAV) has to localise itself in, can be seen in Figure 2. The experimental UAV was equipped with an RGB-D camera which generates RGB images and 3D point clouds at a frequency of 15 Hz. To estimate the position of the UAV, the point cloud stemming from the sensor is matched to the map point cloud. The registration is based on TEASER++ [11] and uses features generated by 3DSmoothNet [9] which utilises a deep learning

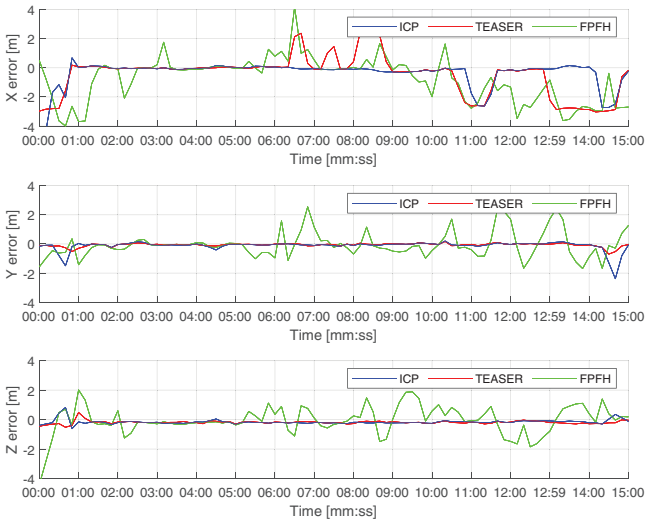

Fig. 3 *Position error for tested methods: ICP registration using: 3DSmoothNet features; TEASER++ using 3DSmoothNet features; and TEASER using FPFH features*

Table 1. The mean and standard deviation (Std.) of the Euclidean position error, in meters, of the tested methods: ICP with 3DSmoothNet features, TEASER++ with FPFH, TEASER++ with 3DSmoothNet, ROVIO, and our EKF filtered absolute localisation solution

	3DSN+ICP	FPFH+TEASER	3DSN+TEASER	ROVIO	Our
Mean	0.715	1.842	1.094	0.1724	0.102
Std.	1.138	1.304	1.153	0.0564	0.050

architecture to extract 3D features from point clouds. To compare the performance of our system, two other setups were tested. The first setup is based on the extraction of FPFH [13] features, and registration using TEASER++. To maximise the quality of the FPFH setup, features were calculated for all points in both the map and the sensor point cloud. The pose estimate from each solution was compared to a groundtruth, which was estimated using ArUco markers mounted inside the mockup-model. From the position error in Figure 3 it can be seen that the FPFH solution has large errors on all three axis and, therefore, cannot register correctly the sensor and map point clouds. To get better matches we, therefore, introduce 3DSmoothNet feature descriptors, while maintaining TEASER as the registration algorithm (shown as the red line in the Figure 3). As the last comparison setup, ICP registration was used based on the features from 3DSmoothNet(3DSN) and the points in map and sensor clouds. The registration for all tests was performed at 0.1 Hz. From Figure 3 it can be seen that both 3DSmoothNet(3DSN) with TEASER++ and the ICP test show sudden errors of around ± 2 meters on the x-axis. Due to the similarity between the two compartments of the WBT, the sensor cloud is being matched to the wrong compartment, which results in large jumps in the matched pose. An example of a correct and wrong match is shown in Figure 2. These jumps are rejected by a motion filter which rejects updates with a velocity larger than 0.3 m/s, before the estimated pose from TEASER is published to the extended Kalman filter, as shown in Figure 1. The extended Kalman filter fuses the pose from TEASER++ with a VIO estimate from ROVIO [14] using a ROS implementation of a generic Kalman filter. The VIO pose has the benefit of a fast update rate but will tend to drift over time. By fusing the accurate but slow pose from the 3D registration with the fast VIO estimate, the final accurate absolute pose can be acquired with the high update rate from the VIO system. The result of the whole localisation pipeline is shown in Figure 4, where it is clear that the position error for ROVIO increases, while the filtered EKF pose error is small. The mean and standard deviation of the tested methods can be seen in Table 1.

Computational Time: An important aspect to all localisation systems is the computational time to acquiring a pose estimate. The computational time for our proposed approach(3DSN+TEASER) is illustrated in

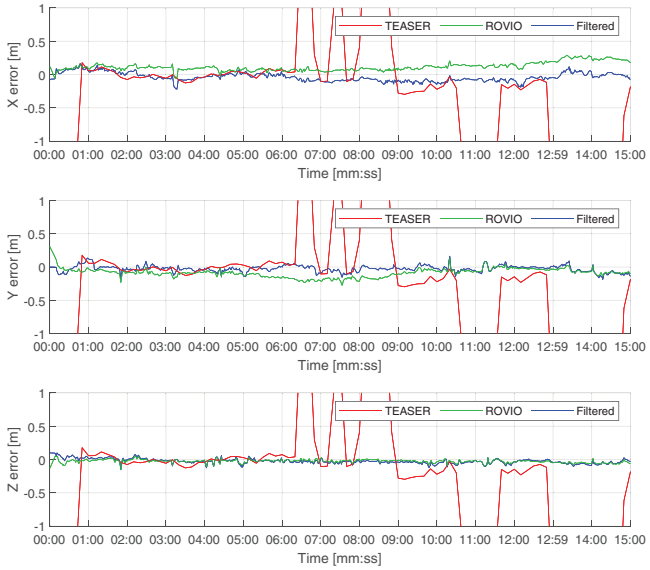

Fig. 4 *The position error of our filtered system (notated as "filtered") compared to the standalone TEASER++ using 3DSmoothNet (notated as **TEASER**), and the ROVIO pose estimate (notated as **ROVIO**). Because of ROVIO's relative localisation, an increasing error is visible on the* x *axis, but due to little movement (±*1*m) in the experiment on the* y *and* z *axis, the drift in ROVIO is limited on these two axes*

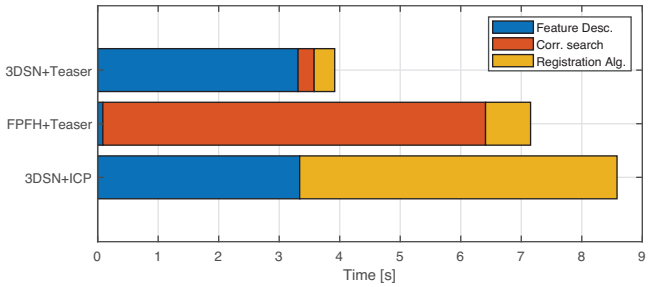

Fig. 5 *The average computational time for each element in the 3D registration methods; (i) ICP registration using 3DSmoothNet features (3DSN+ICP), (ii) TEASER++ registration using FPFH features (FPFH+TEASER), (iii) TEASER++ using 3DSmoothNet features (3DSN+TEASER)*

Figure 5 and is compared to the FPFH and ICP setups. To have a fair comparison, all the tests were conducted on a CPU, and no GPU was used for 3DSmoothNet, despite the fact that an even better performance would be expected by using a GPU. In Figure 3 it can be seen that the ICP registration with 3DSmoothNet features has fewer outliers than 3DSN+TEASER++, but this accuracy comes with a computational cost. The computational costs of the three methods are shown in Figure 5, where it can be seen that the combination of 3DSmoothNet(3DSN) and teaser is more than twice as fast as the ICP solution. It is worth noticing that the correspondence search is significantly slower with the FPFH solution since it is based on all points in the point cloud. This was necessary to get the best registrations with TEASER++ and, therefore, it provides the lowest possible error for the 3D point matches.

Conclusion: In this letter we have proposed an absolute localisation system for confined spaces using VIO and deep geometric features, which is able to accurately estimate the position of an UAV in a confined space. The experimental results in real environment indicate that our absolute localisation system outperforms standard feature detection approaches such as FPFH, and lowers the computational time from scan to registration, compared to previous methods. Furthermore, fusing the pose of the registration with a VIO estimate in an extended Kalman filter increases the robustness in similar structural environments and can increase the update-rate of the absolute pose estimation. Future improvements to the system can be achieved by the inclusion of faster deep geometric feature descriptors which will further robustify the 3D point cloud registration.

Acknowledgements: This work has been supported by the Inspectrone (Autonomous and high-level commanded system for remote inspection of marine vessels to support classification and commercial operations) project, funded by the Innovation Fund Denmark (IFD) under contract number 8090-00080B.

Received: *1 March 2021* Accepted: *8 April 2021*
doi: 10.1049/ell2.12199

References

1 Kostavelis, I., et al.: Stereo-based visual odometry for autonomous robot navigation. *Int. J. Adv. Rob. Syst.* **13**(1), 21 (2016)
2 Ramezani, M., Khoshelham, K.: Vehicle positioning in gnss-deprived urban areas by stereo visual-inertial odometry. *IEEE Trans. Intell. Veh.* **3**(2), 208–217 (2018)
3 Brogaard, R.Y., et al.: Towards UAV-based absolute hierarchical localization in confined spaces. In: 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 182–188. IEEE, Piscataway, NJ (2020)
4 Tiemann, J., et al.: Enhanced UAV indoor navigation through slam-augmented uwb localization. In: 2018 IEEE International Conference on Communications Workshops, pp. 1–6. IEEE, Piscataway, NJ (2018)
5 Boukas, E., Gasteratos, A., Visentin, G.: Towards orbital based global rover localization. In: 2015 IEEE International Conference on Robotics and Automation, pp. 2874–2881. IEEE, Piscataway, NJ (2015)
6 Boukas, E., Gasteratos, A., Visentin, G.: Introducing a globally consistent orbital-based localization system. *J. Field Rob.* **35**(2), 275–298 (2018)
7 Nalpantidis, L., Sirakoulis, G.C., Gasteratos, A.: A dense stereo correspondence algorithm for hardware implementation with enhanced disparity selection. In: Hellenic conference on Artificial Intelligence, pp. 365–370. Springer, Berlin, Heidelberg (2008)
8 Nalpantidis, L., Sirakoulis, G.C., Gasteratos, A.: Non-probabilistic cellular automata-enhanced stereo vision simultaneous localization and mapping. *Meas. Sci. Technol.* **22**(11), 114027 (2011)
9 Gojcic, Z., et al.: The perfect match: 3d point cloud matching with smoothed densities. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5545–5554. IEEE, Piscataway, NJ (2019)
10 Qi, C.R., et al.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Red Hook, NY (2017)
11 Yang, H., Shi, J., Carlone, L.: Teaser: Fast and certifiable point cloud registration. *IEEE Trans. Rob.* **37**(2), 314–333 (2020)
12 Zhou, Q.-Y., Park, J., Koltun, V.: Fast global registration. In: European Conference on Computer Vision, pp. 766–782. Springer, Berlin, Heidelberg (2016)
13 Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3d registration. In: 2009 IEEE International Conference on Robotics and Automation, pp. 3212–3217. IEEE, Piscataway, NJ (2009)
14 Bloesch, M., et al.: Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *Int. J. Rob. Res.* **36**(10), 1053–1072 (2017)