Robust Visual SLAM with Auxiliary Range Factor for Degenerate Construction Environments

Jungwoo Lee[1], Juwon Kim[1], Seung-Woo Ko[2], Inwook Shim[2] and Younggun Cho[1†]

***Abstract*— Robust localization in degenerate environments such as construction sites remains challenging due to frequent visual degradation and intermittent range sensing. To address this challenge, we propose a visual Simultaneous Localization and Mapping (SLAM) framework that tightly integrates visual-inertial odometry (VIO), visual loop closure, and Ultra-Wideband (UWB) range measurements through a factor graph optimization. The proposed system introduces an interpolated range factor, fully utilizing sparse and asynchronous UWB data by leveraging continuous-time pose interpolation. We evaluate the method on the public dataset under simulated degenerate scenarios, including visual degradation and partial UWB signal loss. Experimental results demonstrate that the proposed method significantly reduces drift and maintains accurate global localization, even under severe sensor degradation. Project page: https://sparolab.github.io/research/icra_2025/vir-construction.**

I. Introduction

Autonomous navigation in construction environments remains challenging due to frequent occlusions, dynamic obstacles, and visual degradation. Although camera-based SLAM is widely adopted for its cost-effectiveness and perceptual richness, it is highly susceptible to drift and failure in environments with sparse or unstable visual features.

To enhance robustness, VIO has been widely adopted as it fuses visual and inertial data to provide more reliable local motion estimation than visual-only systems [1]. While VIO improves short-term accuracy and stability, it still lacks the measurements required for global accuracy, eventually leading to incremental drift. Visual loop closures are often introduced to address this limitation by recognizing revisited locations and enforcing global consistency. However, loop closures heavily rely on the re-observation of scenes, which is vulnerable to limited revisits, viewpoint changes, or dynamic occlusions.

As an alternative to Global Navigation Satellite Systems (GNSS), UWB sensors have emerged in GNSS-denied environments such as indoor or urban canyons. Consequently, integrating VIO with UWB is a natural approach to alleviate drift errors. Early approaches adopted loosely coupled strategies, fusing UWB-based positions with VIO poses [2]. Although these methods are simple and modular, they typically require four-range measurements of minimum to obtain 3d position, making them vulnerable in such environments. Thus, recent studies have focused on tightly coupled frameworks that directly fuse raw range data with other modalities within a unified optimization, showing improved accuracy and robustness [3], [4], [5].

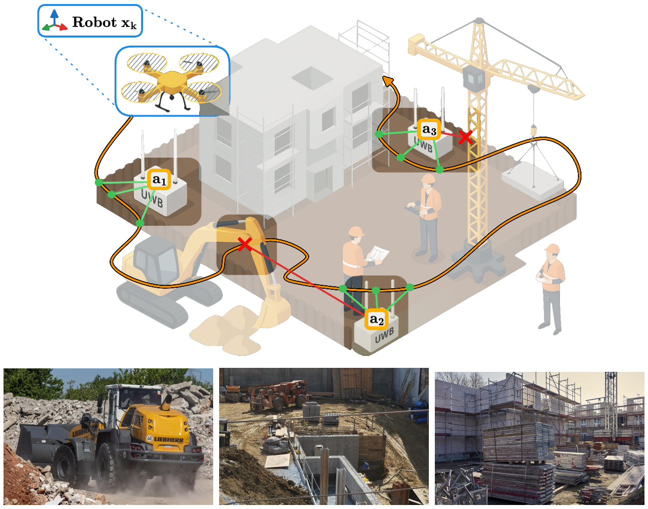


Fig. 1. Illustration of a degenerate construction environment where both visual degradation and UWB signal blockage can occur.

We present a robust visual SLAM system that integrates VIO with UWB range measurements through an interpolated range factor within a factor graph optimization framework. The proposed system is capable of operating reliably even in such challenging environments as illustrated in Fig. 1. Visual noise is prevalent, and stable UWB signal reception cannot always be guaranteed in these environments [6]. The main contributions of this work are summarized as follows:

- **A factor graph-based SLAM framework** that tightly integrates VIO poses, visual loop closures, and UWB range measurements.
- **Interpolated range factor** enables fully utilizing asynchronous and sparse UWB data through continuous-time pose interpolation.
- **Evaluation on simulated degenerate environments** using public dataset [7] shows that the proposed method suppresses large VIO drift and maintains robust localization under degenerate conditions.

II. Method

A. System Overview

As shown in Fig. 2, we fuse camera and IMU measurements to estimate robot poses using the VINS-Fusion framework [8].

This work was supported by the Starting growth Technological R&D Program(RS-2024-00448695 and RS-2024-00448642) funded by the Ministry of SMEs and Startups(MSS, Korea) and the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00448/RS-2022-II220448)

[1]Jungwoo Lee, [1]Juwon Kim and [1†]Younggun Cho are with the Dept. of Electrical and Computer Engineering, Inha University, Incheon, South Korea. [pihsdneirf, marimo117]@inha.edu, yg.cho@inha.ac.kr

[2]Seung-Woo Ko and [2]Inwook Shim are with the Dept. of Smart Mobility Engineering, Inha University, Incheon, South Korea. [swko, iwshim]@inha.ac.kr

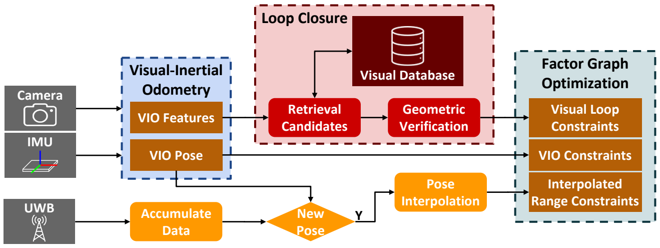


Fig. 2. Overview of the proposed SLAM pipeline.

For loop closure, we perform place recognition based on bag-of-words (BoW) scheme and estimate the relative pose with feature correspondences. In addition, we propose interpolated range factors to tightly integrate UWB measurements into the factor graph. This approach enables full exploitation and effective fusion of asynchronous range data.

B. Visual-Inertial Odometry

We adopt VINS-Fusion [8] to estimate sequential robot poses by fusing visual-inertial data. While visual odometry (VO) can also be used for this purpose, we base our system on VIO to ensure robustness in degenerate environments where visual degradation is common. The resulting relative motion estimates are incorporated into the factor graph as motion factors. The sequential VIO residual is defined as:

$$\begin{aligned} \mathbf{r}_k^{\mathrm{VIO}} &= f(\mathbf{x}_k, \mathbf{x}_{k+1}) \boxminus \hat{\mathbf{z}}_{k,k+1} \\ &= \log\left(\hat{\mathbf{z}}_{k,k+1}^{-1} \cdot f(\mathbf{x}_k, \mathbf{x}_{k+1})\right) \end{aligned} \tag{1}$$

where $\mathbf{x}_k = [\mathbf{p}_k, \mathbf{R}_k]$ is the robot's pose at camera frame k, and $\mathbf{p}_k \in \mathbb{R}^3, \mathbf{R}_k \in \mathrm{SO}(3)$ are the position and orientation of robot. $f(\,\cdot\,)$ denotes the relative pose function computed from current state variables, $\hat{\mathbf{z}}_{k,k+1}$ is the estimated relative pose between frames k and $k+1$, and the operator $\boxminus$ denotes pose difference in the tangent space of the Lie group.

C. Visual Loop Constraint

We adopt a BoW-based scheme from VINS-Mono [1]. Loop candidates are initially retrieved using DBoW2 [9]. Descriptor-based feature correspondences are then established, followed by geometric verification and relative pose estimation using 2D-2D fundamental matrix random sample consensus (RANSAC) and 2D-3D Perspective-n-Point (PnP) RANSAC.

The estimated relative transformation between frame i and loop candidate j is denoted as $\hat{\mathbf{z}}_{i,j}^{\mathrm{LC}}$, and is incorporated into the factor graph through the visual loop residual defined as:

$$\mathbf{r}_{(i,j)}^{\mathrm{LC}} = f(\mathbf{x}_i, \mathbf{x}_j) \boxminus \hat{\mathbf{z}}_{i,j}^{\mathrm{LC}} \tag{2}$$

D. Interpolated Range Constraint

We introduce an UWB range constraint to address the drift errors and limitations of visual loop closure. UWB range measurements are generally accurate under Line-of-Sight (LoS) conditions, but may be significantly biased under Non-Line-of-Sight (NLoS) due to multipath propagation and signal attenuation [6]. We assume that the measurements are acquired under LoS conditions, and apply robust kernels to mitigate the impact of occasional NLoS-induced noise. However, modeling or detecting NLoS conditions explicitly remains out of scope in this work and is left as future work.

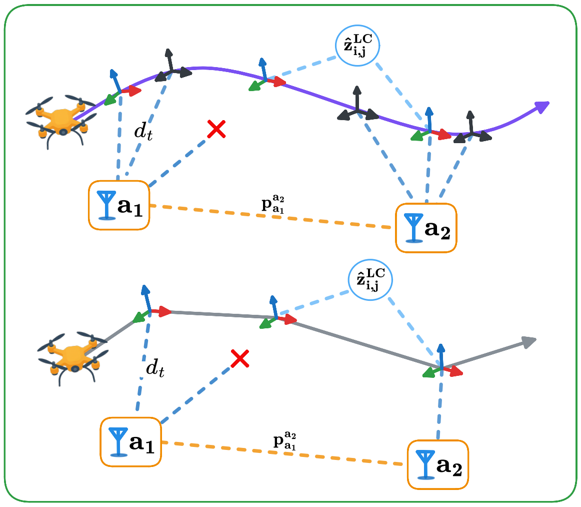

$\hat{\mathbf{z}}_{i,j}^{\mathrm{LC}}$
d_t
$\mathbf{a_1}$
$\mathbf{p}_{\mathrm{a}_1}^{\mathrm{a}_2}$
$\mathbf{a_2}$
$\hat{\mathbf{z}}_{i,j}^{\mathrm{LC}}$
d_t
$\mathbf{a_1}$
$\mathbf{p}_{\mathrm{a}_1}^{\mathrm{a}_2}$
$\mathbf{a_2}$

Fig. 3. Comparison of range factor application with (up) and without (down) pose interpolation. Continuous time pose interpolation (purple curve) increases the chance of asynchronous UWB range measurement (blue dash) utilization. In addition, increased UWB range measurement and visual loop closure (turquoise dash) leads to increased accuracy. Each UWB stations relative location are also given as a between factor (yellow dash).

At time t, the measured distance from ith anchor to the robot's antenna is modeled as:

$$d_t = \left\|\mathbf{p}_t + \mathbf{R}_t \mathbf{o} - \mathbf{a}^i\right\|_2 + b^i + \epsilon \tag{3}$$

where $\mathbf{p}_t$ and $\mathbf{R}_t$ are the robot's position and orientation at time t, respectively, $\mathbf{o}$ denotes the antenna's position in the robot's body frame, $\mathbf{a}^i$ and b^i are the position and bias of the ith anchor, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ represents Gaussian noise.

As illustrated in Fig. 3, we interpolate the robot pose at the time of each range measurement to compensate the temporal discrepancy between UWB and VIO. This enables the system to fully utilize asynchronous UWB measurements by associating them with continuous-time poses. To achieve this, the robot pose is interpolated between adjacent states $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$ as:

$$\begin{aligned} \mathbf{p}_t &= (1-s)\mathbf{p}_{k-1} + s\mathbf{p}_k \\ \mathbf{R}_t &= \mathbf{R}_{k-1} \exp\left(s \log\left(\mathbf{R}_{k-1}^{\top} \mathbf{R}_k\right)\right) \end{aligned} \tag{4}$$

where $\mathbf{x}_t = [\mathbf{p}_t, \mathbf{R}_t]$ is the robot's pose at time t, and $s = (t - t_{k-1})/(t_k - t_{k-1})$, with $t_{k-1} < t \leq t_k$.

Thus, the interpolated range residual is defined as:

$$\mathbf{r}_{(t,k,i)}^{\mathrm{UWB}} = \left\|\mathbf{p}_t + \mathbf{R}_t \mathbf{o} - \mathbf{a}^i\right\|_2 + b^i - d_t \tag{5}$$

E. Factor Graph Optimization

The sequential VIO poses, visual loop constraints, and interpolated range constraints in the factor graph are optimized using iSAM2 [10]. For all variables $\mathcal{X} = \{\mathbf{x}_0, \ldots, \mathbf{x}_k, \mathbf{a}^0, \ldots, \mathbf{a}^i, b^0, \ldots, b^i\}$ including robot poses, anchor positions, and anchor biases, we define the cost function of our system as:

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} \{\mathcal{F}(\mathcal{X})\} \tag{6}$$

TABLE I

ABSOLUTE TRAJECTORY ERROR (ATE) (M) COMPARISON OF SLAM PIPELINES ON NTU VIRAL [7] DATASET.

Sequence	VINS-Mono	VINS-Fusion	VINS-Fusion + LC	VIR-SLAM	Chao Hu el al.	DC-VIRO	Ours (w/o LC)	Ours
eee_01	1.305	0.558	0.437	1.298	0.781	0.524	**0.321**	$\underline{0.322}$
eee_02	0.854	0.640	0.331	0.443	0.678	0.382	$\underline{0.322}$	**0.320**
eee_03	1.065	0.482	0.421	0.657	0.315	0.331	$\underline{0.305}$	**0.304**
nya_01	0.915	0.510	0.492	0.869	0.604	0.412	$\underline{0.305}$	**0.302**
nya_02	0.554	0.463	0.386	0.520	**0.212**	$\underline{0.217}$	0.260	0.253
nya_03	1.445	0.841	0.664	0.761	0.513	$\underline{0.263}$	$\underline{0.253}$	**0.248**
Average	1.023	0.582	0.455	0.758	0.517	0.355	$\underline{0.294}$	**0.292**

To be specific, we incorporate the residual functions into the factor graph $\mathcal{F}(\mathcal{X})$:

$$\mathcal{F}(\mathcal{X}) = \sum_{k} \left\|\mathbf{r}_k^{\mathrm{VIO}}\right\|_{\Sigma_k}^2 + \sum_{(i,j)\in\mathbb{L}} \rho^{\mathrm{LC}}\left(\left\|\mathbf{r}_{(i,j)}^{\mathrm{LC}}\right\|_{\Sigma_{i,j}}^2\right) + \sum_{(t,k,i)\in\mathbb{U}} \rho^{\mathrm{UWB}}\left(\left\|\mathbf{r}_{(t,k,i)}^{\mathrm{UWB}}\right\|_{\Sigma_{t,k,i}}^2\right) \tag{7}$$

where $\rho^{LC}(\cdot)$ and $\rho^{UWB}(\cdot)$ are robust loss functions for visual loop constraints and interpolated range constraints, respectively. $\mathbb{L}$ and $\mathbb{U}$ denote the sets of loop closures and UWB constraints.

In our implementation, we do not estimate the robot's position based on fixed global anchor locations. Instead, we incorporate the known relative positions between UWB anchors as additional constraints in the factor graph, while treating their absolute positions as optimization variables. These inter-anchor constraints regularize the optimization and preserve spatial consistency among anchors during joint estimation.

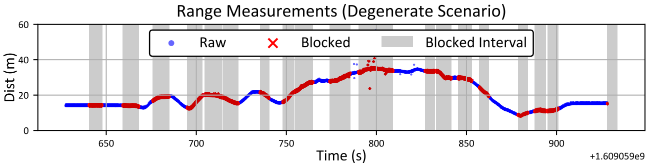
III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we conducted experiments using the NTU VIRAL dataset [7], which was collected in outdoor urban canyons and indoor environments. This dataset includes a wide range of sensor modalities such as LiDAR, stereo cameras, IMU, and UWB-based range data acquired from 3 anchors and 4 antennas.

A. Evaluation in Standard Conditions

We evaluated our method on two representative sequences from the dataset: *EEE* (urban canyon) and *NYA* (indoor). To benchmark performance, we compared our approach against the following baselines: VINS-Mono [1], VINS-Fusion (stereo VIO) [8], VINS-Fusion + LC (with loop closure), VIR-SLAM [3], Hu et al. [4], DC-VIRO [5], our method without loop closure (VIO + UWB), and our full method (VIO + LC + UWB). For evaluation, we computed the Absolute Trajectory Error (ATE) using the evo toolkit [11].

Table I presents the ATE results across all sequences. The proposed method achieves the lowest average ATE of 0.292 m, demonstrating robust localization across both indoor and outdoor environments. Notably, our system maintains relatively uniform performance, suggesting consistent behavior under varying conditions. To assess the contribution of each sensing modality, we compare the loop-closure-only approach (VINS-Fusion + LC) against the UWB-only method (ours w/o LC).



	eee_01	eee_02	eee_03	nya_01	nya_02	nya_03
total duration (s)	378.67	300.92	161.34	376.19	408.61	391.17
1-anchor blocked (s)	125.06	112.53	36.14	113.93	120.45	176.99
2-anchor blocked (s)	165.88	125.22	54.20	115.75	162.40	144.97
3-anchor blocked (s)	66.26	50.25	52.60	65.26	84.37	32.32
total msg	22126	16622	9641	17214	23905	17417
blocked msg	12720	9319	6161	9608	14427	8735
total loss rate	57.49%	56.06%	63.90%	55.82%	60.35%	50.15%

Fig. 4. Simulated UWB signal dropout for degenerate environment evaluation. (up) Example of range measurements from anchor 101 in *eee_02* sequence. (down) Summary of blocking statistics, including anchor-specific blocking duration and total message loss rate across sequences.

Loop closure effectively reduces drift in sequences with strong appearance overlap (e.g., *eee_02*), whereas UWB constraints provide more reliable improvements in sequences with limited loop detection opportunities or poor visual conditions. These results confirm that UWB is essential for accurate localization when loop closure is unreliable. Moreover, the combination of both modalities yields the most robust overall performance.

B. Evaluation under Degenerate Scenarios

To evaluate the robustness of the proposed system under challenging conditions, we simulate three types of degenerate scenarios with the NTU VIRAL dataset [7].

- **Visual degradation**: In construction environments, visual degradation frequently occurs due to airborne dust, metallic structures, and textureless surfaces. These conditions reduce feature quality and lead to significant drift in VIO. To simulate this scenario, we apply a Gaussian blur to the image frames. In the subsequent experiments, we refer to this scenario as (drift).
- **Partial UWB visibility**: Construction sites also present challenges for incorporating UWB range constraints, as signals can be blocked by structural obstacles or dynamic occlusions from moving machinery and workers. Following the approach in [6], we simulate partial UWB observability by randomly dropping range data from specific anchors for 5–10 seconds with a 1% probability per

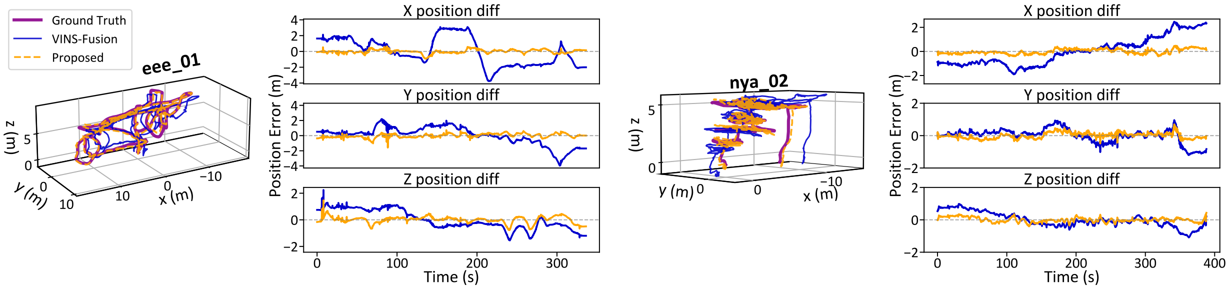


Fig. 5. Qualitative results of *eee_01* and *nya_02* sequences with degenerate scenarios (drift + drop).

TABLE II
ATE ROOT MEAN SQUARE ERROR (RMSE) (M) RESULTS FOR DEGENERATE SCENARIOS.

Sequence	VINS-Fusion	VINS-Fusion (drift)	Ours (drift)	Ours (drop)	Ours (drift + drop)
eee_01	0.558	2.112	0.331	0.347	0.423
eee_02	0.640	0.766	0.342	0.334	0.386
eee_03	0.482	1.036	0.319	0.295	0.380
nya_01	0.510	1.088	0.432	0.318	0.478
nya_02	0.463	0.825	0.263	0.289	0.258
nya_03	0.841	1.262	0.282	0.273	0.298
Average	0.582	1.181	0.328	0.309	0.371

message. This scenario is referred to as (drop) in the following experiments. Fig. 4 illustrates blocked intervals. The accompanying table reports dropout statistics, including the durations with 1-3 anchors simultaneously blocked and total loss rates ranging from 50% to 64%.

- **Combined scenario**: This is the most challenging case, with both visual degradation and UWB dropout applied simultaneously. It evaluates whether the proposed method can maintain localization performance despite degradation in multiple sensor modalities. We refer to this scenario as (drift + drop) in the following experiments.

Table II reports the ATE results under simulated degenerate scenarios. We compared the baseline VINS-Fusion with our proposed method under three conditions: visual degradation only (drift), UWB signal dropout only (drop), and both degradations simultaneously (drift + drop).

Under the visual degradation scenario (drift), VINS-Fusion shows a substantial performance drop, with average ATE increasing by about 2×. In contrast, our method (Ours (drift)) achieves significantly lower ATE, demonstrating the effectiveness of the proposed range constraints in mitigating visual drift.

When UWB signals are only degraded (drop), our system maintains strong localization accuracy despite severe signal loss conditions. As shown in Fig. 4, signal dropouts were simulated with up to three anchors blocked and total message loss rates exceeding 50%. These results demonstrate that the proposed range factor, combined with visual loop closure, provides robust performance even under intermittent UWB observability.

In the most challenging condition (drift + drop), where both visual and range sensing are degraded, our method demonstrates an average ATE of 0.371m, as shown in Table II. Although slightly worse than the individual degradation cases, it still significantly outperforms VINS-Fusion, confirming that tightly coupled multi-modal fusion ensures robust localization under severe sensor degradation. As illustrated in Fig. 5, the proposed method yields lower trajectory error than the baseline.

IV. Conclusion and Future Work

This paper presented a robust visual SLAM system integrating VIO, loop closure, and UWB measurements via an interpolated range factor within a factor graph framework. The proposed system reliably operates in challenging construction environments, characterized by visual degradation and intermittent UWB signals. Extensive experiments demonstrated a significant reduction of trajectory drift under sensor degradation conditions. To facilitate reproducibility and comparative benchmarking, we will publicly release the simulated UWB-degraded dataset and evaluation setup used in our experiments.

In future work, we plan to incorporate a dedicated anchor initialization step, as anchor positions currently rely solely on joint graph optimization. Furthermore, we aim to explicitly model and handle NLoS effects in UWB measurements to enhance robustness. Field evaluations will be conducted at real-world construction sites to validate both practicality and scalability under varying UWB anchor configurations.

References

[1] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[2] C. Wang, H. Zhang, T.-M. Nguyen, and L. Xie, "Ultra-wideband aided fast localization and mapping system," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.* IEEE, 2017, pp. 1602–1609.

[3] Y. Cao and G. Beltrame, "Vir-slam: Visual, inertial, and ranging slam for single and multi-robot systems," *Autonomous Robots*, vol. 45, no. 6, pp. 905–917, 2021.

[4] C. Hu, P. Huang, and W. Wang, "Tightly coupled visual-inertial-uwb indoor localization system with multiple position-unknown anchors," *IEEE Robot. and Automat. Lett.*, vol. 9, no. 1, pp. 351–358, 2023.

[5] S. Jia, R. Xiong, and Y. Wang, "Distributed initialization for visual-inertial-ranging odometry with position-unknown uwb network," in *Proc. IEEE Intl. Conf. on Robot. and Automat.* IEEE, 2023, pp. 6246–6252.

[6] R. Maalek and F. Sadeghpour, "Accuracy assessment of ultra-wide band technology in tracking static resources in indoor construction scenarios," *Automation in Construction*, vol. 30, pp. 170–183, 2013.

[7] T.-M. Nguyen, S. Yuan, M. Cao, Y. Lyu, T. H. Nguyen, and L. Xie, "Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint," *Intl. J. of Robot. Research*, vol. 41, no. 3, pp. 270–280, 2022.

[8] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," 2019. [Online]. Available: https://arxiv.org/abs/1901.03642

[9] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, 2012.

[10] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "isam2: Incremental smoothing and mapping using the bayes tree," *Intl. J. of Robot. Research*, vol. 31, no. 2, pp. 216–235, 2012.

[11] M. Grupp, "evo: Python package for the evaluation of odometry and slam." https://github.com/MichaelGrupp/evo, 2017.