Towards the determination of an algorithm for Simultaneous Localization and Mapping in nuclear facilities

Maugan Michel, Yoann Moline, Gwenolé Corre, Frédérick Carrel

Towards the determination of an algorithm for Simultaneous Localization and Mapping in nuclear facilities

Andréa Macario Barros, Maugan Michel, Yoann Moline, Gwenolé Corre and Frédérick Carrel

***Abstract*—Establishing an accurate radiological mapping is an essential step for the Dismantling & Decommissioning (D&D) processes since it allows the localization of possible hotspots. Nowadays, this work is usually done manually by operators who construct matrices and place each measured value in its respective position. However, this method presents many drawbacks, leading the literature to increasingly explore alternative solutions, such as combining radiological measurements with SLAM techniques (where SLAM stands for Simultaneous Localization and Mapping). SLAM technology allows the simultaneous mapping of the surroundings and indoor localization of a sensor without needing a GPS (usually inoperative in indoor nuclear facilities). Nonetheless, the current solutions based on SLAM algorithms are not suited to different nuclear measurement devices, are often bulky, and require post-processing. To address this matter, we started the development of a modular device for online 3D environment reconstruction and radioactivity measurement localization. This work presents the selection of the most suitable SLAM algorithm for the operation of nuclear installations under dismantling. As far as we know, this is the only study in the literature that evaluates the different SLAM algorithms regarding the nuclear facilities' characteristics. To address this issue, we established comprehensive state-of-the-art visual-sensor-based SLAM algorithms. Then, we critically assessed the algorithms from the literature regarding the characteristics present in dismantling environments. We selected and implemented five: Direct Sparse Odometry, Visual-Inertial Direct Sparse Odometry, Large Scale Direct Monocular SLAM, Semi-direct Visual Odometry, and Visual Inertial Semi-direct Visual Odometry. In order to compare them, we constructed a dataset not found in the literature representing the radiological mapping operations in nuclear facilities. Our dataset relies on stereo images from two cameras in a stereo configuration, inertial data from an Inertial Measurement Unit, and spherical information from a Lidar. This dataset allowed us to benchmark the algorithms considering algorithms' tracking and mapping accuracies. The Visual Inertial Semi-direct Visual Odometry presented the lowest average errors for the tracking and an equivalent performance as the other algorithms for the mapping, being the most suitable algorithm for the dismantling operation.**

***Index Terms*—dismantling, radiological mapping, simultaneous localization and mapping, SLAM**

I. Introduction

RADIOLOGICAL MAPPING plays a crucial role in dismantling and decommissioning nuclear installations. It involves determining the facilities' radiological condition, which helps identify potential contamination areas. In France, most radiation-protection operators (RPOs) still rely on manual radiological mapping, which is time-consuming, exposes them to potentially contaminated environments, and is prone to human errors. This is mainly because the existing literature lacks a system that can reliably perform radiological mapping during the dismantling of nuclear installations, considering the French facilities' characteristics and the procedures followed by RPOs in various scenarios and nuclear measurements.

As an alternative to manual radiological mapping, researchers have increasingly explored the combination of radioactivity monitors with Simultaneous Localization and Mapping (SLAM) algorithms. For instance, [1] developed a system that enables real-time 3D modeling of the environment, gamma-ray imaging, and localization. Their system incorporates LiDAR, RGB, and Compton camera technologies. They created a handheld device capable of performing 3D reconstruction and localization in indoor and outdoor environments. Additionally, their system can generate gamma-ray 3D images. However, the system they built weighs 6 kg, which poses a significant challenge for long dismantling procedures, unlike the maximum of 3 minutes considered by the authors. Moreover, their solution does not address real-case scenarios found in certain French dismantling sites, which will be depicted later.

A handheld imaging system for localizing gamma-ray sources in a 3D reconstructed environment was developed by [2]. They combined gamma-ray imaging with visual data to implement the SLAM algorithm to achieve this. They utilized a Kinect sensor, a High-Efficiency Multi-mode Imager (HEMI), a control tablet, and a notebook for processing the SLAM algorithm. Their system allowed for dense colored 3D reconstructions while integrating gamma-ray localization. Regarding the main drawbacks, their system weighed 3.6 kg and had large dimensions, making it unsuitable for prolonged operations. Additionally, they selected a sensor unsuitable for reconstructing certain nuclear environments, such as laboratories equipped with reflective surfaces and telemanipulators.

In summary, previous studies have led to the developing of portable systems capable of simultaneously localizing radioactivity measurements and providing contextual information. However, many systems are not compact, hindering their use in long-duration operations for mapping large facilities. In addition, the presented references are based only on gamma imaging measurements; therefore, remote measurements of hot spots, not including other types of nuclear measurements, are used as contact measurements for contamination detec-

Andréa Macario Barros, Maugan Michel, Yoann Moline, Gwenolé Corre and Frédérick Carrel are with Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France (e-mail: andrea.barros@cea.fr, maugan.michel@cea.fr, yoann.moline@cea.fr, gwenole.corre@cea.fr, frederick.carrel@cea.fr).

tion. Furthermore, SLAM algorithms have their performance dependent on different scene aspects, such as the presence of reflective and reflexive surfaces. This feature raises the necessity of selecting a SLAM algorithm robust to the environment constraints. The presented works do not perform such selection or consider characteristics imposed by the French nuclear installations.

In this context, for the first time in the literature, this study selects and benchmarks SLAM algorithms for 3D localization and reconstruction of real-case scenarios of nuclear installations under dismantling.

This paper is organized as follows: Section II describes the challenging characteristics of French dismantling facilities that hinder the well-functioning of certain sensors and SLAM algorithms and a theoretical selection of algorithms in the literature robust to those characteristics. Section III describes the construction of the first nuclear dismantling dataset in the literature, which enables the benchmarking of different SLAM algorithms. Section IV depicts the metrics adopted in this paper for the algorithms' evaluation. Section V shows the results obtained for determining the most suitable algorithm for localization in French nuclear facilities. Finally, Section VI highlights our conclusions.

II. Nuclear Dismantling Facilities

In order to employ a SLAM algorithm for reconstruction and localization, it is necessary to characterize the scene and the application, as some scenarios may have specific attributes that directly affect the performance of the algorithms [3]. Considering this scenario dependency, four criteria were defined, which depend on the actual condition of several representative nuclear facilities under dismantling. These criteria have the potential to influence the effective functioning of certain SLAM algorithms significantly; they are:

- *Texture absence*: this criterion was selected considering the techniques employed by many SLAM algorithms, especially the feature-based ones. For instance, the Harris corner detector [4], employed to identify the features in an image, the lack of texture can lead to bad algorithm functioning. Some examples of algorithms based on texture detection are ORB-SLAM [6] and OKVIS [5].
- *Environment Size*: the size of a nuclear facility can be important for some algorithms because it affects how much data needs to be processed and what sensors are needed. For example, the KinectFusion [7] algorithm can only be used in room-sized spaces, while the LSD-SLAM [8] algorithm can be used in large-scale areas. Nuclear facilities can range from a single room to a large contaminated building.
- *Reflective and Transparent Surfaces*: some algorithms, especially those based on RGB-D and LiDAR sensors, may have difficulty mapping reflective and transparent surfaces. Nuclear facilities often contain reflective metal and glass equipment, and contaminated facilities in France may be isolated with glass and transparent materials or contain gloveboxes.
- *Luminosity variations*: some SLAM algorithms, especially those that rely only on luminosity information, may have difficulty mapping in low-light conditions. For instance, some visual-based algorithms, such as DTAM [9], LSD-SLAM, and SVO [10], use the intensity of pixels to estimate depth, meaning they are particularly susceptible to problems caused by low light.

When choosing a SLAM algorithm for radiological mapping in a decommissioning scenario, it is important to consider algorithms that are robust to the challenging conditions of these environments, such as large, textureless spaces with reflective and transparent objects and changes in lighting. Following, we present a theoretical analysis considering the nuclear facilities' characteristics and the main SLAM algorithms [3] and sensors.

A. Comparison of SLAM Sensor technologies for nuclear dismantling facilities

SLAM algorithms may process data from one or more sensors to localize it within a 3D reconstruction. The main employed sensors are:

- LiDAR (Light Detection and Ranging): LiDAR sensors emit laser beams and measure the time it takes for the laser to bounce back from objects in the environment. By analyzing the reflected laser beams, LiDAR sensors can create accurate 3D point cloud representations of the surroundings. LiDAR data is commonly used in SLAM algorithms for precise mapping and localization. However, due to its functioning physics, LiDAR is not ideal for reconstructing dismantling facilities with reflexive and transparent surfaces since this feature significantly changes the reflected laser beam.
- Inertial Measurement Unit (IMU): IMUs are devices able to measure acceleration, angular velocity, and orientation. IMU data helps estimate the device's motion and is often fused with other sensor data to improve position tracking and motion estimation in SLAM algorithms. However, it must be combined with other types of sensor, such as cameras, to allow the 3D reconstruction and increase the system calibration's complexity.
- Cameras: cameras capture visual information from the environment and can be used for various purposes in SLAM. Visual SLAM algorithms leverage camera images to extract visual features, track their motion across frames, and estimate the device's position and environment structure. Cameras can provide rich visual information but may be sensitive to lighting conditions and textureless environments, depending on the employed algorithm.
- Depth Sensors: usually combined with RGB cameras, depth sensors are typically designed to capture depth information by measuring the time-of-flight or structured light patterns of the infrared (IR) sensor, which bounces off objects and then returns to the sensor. As a result, they are generally unable to capture depth information for transparent or highly reflective surfaces because most of the IR light passes through or bounces away from the surface without being reflected to the sensor. In addition, they present a limited range that may vary from 0.2 m to 10 m (usually less than 10 m) [11]. Those reasons limit

the use of this kind of sensor in large environments and reconstructing transparent and reflexive surfaces.

- Global Positioning System (GPS): GPS receivers provide global positioning and timing information. While GPS signals are widely used for outdoor localization, they have limitations in indoor or urban environments where the signals may be weak or obstructed, such as an indoor dismantling scene.

Considering the application of this paper, the only sensors able to operate in facilities with the presented characteristics are the cameras and IMU. For this reason, this study limits the algorithms' benchmarking to visual or visual-inertial SLAM algorithms.

B. Selection of the visual-SLAM algorithms for nuclear facilities

We could analyze the robustness of the main visual SLAM algorithms according to the nuclear facilities' conditions. A recent review [3] assembled the main visual and visual-inertial algorithms. Considering the algorithms presented by the authors, Tables I and II summarize the theoretical analysis realized for the presented SLAM techniques. In these Tables, *Textureless* refers to the robustness to uniform environments, *Size* refers to the ability of the algorithm to perform the construction of large-scale maps, *Refle./Transp. Surf.* criteria evaluate the algorithm's robustness to reflective and transparent surfaces and *Lum. Const.* refers to the performance of the algorithms under luminosity variation. They were classified in *Good* (✓), *Average* (-), and *Bad* (✗) according to their theoretical characteristics presented in the papers.

Table I details the assessment of visual-only SLAM techniques. Regarding their ability to handle textureless environments, feature-based methods yielded unsatisfactory results due to their heavy reliance on visual textures in the surroundings. Hybrid and direct approaches, as highlighted by [9], exhibited an inferior performance in textureless environments, but they remained functional—a contrast to feature-based algorithms. The only monocular algorithm capable of dealing with featureless environments is CNN-SLAM, which relies on depth predictions via convolutional neural networks (CNNs). However, regarding the *Size* criterion, CNN-SLAM and DTAM faced challenges, as they were primarily designed and tested for small-scale spaces, demanding significant computational resources. MonoSLAM also did not address this criterion, as it operates based on the Kalman Filter, and its complexity scales with the environment's size. As for the PTAM algorithm, it was originally designed for augmented reality and confined workspaces, which limits its effectiveness in larger environments.

Feature-based algorithms exhibit resilience to changes in lighting conditions when they can gather sufficient visual data to recognize distinctive features. In this regard, the SVO, LSD-SLAM, and DSO algorithms were assessed to deliver an *Intermediate* performance level, as they consider both photometric and geometric data during estimation. However, the DTAM algorithm stood out as the only algorithm with poor performance regarding this criterion, relying exclusively on photometric errors to estimate depth information.

TABLE I: Theoretical evaluation of visual-only based SLAM techniques.

Method	Texture	Size	Reflective Transparent	Luminosity
PTAM [17]	✗	✗	✓	✓
DTAM [9]	-	✗	✗	✗
SVO [10]	-	✓	-	-
LSD-SLAM [8]	-	✓	-	-
ORB-SLAM 2.0 [6]	✗	✓	✓	✓
CNN-VSLAM [18]	✓	✗	✓	✓
DSO [19]	-	✓	-	-

Table II summarizes the evaluation based on considered criteria in the context of visual-inertial-based algorithms. Since most of these algorithms rely on feature extraction, their ability to reconstruct textureless environments is constrained, even though they exhibit greater resilience than visual-only-based algorithms, as they can maintain motion tracking even when visual information is limited. An exception to this pattern is the VI-DSO algorithm, which, like DSO, primarily focuses on minimizing photometric errors for reconstruction. While most algorithms have demonstrated strong performance in large-scale environments, exceptions include MCSKF and ROVIO, whose complexity is directly influenced by the size of the scenario and the number of features. All algorithms were theoretically expected to perform similarly to the last two parameters. Even though VI-DSO is based on minimizing photometric error, incorporating IMU data and coefficients accounting for lighting variations in the optimization enhances the algorithm's robustness in low-light situations and when dealing with reflective surfaces. As for the visual-only feature-based methods, the other algorithms can operate effectively in low-light conditions.

TABLE II: Theoretical evaluation of visual-inertial based SLAM techniques.

Method	Texture	Size	Reflective Transparent	Luminosity
MSCKF [20]	✗	✗	✓	✓
OKVIS [5]	✗	✓	✓	✓
ROVIO [21]	✗	✗	✓	✓
VINS [22]	✗	✓	✓	✓
VI-DSO [23]	-	✓	✓	✓

The presented theoretical assessment has enabled the selection of algorithms well-suited for the specific conditions imposed by nuclear facilities under dismantling. These selected algorithms include SVO, LSD-SLAM, DSO, VI-SVO, and VI-DSO. Figure 1 provides a visual synthesis of the information presented in Tables I and II. In this representation, each set contains the algorithms that exhibit robustness to a particular characteristic of nuclear facilities. For instance, the DTAM algorithm demonstrates robustness in textureless environments but is not resilient to large-scale facilities, reflective or transparent surfaces, and variations in lighting conditions. Therefore, it is exclusively placed in the *Textures* set.

Nonetheless, it is necessary to benchmark the selected algorithms on representative D&D datasets to reinforce the presented theoretical selection. However, no dataset in the literature currently represents a nuclear facility under disman-

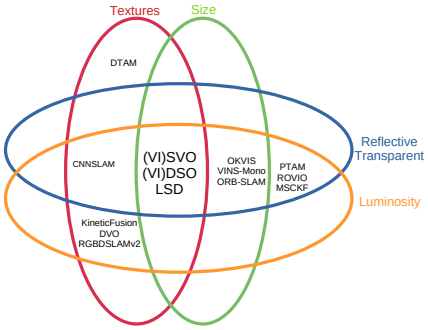


Figure 1: Set diagram representing the selected algorithms that are theoretically robusts to the conditions imposed by the decommissioning nuclear facilities.

Figure 2: Conceived system to data acquisition in real-case scenarios of dismantling environments.

tling. Hence, there is a need to conceive a new dataset of nuclear facilities under dismantling for benchmarking visual and visual-inertial SLAM algorithms.

III. Dataset construction

In order to obtain a practical and real-case evaluation of the selected algorithms, we constructed a dataset one of its kind in CEA Fontenay-aux-Roses facilities. This CEA site has several installations under dismantling, presenting the different characteristics mentioned earlier. The data was acquired based on the hardware synchronized acquisition system shown in Figure 2. It comprises two cameras Blackfly S [14] synchronized with one IMU Xsens MTi-610 [15]. In addition, a LiDAR Velodyne VLP-16 [16] provides additional data representing the 3D reconstruction ground truth. The system was mounted in a NuCoMo-100 $\alpha - \beta$ contamination monitor. The radiological data was not acquired in the context of the work presented in this article since the main objective is selecting the most suited SLAM algorithm. So, the contamination monitor is mostly used here to illustrate the concept of employment. A laptop processed the data with Intel Core I7 10^{th} generation, 32 Go of RAM.

The algorithms were benchmarked considering three different sequences with different levels of difficulty to track. They are presented in the following subsections.

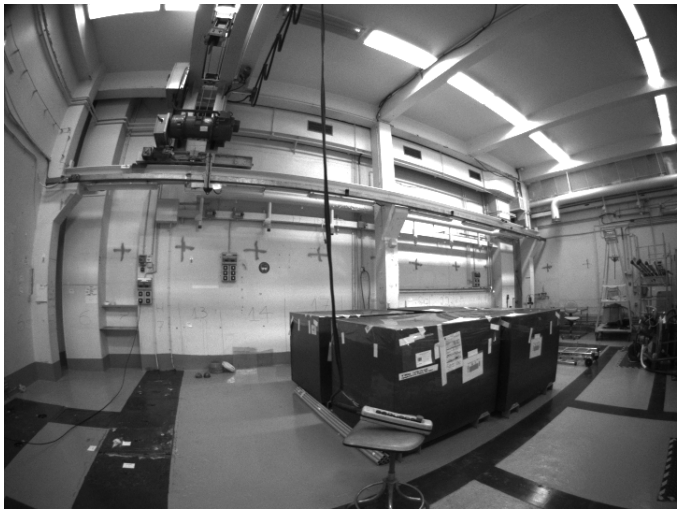


Figure 3: General view of the easy-to-track sequence.

Figure 4: General view of the medium-to-track sequence.

A. Easy-to-track sequence

The *easy-to-track* sequence was recorded in a large-size contaminated environment, and it may be considered an easy-to-track sequence since it contains different types of elements, being a texture-rich environment. It does not contain reflexive, transparent surfaces nor luminosity variation. A general view of this sequence scenario is shown in Figure 3.

B. Medium-to-track sequence

The *medium-to-track* sequence is considered a medium-difficult sequence to track. It was recorded in the same environment as the *easy-to-track*, but with a forced lightning variation, which increased the sequence difficulty to track. It is a texture-rich environment and does not contain reflexive or transparent surfaces, as presented in Figure 4.

C. Difficult-to-track

Finally, the last recorded sequence is a difficult-to-track environment since it is an isolated and contaminated installation.

Figure 5: General view of the difficult-to-track sequence.

This facility is isolated with a translucent material and poor in texture presence. It counts with good luminosity conditions and is equivalent to a small room (4.85m×2.87m), as shown in Figure 5.

IV. Evaluation criteria

This section presents the metrics used to compare the selected SLAM algorithms regarding their tracking and mapping performances in nuclear facilities under dismantling. Usually, the algorithms' evaluation is based on the error estimation over the estimated trajectory; this error is computed based on a ground-truth trajectory from a Motion Capture (MoCap) device. As the MoCap device was unavailable, this paper proposes alternative error metrics. We used the metrics proposed in [12] for the tracking evaluation: the loop drift estimation and alignment error evaluations. Additionally, two metrics are proposed to evaluate the 3D reconstructions: the evaluation of the reconstructed wall distances compared to the actual wall distances and the comparison between the 3D reconstruction and the 3D reference point cloud from the LiDAR.

A. Loop drift

The recorded sequences started and finished at the same position, enabling the computation of the loop drift error. This error refers to the drift accumulated during the sensor movement. Some algorithms can perform loop-closing techniques, identifying the images previously detected by the algorithm to estimate and correct the accumulated drift. For those algorithms, as affirmed in [12], the loop closure detection must be disabled to enable this evaluation. The computing of the Euclidean distance between the first and end points of the estimated trajectory provides the loop drift value.

B. Alignment Error

Another error metric proposed in [12] allowing the evaluation of the scale, rotation, and translation errors is the *alignment error* computation. The authors propose the sequences recording starting and finishing by easy-to-track frames enabling the reliable tracking by a SLAM algorithm well-established in the literature; the authors in [12] employed the LSD-SLAM, and we employed the ORB-SLAM. Initially, the authors propose the computation of two relative transformations, T_s^{gt} and T_e^{gt} for the start- and end-segments, respectively, and expressed by equations (1) and (2).

$$T_s^{gt} := \underset{T \in Sim(3)}{\arg\min} \sum_{i \in S} (T p_i - \hat{p}_i)^2, \tag{1}$$

$$T_e^{gt} := \underset{T \in Sim(3)}{\arg\min} \sum_{i \in E} (T p_i - \hat{p}_i)^2. \tag{2}$$

Considering the tracking of frames 1 to n with the tracked positions $p_1 ... p_n \in \mathbb{R}^3$, $S \subset [1; n]$ and $E \subset [1; n]$ correspond to the frames indices of the start- and end-segments containing an aligned ground-truth position $\hat{p} \in \mathbb{R}^3$. From equations (1) and (2) computations, the authors propose the estimation of the accumulated drift by the computation of $T_{drift} = T_e^{gt}(T_s^{gt})^{-1} \in Sim(3)$, and propose the alignment error computation. A combined error metric considering the error caused by the scale, rotation, and translation over the full trajectory. This error is expressed by equation (3):

$$e_{align} := \sqrt{\frac{1}{n} \sum_{i=1}^{n} ||T_s^{gt} p_i - T_e^{gt} p_i||_2^2} \tag{3}$$

C. 3D reconstruction errors

In addition to the two tracking evaluation metrics, we introduce two supplementary metrics to appraise the 3D reconstruction performance. The first metric assesses the reconstructed walls' precision compared to their actual dimensions. This measurement gauges the SLAM algorithm's ability to reconstruct the environment, and it is important to note that we adjust the scale of the reconstructed trajectory for monocular algorithms to enable this evaluation. The second metric for 3D reconstruction evaluates the disparity between randomly selected points in the reconstructed point cloud and the corresponding points in the reference LiDAR point cloud. It is important to highlight that this latter metric is exclusively applied in evaluating easy-to-track and medium-to-track sequences, as LiDAR cannot effectively operate in reconstructing the difficult-to-track sequences.

V. Results

This section compares the visual and visual-inertial SLAM algorithms to determine the most suitable one for 3D reconstruction and localization in nuclear facilities under dismantling. Therefore, we implemented the selected algorithms (SVO, DSO, LSD-SLAM, VI-DSO, and VI-SVO) and compared them regarding the criteria presented in section IV and real-case scenario dataset, described in section III.

It is worth noting that the results for LSD-SLAM are omitted due to its inability to track and map the recorded sequences successfully. LSD-SLAM requires a high frame rate for accurate performance, with the authors suggesting a minimum of 30 frames per second (fps). The employed 10

TABLE III: Average alignment error e_{align} and standard deviation σ obtained by the selected SLAM algorithms for the trajectory estimation of the conceived dataset.

$e_{align} \times 10^{-2}$ ($\sigma \times 10^{-2}$)	DSO	VI-DSO	SVO	VI-SVO
Easy-to-track	14.84 (0.58)	5.20 (0.53)	4.25 (0.90)	**1.83 (0.48)**
Medium-to-track	1.75 (0.42)	2.14 (0.64)	5.92 (0.51)	**0.70 (0.26)**
Difficult-to-track	-	-	-	**2.59 (0.15)**

TABLE IV: Average RMSE error and standard deviation in meters of the loop drift estimation obtained by the selected SLAM algorithms for the localization of the conceived dataset.

RMSE(σ)	DSO	VI-DSO	SVO	VI-SVO
Easy-to-track	1.22 (0.12)	2.82 (0.32)	**0.41 (0.25)**	0.74 (0.28)
Medium-to-track	0.83 (0.06)	1.12 (0.38)	0.39 (0.18)	**0.25 (0.09)**
Difficult-to-track	-	-	-	**0.15 (0.05)**

fps frame rate results in fast inter-frame motion, consequently, substantial displacements between frames, increasing the difficulty for LSD-SLAM to identify matching features in the images, leading to errors in the computation of the Jacobian matrix (utilized for minimizing the reprojection errors). Consequently, the algorithm may fail to converge towards an accurate solution, resulting in an inaccurate map.

A. Alignment error evaluation

This section conducts a performance comparison of the selected SLAM algorithms, focusing on the alignment error. This metric has proven its reliability to assess the effectiveness of SLAM algorithms, as it provides a more comprehensive evaluation of the algorithm's accuracy compared to measurements such as translational drift or joint RMSE, as highlighted in [12]. Table III displays the average alignment errors the selected SLAM algorithms obtained over ten estimations for the three conceived sequences.

As observed, the two monocular algorithms exhibited the highest average alignment error across all sequences, and they faced a complete tracking failure in the last sequence. This is mainly due to their susceptibility to scale error since they cannot directly measure distances to objects in the scene due to the absence of absolute scale. Furthermore, they demonstrated a significant sensitivity to outliers, which decreases the algorithms' performance in terms of alignment error. Despite the first sequence being relatively easy to track, it contains frames with a radiation protection operator at work, which can be regarded as a dynamic outlier object. While this did not affect most algorithms' functionality, it contributed to increased errors compared to other sequences, emphasizing the DSO algorithm's lack of robustness compared to the other considered algorithms.

On the other hand, VI-DSO demonstrated more resilience to outliers than its monocular counterpart, as evidenced by the results in the easy-to-track sequence. However, their performances, in terms of alignment error, became comparable when processing the medium-to-track dataset. Additionally, VI-DSO failed to completely track the challenging-to-follow sequence.

Finally, VI-SVO emerges with the slightest average alignment error across all sequences, indicating that it stands out as the most accurate SLAM algorithm for trajectory estimation within the conceived dataset. Beyond the average alignment error, Table III provides the standard deviation of the alignment error. This data serves as a measure of the SLAM algorithms' resilience to noise and other disruptions. The results reinforce that VI-SVO exhibits the highest level of robustness among the SLAM algorithms for trajectory estimation regarding the alignment error.

B. Loop drift evaluation

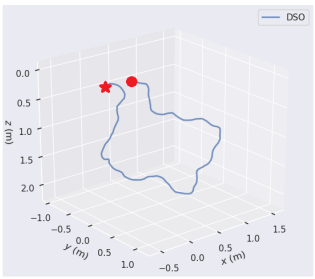
Loop drift stands as a crucial metric in the assessment of SLAM algorithms. A low loop drift indicates an SLAM algorithm's ability to accurately monitor the camera's motion over an extended duration, significantly influencing the accuracy of the estimated 3D structure of the environment. In contrast, a significant loop drift leads to inaccuracies in the estimated 3D structure of the environment. Moreover, loop drift can impact the resilience of a SLAM algorithm, as a substantial loop drift can increase the likelihood of losing track of the camera's position.

Hence, loop drift is a fundamental criterion for evaluating SLAM algorithms. This section provides an evaluation of the selected algorithms concerning loop drift. Table IV presents the average loop drift estimation from ten runs for tracking within the three considered sequences using the DSO, SVO, VI-DSO, and VI-SVO algorithms.

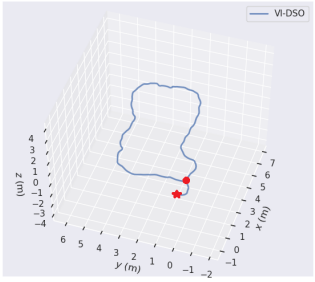
Table IV displays the average root mean square error (RMSE) and standard deviation for loop drift estimation for the DSO, VI-DSO, SVO, and VI-SVO algorithms across the recorded sequences. The results indicate that VI-SVO consistently achieves the lowest RMSE and standard deviation in most sequences, implying its superior accuracy regarding loop drift estimation. Monocular SVO is the second most accurate algorithm, followed by DSO and VI-DSO.

A primary reason justifying the enhanced accuracy of VI-SVO and SVO compared to DSO and VI-DSO is the different employed image alignment techniques. The latter two employ traditional image alignment between the current frame and the previous keyframe. At the same time, SVO and VI-SVO align the camera poses with the map rather than prior frames. This approach helps mitigate drift, notably enhancing loop drift estimation accuracy, particularly in challenging scenarios such as low-light conditions or high-speed motion.

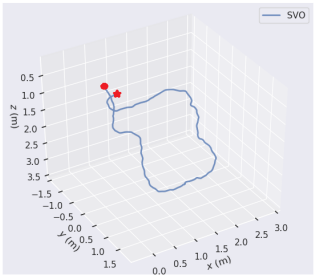
Furthermore, the initialization method used in the DSO/VI-DSO algorithms requires more processing time than SVO/VI-SVO, which results in the loss of initial frames and inertial data for tracking. DSO and VI-DSO employ a two-step initialization procedure where they first estimate the camera's pose and a sparse set of 3D points in the scene. Subsequently, these algorithms utilize joint optimization to refine the estimations of the camera pose and 3D points. In contrast, SVO and VI-SVO employ a one-step initialization process where they estimate the camera pose and a sparse set of 3D points through stereo matching, or in the case of monocular configuration, by



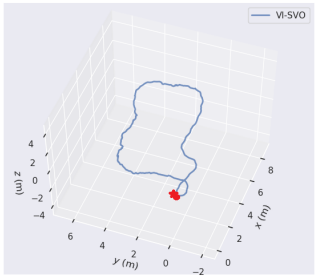
(a) Trajectory estimated by the DSO algorithm.



(b) Trajectory estimated by the VI-DSO algorithm.



(c) Trajectory estimated by the SVO algorithm.



(d) Trajectory estimated by the VI-SVO algorithm.

Figure 6: Estimated trajectories for the medium-to-track sequence by the DSO, VI-DSO, SVO and VI-SVO algorithms.

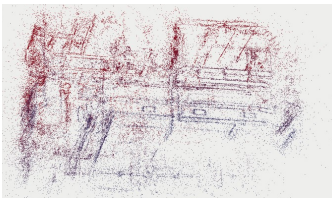
estimating the camera's pose through solving linear equations and matching 2D-3D correspondences.

Figure 6 illustrates the estimated trajectories for the medium-to-track sequence, highlighting the loop drift resulting from the initialization process. In order to ease the visualization, a red dot indicates the trajectory's start, and a red star indicates the end. The presented results suggest the accuracy and robustness of VI-SVO compared to the other implemented algorithms.

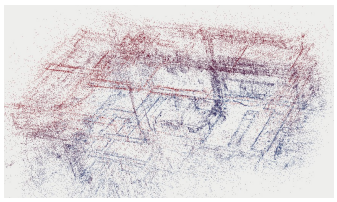
C. Mapping evaluation

This section compares the selected SLAM algorithms regarding their 3D reconstruction capabilities. The resulting reconstructions were assessed by contrasting them with the actual measurements derived from the 2D plans of the facilities. A lower error means a more consistent reconstruction, and this metric holds significance for assessing the quality of 3D reconstructions. A 3D reconstruction aligned with the real-world environment is more likely accurate and valuable for localizing nuclear measurements in dismantling facilities. Figure 7 illustrates the resulting 3D reconstructions for the medium-to-track sequence processing.

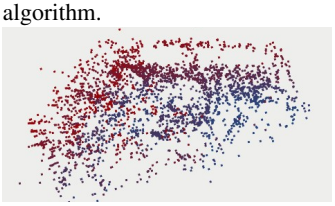
To assess the 3D reconstructions, we have introduced a metric known as the reconstruction error ratio (e_{rec}). This metric involves calculating the ratio between the height and width of the 2D plan compared to the estimated height and width ratio. This metric aims to evaluate the consistency of the 3D reconstruction without factoring in scale errors, which are predominantly observed in monocular algorithms. It relies on the ratio between measurements, making it insensitive to



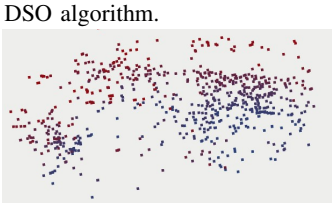
(a) 3D reconstruction by the DSO algorithm.



(b) 3D reconstruction by the VI-DSO algorithm.



(c) 3D reconstruction by the SVO algorithm.



(d) 3D reconstruction by the VI-SVO algorithm.

Figure 7: 3D reconstructions obtained for the medium-to-track sequence by the DSO, VI-DSO, SVO and VI-SVO algorithms.

TABLE V: Average reconstruction error ratio e_{rec} and standard deviation σ obtained by the selected SLAM algorithms for the 3D reconstruction of the conceived dataset.

$e_{rec} \times 10^{-2}$ ($\sigma \times 10^{-2}$)	DSO	VI-DSO	SVO	VI-SVO
Easy-to-track	**2.88 (2.3)**	4.55 (2.38)	10.15 (6.65)	4.37 (3.61)
Medium-to-track	5.96 (2.45)	3.90 (2.44)	9.78 (7.35)	**3.66 (2.42)**
Difficult-to-track	-	14.98 (8.07)	-	**7.12 (6.94)**

scale distortions and ensuring a fair assessment across the different algorithms. The estimated error values are presented in Table V.

As one may observe, DSO and VI-SVO produced the lowest values for the reconstruction error ratio, with DSO excelling in the first sequence and VI-SVO outperforming in the last two sequences. Despite the partial processing of the difficult-to-track sequence by VI-DSO due to its failure to complete the entire sequence, we could evaluate the partial reconstruction based on the initial frames. This partial reconstruction contained several outliers, leading to an error ratio value exceeding the average.

Based on the presented results, we conclude that the VI-SVO algorithm obtained the best results for the 3D reconstruction criterion in the specified dataset. It consistently produces the lowest reconstruction error ratio in most sequences. The DSO and VI-DSO provided e_{rec} similar to the VI-SVO, yet this last one demonstrates superior consistency regarding the average reconstruction error ratio. The SVO presented the highest e_{rec} values.

In addition to e_{rec}, we propose an additional metric for 3D reconstruction evaluation, the average reconstruction error e_{ref}, based on the reference point cloud from the LiDAR device. Each algorithm was run on the easy- and medium-to-track sequences five times, and we compared the resulting 3D point cloud reconstructions to a reference 3D point cloud obtained from LiDAR. However, it is important to note that this metric cannot evaluate the results for difficult-to-track

TABLE VI: Average reconstruction error e_{ref} and standard deviation σ obtained by the comparison between the SLAM algorithms' reconstructions with the reference point cloud.

$e_{ref}(m)$ ($\sigma(m)$)	DSO	VI-DSO	SVO	VI-SVO
Easy-to-track	**0.15 (0.09)**	0.18 (0.10)	0.16 (0.20)	0.31 (0.10)
Medium-to-track	0.19 (0.16)	**0.15 (0.10)**	0.16 (0.21)	0.44 (0.28)

sequence, as the LiDAR is unsuitable for operation in this particular environment.

The error assessment involves aligning and scaling the estimated point cloud to match the reference point cloud, enabling the computation of the Euclidean distance between ten randomly selected points in the estimated point cloud and their corresponding points in the reference point cloud. These operations were carried out using the CloudCompare tool [13], which implements point cloud alignment based on the Iterative Closest Point (ICP) algorithm. A lower error value indicates a higher level of accuracy in the reconstruction. The error values, denoted as e_{ref}, are presented in Table VI, comparing the reconstructions of SLAM algorithms and the reference point cloud.

As observed in Table VI, DSO yielded the lowest reconstruction error in the easy-to-track sequence, followed by SVO, VI-DSO, and VI-SVO. VI-DSO achieved the lowest error for the second sequence, with SVO, DSO, and VI-SVO following. e_{rec} analysis revealed a consistent 3D reconstruction quality for the five algorithms, yet the VI-SVO presented a relatively higher reconstruction error, possibly due to its mapping approach. VI-SVO initiates point estimation with a large depth uncertainty and refines depth values in subsequent frames. In contrast, DSO and VI-DSO employ a more sophisticated method of identifying and tracking candidate points over several frames, generating an initial coarse depth value for joint optimization. Additionally, they incorporate an adaptive threshold to identify and mitigate outliers by removing candidate points with photometric errors exceeding a set limit. The mapping strategy adopted by DSO and VI-DSO contributes to a reliable 3D reconstruction of the environment, resulting in lower e_{rec} values and a more detailed reconstruction compared to SVO and VI-SVO, as evidenced in Figure 7.

Based on the e_{rec} and e_{ref} results, it is possible to verify the equivalent performance obtained by the algorithms regarding the mapping. While VI-SVO presented a superior performance regarding the general consistency of the 3D reconstruction by presenting the lowest e_{rec} values across most of the sequences, the DSO and VI-DSO also demonstrated to generate accurate 3D reconstructions as demonstrated by their low e_{ref} values across the easy- and medium-to track sequences.

VI. Conclusion

In conclusion, this paper analyzed and compared various SLAM algorithms within the context of real-case scenarios and constraints associated with radiological mapping for the first time in the literature. In order to do so, a first theoretical analysis was conducted, enabling the selection of five algorithms based on their responses to the nuclear facilities' characteristics. Then, we constructed a new dataset containing different data types and features, allowing the benchmark of the selected algorithms. The algorithms' benchmarking employed criteria encompassing alignment errors, loop drift assessment, and 3D reconstruction evaluation.

In the context of tracking errors, the results highlight VI-SVO as a promising SLAM algorithm for trajectory estimation in nuclear facilities undergoing dismantling. VI-SVO consistently demonstrated superior performance, obtaining an average error ranging from 24% to 53% lower than the other algorithms in terms of alignment error values for the easy- and medium-to-track sequences, and it was the only tested algorithm able to fully operate in the difficult-to-track sequence, as presented in Table III. Moreover, regarding the loop-drift evaluation, VI-SVO surpassed the other algorithms by 1% to 76%, as presented in Table IV. In light of these results, VI-SVO can be considered the most suitable algorithm for tracking nuclear dismantling facilities, among the selected algorithms.

Concerning the 3D reconstruction assessment based on the e_{rec} analysis, aside from the SVO algorithm, all implemented algorithms achieved comparable results, generating a consistent 3D reconstruction, as depicted in Table V. VI-SVO marginally outperformed the others, yielding reconstructions that were, on average, 3% to 5% more consistent than those produced by the DSO and VI-DSO algorithms.

On the other hand, in an overall evaluation of the reconstruction quality based on the e_{ref} analysis, DSO and VI-DSO outperformed the SVO and VI-SVO algorithms. While exhibiting a low average error value, SVO demonstrated a relatively high standard deviation, indicating inconsistency in the reconstruction. In contrast, VI-SVO registered the highest error values, as shown in Table VI. Despite this, only VI-SVO could complete the tracking and mapping of the difficult-to-track sequence successfully. VI-DSO managed to track a portion of the last sequence's trajectory and obtained a reconstruction significantly affected by outliers. The monocular algorithms could not process the final sequence, rendering them unsuitable for localization and mapping in nuclear facilities undergoing dismantling. Considering the presented analysis, it is possible to conclude the superior performance and suitability of the VI-SVO algorithm to reconstruct and localize nuclear facilities under dismantling. Further analysis must be conducted to analyze the embeddability of this algorithm for a modular handheld implementation.

Acknowledgment

The authors would like to acknowledge the valuable contributions of Romuald Woo, from CEA List, for his exceptional support and expertise throughout the prototype's design and the colleagues from the DIASI department.

References

[1] Hellfeld, Daniel and Barton, Paul and Gunter, Donald and Haefner, Andrew and Mihailescu, Lucian and Vetter, Kai, *Real-Time Free-Moving Active Coded Mask 3D Gamma-Ray Imaging*, *IEEE Transactions on Nuclear Science*, vol. 66, no. 10, pp. 2252-2260, 2019, 10.1109/TNS.2019.2939948.

[2] Andrew Haefner and Ross Barnowski and Paul Luke and Mark Amman and Kai Vetter, *Handheld real-time volumetric 3-D gamma-ray imaging*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 857, pp. 42-49, 2017, 10.1016/j.nima.2016.11.046.

[3] Macario Barros, Andréa and Michel, Maugan and Moline, Yoann and Corre, Gwenolé and Carrel, Frédérick, *A Comprehensive Survey of Visual SLAM Algorithms*, *Robotics*, vol. 11, no. 24, 2022, 10.3390/robotics11010024.

[4] C. G. Harris and M. Stephens, *A Combined Corner and Edge Detector*, *Alvey Vision Conference*, 1988.

[5] Leutenegger, Stefan and Lynen, Simon and Bosse, Michael and Siegwart, Roland and Furgale, Paul, *Keyframe-Based Visual-Inertial Odometry Using Nonlinear Optimization*, *The International Journal of Robotics Research*, 2014, 10.1177/0278364914554813.

[6] Mur-Artal, Raúl and Tardós, Juan D., *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, 10.1109/TRO.2017.2705103

[7] R. A. Newcombe et al., "KinectFusion: Real-time dense surface mapping and tracking," 2011 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, 2011, pp. 127-136, doi: 10.1109/ISMAR.2011.6092378.

[8] Engel, J., Schöps, T., Cremers, D. (2014). LSD-SLAM: Large-Scale Direct Monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8690. Springer, Cham. https://doi.org/10.1007/978-3-319-10605-2_54

[9] R. A. Newcombe, S. J. Lovegrove and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," 2011 International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 2320-2327, doi: 10.1109/ICCV.2011.6126513.

[10] C. Forster, M. Pizzoli and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 2014, pp. 15-22, doi: 10.1109/ICRA.2014.6906584.

[11] Vit A, Shani G. Comparing RGB-D Sensors for Close Range Outdoor Agricultural Phenotyping. Sensors (Basel). 2018 Dec 13;18(12):4413. doi: 10.3390/s18124413.

[12] Engel, Jakob and Usenko, Vladyslav and Cremers, Daniel. A Photometrically Calibrated Benchmark For Monocular Visual Odometry. 2016.

[13] CloudCompare (version 2.11.1) [GPL software]. Retrieved from: http://www.cloudcompare.org/

[14] *Blackfly S USB3*. [Online]. Available: https://www.flir.fr/products/blackfly-s-usb3/?vertical=machine+vision&segment=iis

[15] *MTi 600-series Datasheet*. [Online]. Available: https://www.xsens.com/hubfs/Downloads/Leaflets/MTi%20600-series%20Datasheet.pdf

[16] *VLP-16 User Manual*. [Online]. Available: https://velodynelidar.com/wp-content/uploads/2019/12/63-9243-Rev-E-VLP-16-User-Manual.pdf

[17] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 2007, pp. 225-234, doi: 10.1109/ISMAR.2007.4538852.

[18] Tateno, Keisuke and Tombari, Federico and Laina, Iro and Navab, Nassir. (2017). CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. 10.1109/CVPR.2017.695.

[19] J. Engel, V. Koltun, D. Cremers, Direct Sparse Odometry, In arXiv:1607.02565, 2016

[20] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," Proceedings 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, 2007, pp. 3565-3572, doi: 10.1109/ROBOT.2007.364024.

[21] M. Bloesch, S. Omari, M. Hutter and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015, pp. 298-304, doi: 10.1109/IROS.2015.7353389.

[22] T. Qin, P. Li and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," in IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004-1020, Aug. 2018, doi: 10.1109/TRO.2018.2853729.

[23] Stumberg, Lukas and Usenko, Vladyslav and Cremers, Daniel. (2018). Direct Sparse Visual-Inertial Odometry Using Dynamic Marginalization. 2510-2517. 10.1109/ICRA.2018.8462905.