Karlsruhe Institute of Technology
Institute for Material Handling and Logistics
Prof. Dr.-Ing. K. Furmans

Bachelor Thesis

Feature-Based Visual Odometry and RGB-D Mapping in Confined Pipe Environments

Submitted by:
Yusuf Berk Açıkel

Karlsruhe, August 2025

Supervised by:
MSc. Azman Azka

Bachelor Thesis AA-82804-10095 at the
Institute for Material Handling and Logistics
Karlsruhe Institute of Technology
Subject: Feature-Based Visual Odometry and RGB-D Mapping in Confined Pipe Environments
Author: Yusuf Berk Açıkel

Yusuf Berk Açıkel
Klosterweg 28
76131 Karlsruhe
email: yusufberkacikel@gmail.com

I hereby truthfully affirm that I have done my work independently, that I have not used any other sources and aids than those indicated, that I have identified the places taken over verbatim or in terms of content as such and that I have observed the statutes of the KIT for the safeguarding of good scientific practice in the respectively valid version.

Karlsruhe, 19.08.2025

.....

Place, Date (Yusuf Berk Açıkel)

To Mom and Dad

Abstract

Maintenance of EU-DEMO's breeding blanket and divertor cassettes must be fully remote due to extreme radiation and spatial constraints inside the vacuum vessel. A key bottleneck lies in the reliable rejoining and inspection of in-bore pipe welds, where current Non-Destructive Testing (NDT) and in-bore localization methods remain immature. Previous approaches using wheel encoder-based mapping suffered from odometry drift due to slippage and pipe network complexity.

This study evaluates the Intel® RealSense™ Depth Camera D405 for visual odometry and 3D surface mapping in confined pipe environments. A feature-based visual odometry pipeline is implemented to estimate inter-frame camera motion, and systematic depth errors are heuristically corrected to support accurate reconstruction of straight pipe segments. Results demonstrate that the camera is capable of producing consistent dense point clouds even without dedicated pose estimation in straight pipe sections by leveraging the constrained one-dimensional linear motion of the pulley-actuator along the pipe axis. While feature-based pose estimation improves short-term alignment, it accumulates drift over time and fails to maintain global consistency in long-term mapping. These findings highlight the potential and current limitations of visual odometry in fusion-relevant confined environments, informing future maintenance strategies for EU-DEMO.

Contents

Abbreviations

DEMO	Demonstration Fusion Power Plant
NDT	Non-Destructive Testing
D405	Intel RealSense Depth Camera D405
ITER	International Thermonuclear Experimental Reactor
VV	Vacuum Vessel
IVC	In-Vessel Component
PFC	Plasma Facing Component
RH	Remote Handling
VMS	Vertical Maintenance Scheme
MMS	Multi Module Segment
TF	Toroidal Field
3D	Three-Dimensional
RGB-D	Red, Green, Blue + Depth
IR	Infrared
ToF	Time-of-Flight
SLAM	Simultaneous Localization and Mapping
SfM	Structure-from-Motion
DN90	Diameter Nominal 90
AISI	American Iron and Steel Institute

IMU	Inertial Measurement Unit
FPS	Frames Per Second

1 Introduction

DEMOnstration Power Plant

The world's largest nuclear fusion experiment to date, ITER, aims to demonstrate the feasibility of fusion energy as a large-scale, carbon-free power source. Demonstrating that net electricity can be produced from fusion represents the next essential breakthrough (International Atomic Energy Agency 2023). Demonstration fusion power plants, often known as DEMOs, will be developed for this purpose. Key mission objectives for DEMO-class power plants are to secure regulatory licensing, ensure stable plasma operation, and maintain continuous full-power performance over several years. Additionally, DEMOs must achieve tritium self-sufficiency, generate net electricity at the scale of several hundred megawatts, and validate all necessary technologies for building a commercial fusion power plant with an adequate level of availability (Federici et al. 2014). The European fast-track strategy envisions DEMO as the single step between ITER and a commercial fusion power plant, thereby positioning it not as an experimental device, but as a near-final point design. In this context, a tokamak configuration has been selected for DEMO (Batistoni et al. 2010).

DEMO Tokamak Architecture

The DEMO tokamak shares its core design features with other superconducting tokamaks, notably ITER. Both the magnet system and the vacuum vessel (VV) are structurally independent components, each supported by a ring structure located within the cryostat. The VV acts as the structural backbone for all in-vessel components (IVCs) and equipment installed in the VV ports, such as port plugs and associated piping. Covering the plasma-facing surface is a protective wall composed of blankets, divertors, and port plug plasma-facing components (PFCs), all engineered to endure intense heat loads (Bachmann et al. 2018).

Remote Maintenance

One of the central challenges for the EU-DEMO reactor is ensuring high availability, a

prerequisite for commercial viability. Early estimates of the fusion power plant suggested a minimum availability of 75% (Crofts and Harman 2014). In this context, maintenance is regarded as a critical consideration due to the high cost associated with downtime and the risk of certain failures rendering the reactor inoperable. The densely packed environment and harsh conditions within fusion reactors further complicate this, requiring that all inspection, maintenance, and replacement activities be performed using remote handling (RH) systems.

The IVCs are subjected to high stress levels due to thermal, electromagnetic, and mechanical loads. As a result, regular replacement of IVCs is necessary (Bachmann et al. 2018). Several maintenance strategies have been proposed to tackle this challenge. The Vertical Maintenance Scheme (VMS) with Multi Module Segment (MMS) has been chosen as the preferred layout. It consists of 16 TF coils. Each toroidal segment includes an upper vertical port, an equatorial port, and a lower divertor port.

Service Joining

A major maintenance challenge in DEMO is the reliable management of service connections for the first wall components. These connections enable the transport of essential breeding and cooling fluids and are critical for maintaining the vacuum integrity of the tokamak, where even minor leaks are unacceptable. Welding has been identified as the most suitable method to ensure the required leak-tightness in the harsh in-vessel environment. The process is performed using tools introduced through the upper port and deployed from a cask, with cutting of the old pipes followed by welding of the new ones to the permanent connections. Service joining tasks account for roughly forty percent of the total upper port maintenance duration, and the timely validation of weld integrity is essential for meeting availability targets. For this purpose, non-destructive testing methods (NDT) are applied immediately after welding to ensure quality without introducing significant delays (Tremethick et al. 2020).

Non-Destructive Testing

As part of the safety inspections the surface condition of the pipes is evaluated to ensure compliance with regulatory criteria. This includes verifying weld integrity and precisely mapping the location and size of surface defects of defined types, with the requirement of detecting a sub-0.5 mm discontinuity (Azka and Mittwollen 2024). To meet these stringent requirements, NDT methods are employed in-bore to validate pipe joints. Visual inspection using optical fiberscopes and volumetric inspection using ultrasonic techniques are among

the proposed methods for assessing weld quality against acceptance criteria (Crofts et al. 2016).

According to the technical risk assessment of the remote handling system's concept design, the most significant risks are shared by all currently proposed in-vessel component maintenance solutions. These risks are associated with the precise handling and control of in-vessel components. The service joining system concept also presents key technical challenges, such as the immaturity of NDT techniques and the currently uneconomical duration of joining operations (Tremethick et al. 2020).

A prior study examined the use of sensors, such as cameras and lasers, for surface inspection. The goal was to generate detailed surface maps which enable analysis, whether conducted manually or automatically, to detect surface defects or critical features that might endanger the power plant's safety and operational reliability. However, it was discovered that when wheel encoders were utilized for localization, the slip between the deployment module's wheel and the wall surface produced an inconsistent measurement result. This limitation highlighted the need for more reliable localization techniques. As a result, there is growing motivation to explore alternative methods, such as visual odometry, which leverages image data to estimate the motion of the inspection module.

Study Objectives and Scope

The primary goal of this study is to assess the feasibility of using RGB-D cameras for visual odometry and mapping in a nuclear fusion environment. Specifically, this work aims to:

- Evaluate the feasibility of using RGB-D cameras for in-bore inspection and localization,
- Identify challenges and peculiarities associated with the inspection of confined pipe geometries in DEMO-like environments,
- Explore methods to improve the accuracy and reliability of RGB-D based localization and mapping.

This study focuses on the experimental evaluation of a single RGB-D sensor (Intel RealSense D405) mounted on a robotic inspection module. Tests are conducted on a custom-built test-bench mimicking the pipe geometry of the EU-DEMO environment. The study is organized as follows:

- **Introduction:** Provides background on DEMO, remote maintenance, and the challenges associated with visual inspection in confined environments.

- **Literature Review:** Surveys prior work on 3D imaging, stereo vision, camera tracking, geometric representations, and autonomous inspection. The mathematical model for RGB-D visual odometry and mapping utilized in this work is also presented..
- **Methodology:** Details the custom test-bench setup, robotic platform, data acquisition procedures, and identified limitations of the experimental approach.
- **Results:** Reports qualitative and quantitative analyses of RGB-D depth and color data, systematic depth error correction, feature-based pose estimation, and reconstructed point clouds.
- **Discussion and Conclusion:** Interprets the findings emphasizing the feasibility, challenges, and potential improvements for RGB-D-based visual odometry and mapping in confined nuclear fusion pipe environments.

2 Literature Review

Three-Dimensional Imaging

Three-dimensional (3D) imaging refers to methods used to capture and reconstruct the spatial structure of objects or scenes, serving as a foundational tool in computer vision and graphics for applications such as scene reconstruction and object recognition, navigation, and manipulation (Pears et al. 2012). Its central challenge lies in recovering the depth information that is lost when a three-dimensional scene is projected onto a two-dimensional image plane. The theoretical foundations for addressing this problem were established through research in multi-view geometry and projective methods, which formalized how correspondences across images can be used to infer scene structure (Longuet-Higgins 1981). Building on these principles, practical systems such as stereo vision and RGB-D cameras have been developed to directly estimate depth, providing reliable and dense 3D data in real time. These systems often capture aligned color images using conventional digital cameras, delivering both the geometric structure and photometric information of the scene.

Numerous sensing modalities have been developed to capture 3D data, each employing different physical principles and hardware to reconstruct scene geometry. Passive methods rely solely on natural illumination and multiple views, extracting depth from 2D images through feature correspondences or photometric consistency. Traditional methods such as binocular stereo and multi-view photogrammetry can recover the 3D structure of a scene using only standard cameras, without requiring extra specialized hardware. In contrast, active approaches rely on dedicated hardware to obtain depth measurements. Time-of-Flight (ToF) cameras project modulated infrared light and estimate distance either by measuring the travel time of light pulses or by analyzing the phase shift of a continuous-wave signal, thus providing dense depth data at video frame rates (Horaud et al. 2016). Active infrared (IR) stereo systems project structured IR patterns onto the scene to create artificial texture, which a stereo-matching algorithm then uses to compute disparity and depth (Jang et al. 2013), improving reconstruction accuracy in areas with little texture. Each modality involves

inherent trade-offs in range, accuracy, resolution, cost, and robustness to environmental conditions, which determine their suitability for specific applications in robotics, augmented reality, and industrial inspection.

Passive Stereo Vision

Stereo depth estimation is a well-established 3D imaging technique that calculates distances using the principle of triangulation. Two cameras, with a known baseline, capture the same scene from slightly different angles, and the depth of objects is inferred by measuring the pixel disparities between the corresponding points in the two images. This method emulates human binocular vision and has been the focus of extensive research in both computer vision and photogrammetry. Traditional techniques, including block matching (Kanade and Okutomi 1994) and semi-global matching (Hirschmüller 2005), are still commonly employed for computing disparities, while more recent approaches (Sarlin et al. 2019) leverage machine learning to enhance correspondence matching in difficult scenarios. In the context of robotic inspection, passive stereo depth estimation is often preferred in scenarios where active sensors may be unreliable or where space constraints limit the use of additional hardware. A representative example is the Intel RealSense D405 (Intel Corporation 2025), a stereo camera tailored for detailed close-range imaging. Its ability to capture fine details at extremely close distances makes it suitable for inspection tasks in confined environments.

Camera Tracking

Camera tracking involves estimating the position and orientation of a camera over time to accurately reconstruct or navigate a scene. Approaches typically rely on establishing correspondences between frames, which can be either sparse feature-based matches or dense, photogrammetric correspondences that exploit intensity or color information across all pixels for precise alignment (Zollhöfer et al. 2018). Camera tracking can be carried out using different strategies: frame-to-frame, which aligns each new frame with the immediately preceding one; frame-to-model, which registers the current frame against an evolving 3D representation of the scene; and global optimization, which adjusts all poses simultaneously to reduce cumulative drift (Zollhöfer et al. 2018). To maintain continuous operation under challenging circumstances, robust tracking must include re-localization strategies that recover the camera's position when tracking is disrupted by occlusions, fast movements, or low-texture areas.

Several related techniques are widely used in camera tracking and scene reconstruction. Optical flow estimates the apparent motion of pixels between consecutive frames, providing

dense 2D motion fields that can assist in ego-motion estimation or scene flow computation, but does not directly produce 3D structure. Structure-from-motion (SfM) (Tomasi and Kanade 1992) recovers both camera motion and 3D scene geometry from a collection of images, typically in an offline batch process, relying on feature correspondences and triangulation. In contrast, Simultaneous Localization and Mapping (SLAM) works by progressively estimating both the camera path and scene geometry, refining the map online with techniques like loop closure and global optimization to reduce drift (Smith et al. 2013).

Geometric Representation

Geometric representation is a fundamental aspect of 3D reconstruction, defining how spatial information is stored, processed, and visualized. The most basic and commonly used 3D representation is the point cloud, which consists of individual 3D points capturing sampled locations on surfaces, typically acquired using depth sensors or stereo camera systems. Mesh-based representations connect points into vertices, edges, and faces to form continuous surfaces, providing richer topological information that is useful for visualization, simulation, and collision detection. Voxel-based representations discretize 3D space into a uniform grid of volumetric elements, allowing efficient fusion of multiple scans and volumetric reasoning. Recent advances have introduced hybrid and learned representations, such as implicit neural fields (Mildenhall et al. 2020) or occupancy networks (Mescheder et al. 2018), which can encode high-fidelity geometry compactly, support smooth surface reconstruction, and generalize across incomplete or noisy observations.

Autonomous Inspection

High-quality 3D reconstruction of complex pipe networks is particularly time-consuming, as capturing a complete representation requires acquiring geometry from multiple viewpoints along densely packed and spatially constrained structures. The path followed by the inspection system directly influences both scanning efficiency and reconstruction quality. Auto-scanning approaches address this challenge by formulating it as an optimization problem, generating control signals for robotic systems to enable automated acquisition of detailed 3D data (Zollhöfer et al. 2018). Assuming a centrally positioned camera, motion during pipe inspection can be described in two components: linear translation along the pipe's axis to scan its entire length, and rotation around the axis (if allowed) at each location to acquire multiple viewpoints, enhancing both surface coverage and the fidelity of the reconstructed 3D model. Inspection efficiency is further enhanced through parallelized operation, where multiple cameras or modules simultaneously scan different pipe sections, accelerating data acquisition. These strategies are crucial for ensuring timely and

comprehensive monitoring of critical components while coping with the spatial constraints inherent to such environments.

Mathematical Problem Description

Sensor and Data Model

A mobile robot is moving within a confined pipe environment. It is equipped with a rigidly attached stereoscopic RGB-D camera system, which, at each discrete time step $t \in \{1, \ldots, T\}$, captures grayscale intensity images from its left and right passive infrared (IR) cameras. Using the stereo intensity pair, the camera's onboard processors compute a corresponding depth image $D_t \in (R_{>0})^{H \times W}$ via disparity calculation, with the left imager serving as the reference frame. For simplicity, this process is considered a black-box, and specific details of depth computation are not discussed here. Additionally, the left imager has an integrated color sensor that provides RGB data, namely $I_t \in R^{H \times W \times 3}$.

Utilizing the camera intrinsics matrix $\mathbf{K} \in R^{3\times3}$ (obtained through calibration) and the depth image D_t, each pixel $\mathbf{u}_{t,i} = (u_{t,i}, v_{t,i})^\top$ with a valid depth reading is reconstructed to its corresponding 3D point $\mathbf{X}_{t,i} \in R^3$ via:

$$\mathbf{X}_{t,i} = D_t(\mathbf{u}_{t,i}) \cdot \mathbf{K}^{-1} \cdot \tilde{\mathbf{u}}_{t,i} \tag{2.1}$$

where $\tilde{\mathbf{u}}_{t,i}$ denotes the homogeneous pixel coordinates pears20123d. This yields the raw point cloud $\mathbf{P}_t = \{\mathbf{X}_{t,i}\}_{i=1}^{N_t}$, with each point colored using its corresponding RGB value from I_t .

Pose Estimation (Visual Odometry)

Features are first extracted from the grayscale-converted color image at time t, then matched to those in frame $t+1$ after applying a mask to exclude pixels with zero depth. The pose change between consecutive frames is estimated by solving the 3D–2D Perspective-n-Point (PnP) problem using Levenberg-Marquardt optimization. Given n correspondences between 3D points $\mathbf{X}_{t,i}$ in frame t and their 2D projections $\mathbf{u}_{t+1,i}$ in frame $t+1$, the objective is to find the rigid transformation $\mathbf{T}_{t\to t+1}$ that minimizes the re-projection error:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{n} \left\| \mathbf{u}_{t+1,i} - \pi\left(\mathbf{R}\mathbf{X}_{t,i} + \mathbf{t}\right) \right\|^2,$$

where $\pi(\cdot)$ denotes the perspective projection function opencv$_p$np.

Depth Correction for Mapping

Given point cloud data and pose information for a single frame, the scene can be reconstructed. However, due to sensor noise and systematic errors in depth estimation, the raw point clouds $\mathbf{P}_t$, representing the observations $\mathbf{y}_{t,i} \in R^3$, contain spatial distortions. To accurately reflect the cylindrical geometry of the straight pipe section, these distortions must be corrected.

Trajectory and Global Map Construction

The robot's trajectory is obtained by accumulating relative transformations $\hat{\mathbf{T}}_{t,t+1}$ from the initial frame to frame t, forming

$$\hat{\mathbf{T}}_{0,t} = \prod_{k=1}^{t} \hat{\mathbf{T}}_{k-1,k}.$$

Each corrected 3D point $\tilde{\mathbf{X}}_{t,i}$ is transformed into the global frame via $\hat{\mathbf{T}}_{0,t}$, producing the global corrected point cloud

$$\tilde{\mathbf{P}}_t^{\text{global}} = \{\hat{\mathbf{T}}_{0,t} \cdot \tilde{\mathbf{X}}_{t,i}\}.$$

Combining these over all frames yields the final 3D map of the environment,

$$\mathcal{M} = \bigcup_{t=1}^{T} \tilde{\mathbf{P}}_t^{\text{global}},$$

representing a refined reconstruction based on the robot's estimated motion and corrected sensor data.

Approach Summary

The outlined approach performs offline visual odometry using raw depth data from a stereo RGB-D system with monocular pose estimation. Depth error correction is applied separately before mapping, and no bundle adjustment or loop closure (i.e., no SLAM) is performed.

3 Methodology

To investigate the applicability of using RGB-D cameras for visual odometry and mapping tasks within a nuclear fusion pipeline, a custom test bench with a straight pipe section was constructed, focusing solely on geometric and lighting considerations to assess the fundamental feasibility of the chosen sensor, without taking into account the effects of radiation, which would further degrade performance and shorten life-cycle.

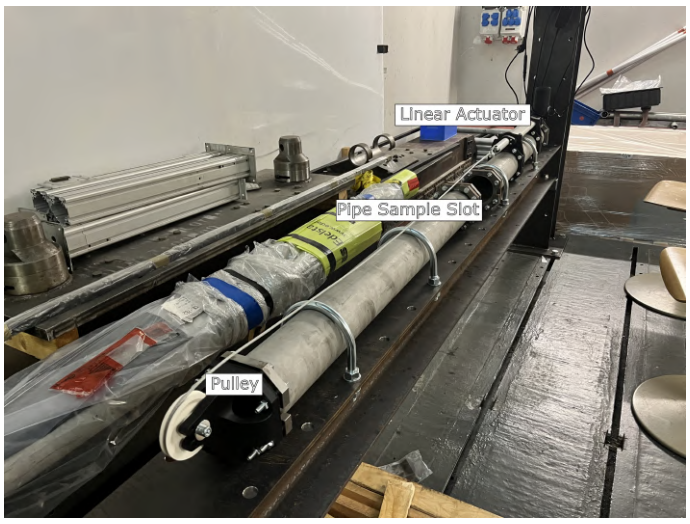

Figure 3.1: The constructed test bench setup.

Pipe Environment

In this study, a straight pipe section was used to establish baseline sensor performance, with additional work planned for bent sections in the future. Pipes used are DN90 pipes with AISI 316L stainless steel as the base material. A pipe sample with a length of 100 mm is placed between the pipe stretches as a test target for the inspection systems. The main pipe stretches are made of rolled steel, while the pipe samples are machined. Although the exact condition of the pipes after operation is unclear, the internal surfaces are presumed to be clean for inspection and welding purposes (Tremethick et al. 2020). The lighting

conditions inside the pipe are predicted to be poor and variable, with reflections further complicating sensor performance. Radiation effects were not considered in this evaluation, as the current focus is on assessing sensor behavior in a confined geometric environment without the additional radiation effects. The test bench pipe structure was designed to allow the robotic platform to be deployed in-line, providing a realistic setting to evaluate its operation within the DEMO pipe environment.

Robotic Platform

A six-legged crawler robotic platform was devised for in-line deployment, with a modular and compact design tailored to DN90 pipes, allowing effective navigation of both horizontal and vertical sections, as well as bend traversal. The platform's modular design supports the in-line deployment of various sensor setups, including laser scanners and cameras. Additionally, although not utilized in this study, the platform is equipped with an inertial measurement unit (IMU) for investigating inertial odometry, alongside a slot for a Raspberry Pi 4B, which is powered by an onboard battery and serves as the data acquisition unit for the IMU. A wheel encoder is also integrated on the platform. The camera system is powered externally via the host computer and is equipped with integrated static LED lighting, which is powered through a separate external cabling. Locomotion is achieved externally by a linear actuator pulley system that pulls the platform through the pipe at a constant speed of 12 mm/s, while the leg structure ensures the camera remains centered within straight pipe sections.

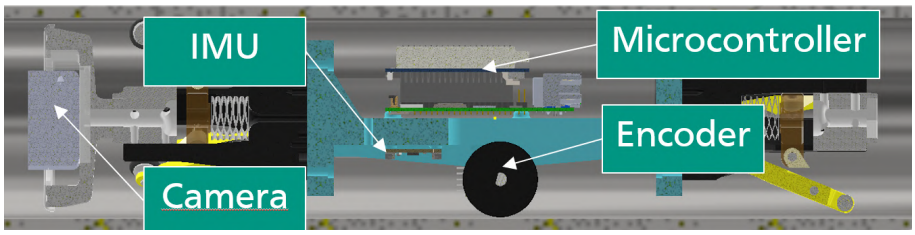

Figure 3.2: The in-line deployed six-legged crawler robotic platform equipped with the Intel RealSense D405 RGB-D camera.

Data Acquisition

The camera and actuator data were collected during the in-line deployment runs. The camera recordings were made using the Intel RealSense Viewer, with the camera connected to the computer via a USB 3.0 cable to ensure reliable power and data transfer. Recordings were made using the high-density preset at 30 FPS with a resolution of 848×480, without

any additional changes to the camera settings. Prior to data collection, the camera was calibrated using the self-calibration routine provided by the RealSense Viewer. During each run, .bag files were saved, from which color and depth images were later extracted for further processing and analysis. In parallel with the camera data, actuator encoder data was recorded.

Identified Limitations

This method, while effective within its defined scope, exhibits several important limitations that constrain its broader applicability and performance.

- Relies on a black-box depth estimation process with limited control over algorithmic assumptions or parameters, limiting adaptability to the confined pipe environment.
- Utilizes only the left imager in a monocular manner, without leveraging both IR imagers for custom stereo-based depth estimation.
- Does not incorporate bundle adjustment or loop closure techniques, which limits the refinement and global consistency of the reconstructed map.
- Operates entirely offline, preventing real-time deployment for inspection or mapping tasks.
- The implementation is computationally expensive compared to typical SLAM pipelines, restricting practical use on embedded systems.
- Depth estimation degrades under sub-optimal or uncontrolled lighting conditions due to a lack of active illumination control.
- The feature-poor nature of pipe environments complicates robust feature extraction and matching, affecting pose estimation accuracy.
- Systematic depth errors are only partially corrected using a simple mirroring approach, leaving residual geometric inaccuracies.
- Random errors were not addressed rigorously; complexity was reduced by filtering points. Rather than processing single frames, random errors need to be handled using a sliding-window approach.
- Provides limited digital twin capabilities, as there is no validated correspondence between the reconstructed map and the real environment.
- Accounts solely for geometric and lighting-related effects, neglecting any radiation-based influences.

4 Results

Qualitative Results

This section provides a detailed qualitative assessment of the raw and corrected depth and color data captured by the D405 RGB-D camera inside a straight pipe segment. Through visual inspection of depth maps, color images, and reconstructed point clouds, the presence and characteristics of systematic depth errors are explored. Additionally, the devised heuristic depth correction method is evaluated by comparing raw and corrected data representations, illustrating how geometric fidelity and data consistency are enhanced, thereby improving mapping accuracy.

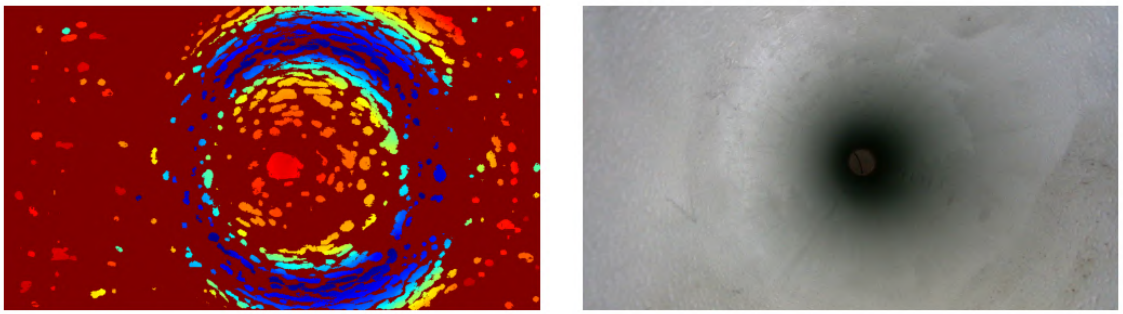

Figure 4.1: Exemplary raw depth and color images obtained using the RealSense D405 RGB-D camera for a single frame inside the pipe environment.

Analysis of Depth and Color Image

Figure 4.1 displays an exemplary depth and color picture obtained by the left imager of the D405 camera inside a straight pipe segment. With an 18mm baseline between the imagers, the scene is effectively observed from left to right. This results in a higher density of depth measurements on the right side of the pipe, creating a slight asymmetry in the otherwise concentric pattern reflecting the cylindrical geometry of the pipe environment. It is observed that, within the densest central section, the top and bottom regions of the

pipe contain a higher number of valid depth pixels compared to the left and right regions. Furthermore, systematic depth estimation errors manifest as a depth coloring gradient that gradually increases towards the center of the pipe before decreasing again, deviating from the expected monotonic increase from periphery to center imposed by the pipe's actual geometry.

Influence of Lighting Conditions on Depth Measurements

During the experiments, lighting conditions varied unpredictably as the camera moved through the pipe. Depth estimation with the D405, which relies on passive stereographic infrared triangulation, is highly sensitive to such variations in illumination. This sensitivity to illumination directly impacts the infrared images captured by the left and right imagers, which serve as the basis for depth estimation through block matching. When pixels become too dark—caused by inadequate lighting or the pipe material's inherent reflectivity—the camera cannot produce valid depth measurements. This limitation is evident in the checkerboard image, where areas with low reflectivity lack sufficient reflected IR light, preventing reliable block matching between the stereo imagers and leading to gaps in the data. Furthermore, the presence of strong reflections creates new challenges that compromise the accuracy and consistency of depth estimation. Glare frequently results in the underestimation of depth values, which causes points to appear closer than their actual locations. As a result, objects affected by this phenomenon may appear larger or distorted during reconstruction. All of these depth estimation errors propagate through the final point cloud as geometric inaccuracies, reducing the overall quality of the reconstructed 3D model.

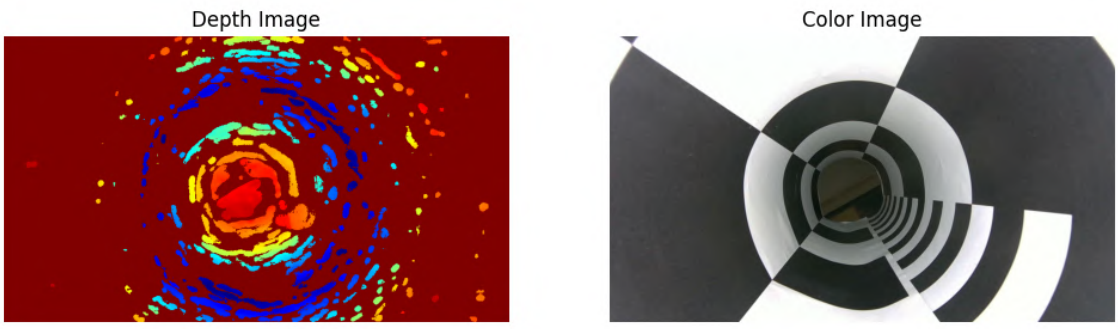

Figure 4.2: Checkerboard pattern with known dimensions placed on the inner surface of the pipe. The depth image shows that regions corresponding to darker areas contain few or even no valid depth pixels.

Systematic Depth Measurement Errors

No measurement, no matter how carefully performed, can be genuinely free of errors. All measurements are subject to measurement deviation, which is the discrepancy between the observed and actual values of the measured variable. This deviation can be caused by the imperfections in the measuring device, the observer, and the environmental factors, and is often analyzed under two categories: random errors, which primarily impact the consistency of repeated measurements under the same circumstances, and systematic errors, which affect the overall accuracy of the measurement in predictable ways.

Estimating the pose change between consecutive frames through feature matching, and aggregating these to reconstruct the camera trajectory while simultaneously mapping the environment, is a complex task that presents unique challenges, particularly in confined, small-diameter pipe environments. It was observed throughout the experiments that the Intel RealSense D405 RGB-D camera produces systematic depth measurement errors due to both intrinsic limitations of the measuring device and environmental factors within the test bench. The depth estimation of the D405 is based on stereographic passive infrared triangulation, with the left imager serving as the primary imaging sensor and the right imager assisting in depth calculation from disparity. Although it is primarily intended for short-range general-purpose indoor/outdoor applications with an optimal operating range of 7 cm to 50 cm, its depth estimation method has inherent limitations when used in confined hollow environments such as small-diameter pipes. In such scenarios, triangulation becomes less reliable due to the small baseline between the left and right imagers.

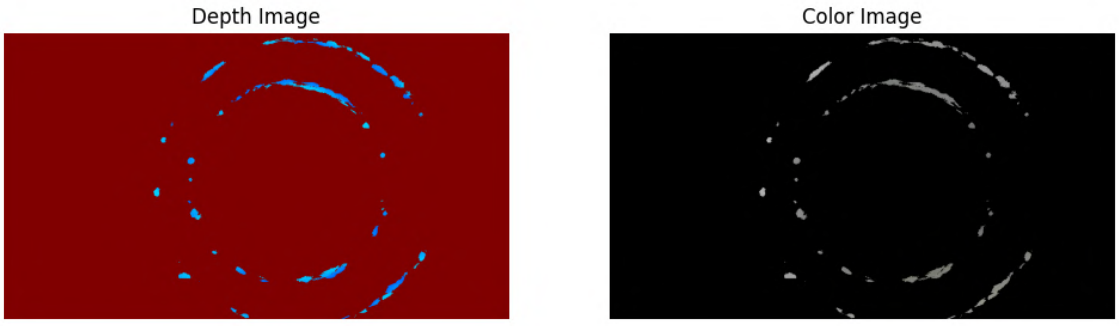

Figure 4.3: Exemplary binned depth and corresponding color images showing pixels within the 150–175 range. Two nearly concentric circular patterns are visible, indicating a systemic depth calculation error, as these regions should have distinct depths along the pipeline rather than appearing at similar values.

Under ideal conditions, the depth visualization is expected to exhibit a coherent, near-continuous gradient, with depth values gradually increasing from the periphery of the image toward the center, reflecting the cylindrical geometry of the pipe. However, during the experiments, the acquired raw depth measurements consistently deviated from this ideal pattern. From a two-dimensional perspective, the binned depth images revealed nearly concentric circular patterns, where pixels corresponding to different z-positions along the pipe axis were assigned similar depth values. When the depth images were reconstructed into a three-dimensional point cloud using the intrinsic matrix, these systematic errors manifested as two nearly concentric conic structures along the pipe's axis. This indicates that the depth estimation method was unable to differentiate between foreground and background regions in the image correctly. Without addressing these systematic depth errors, accurate mapping becomes fundamentally compromised. Therefore, in order to solve the high-level tasks of mapping, the downstream task of depth correction becomes essential.

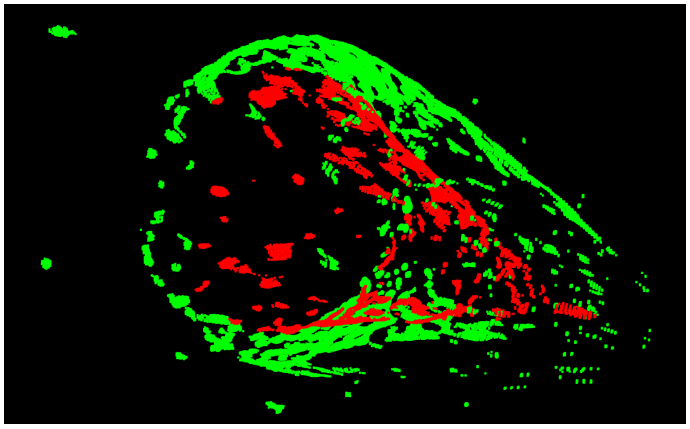

Figure 4.4: Visualization of the raw point cloud data from the exemplary frame reconstructed using the known intrinsics matrix viewed from an angle. Pixels are colored to indicate foreground (green) and background (red). Two conic patterns are visible, highlighting systemic depth errors.

When addressing systematic depth measurement errors, heuristic-based correction methods provide a practical solution, especially in cases where detailed analytical modeling of the error source is infeasible. The proposed depth correction method in this work is designed to be conceptually simple and intuitive, though the current implementation may not be computationally inexpensive, and its generalizability is currently limited to straight pipe

sections under favorable lighting conditions.

To produce a corrected point cloud that accurately represents the pipe's internal geometry, two conceptual approaches were considered: correcting the depth image before point cloud reconstruction or directly manipulating the three-dimensional point cloud coordinates. In this work, the former approach was chosen, as correcting the raw depth image maintains consistency with camera intrinsics and preserves geometric integrity. The first step in the proposed method involves selecting which pixels require correction. This is achieved through a foreground/background classification process. To perform this classification, a circle is first fitted to the set of pixels corresponding to the maximum observed depth value in the raw depth image. Each pixel is then classified by calculating its radial distance from the estimated center and comparing this distance to the fitted radius. Pixels located outside the fitted circle are classified as foreground, whereas those inside are classified as background and subsequently flagged for depth correction. The correction itself is performed by mirroring the depth values of these background pixels with respect to the reference depth (max depth). By effectively reflecting the incorrect depth measurements around this reference plane, the corrected depth image produces a point cloud reconstruction that better reflects the expected cylindrical geometry of the pipe, thereby enabling downstream tasks to operate on a more geometrically faithful representation of the environment.

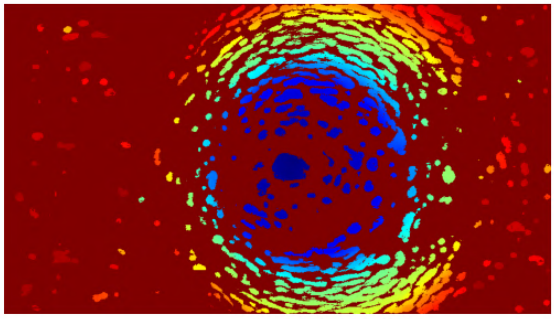

Figure 4.5: Exemplary corrected depth image produced using the proposed heuristic-based mirroring method. The background pixels have been adjusted relative to the reference plane, resulting in a depth map that better reflects the true cylindrical geometry of the pipe environment.

While heuristic methods such as this are inherently approximate, they provide an effective means to correct systematic measurement errors. Nonetheless, care must be taken when applying such corrections, as they manipulate raw measurement data and may introduce

unintended distortions if used without rigorous validation.

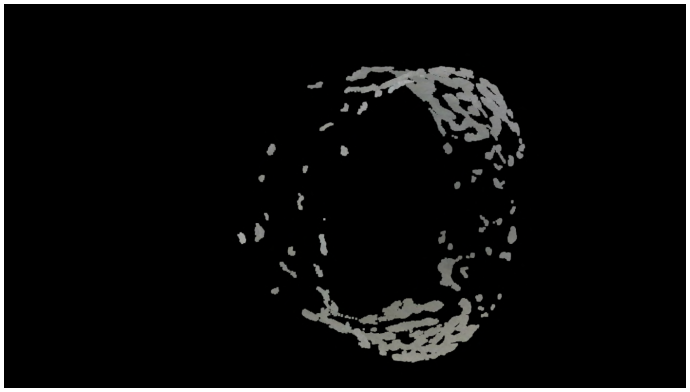

Figure 4.6: Final corrected and filtered local point cloud for the exemplary frame viewed from an angle. The point clouds are reconstructed from color and depth images after applying heuristic-based depth correction, depth-based clipping and consecutive radial filtering, resulting in a geometrically consistent representation of the pipe's internal surface.

Quantitative Results

This section presents a comprehensive quantitative evaluation of the proposed RGB-D visual odometry and mapping algorithm. The analysis is structured to assess multiple aspects of the pipeline, including depth quality, feature detection and matching performance, pose estimation consistency, and overall reconstruction outcomes.

Depth Quality Metrics

RGB-D perception within a constrained pipe environment is not a trivial task for any off-the-shelf camera, especially right out of the box. While space constraints imposed by the pipe dimensions limit the available camera options to choose from, the depth estimation characteristics of the camera play a significant role in the acquisition of data that reflects the real world. As most of the cameras in the market are not designed to inspect hollow objects from the inside but rather to inspect objects on flat planes or feature rich and complex environments, when these cameras are utilized in the DEMO piping environment, the effects of decisions made in the software implementation becomes apparent as systemic depth estimation errors and low fill rate. Moreover, as the regions on the image periphery are too close and do not overlap in the field of view of both left and right imagers, we

get very minimal depth readings from the pixels on the image periphery. It becomes essential to obtain a quantitative measure of the actual data quality to understand the extent of these limitations and their impact on subsequent processing steps. A quantitative analysis of all frames across `sample_run` run was conducted to derive representative metrics characterizing the depth quality throughout the dataset and to evaluate the correction and filtering characteristics of the implemented algorithm.

Table 4.1: Depth Quality Metrics for `sample_run`.

Metric	**Value [%]**
Average Valid Depth Pixel Rate	18.8
Average Corrected Rate	44.5
Average Filtered Rate	50.2
Average Point Cloud Fill Rate	9.4

Overall, the reported depth quality metrics highlight the challenges of reliable data acquisition in the constrained pipe environment and demonstrate the characteristics of the implemented correction and filtering pipelines. Besides the low rate of valid pixel readings with depth > 0, we also see that 44.5% of valid pixels (background) suffer from systemic depth estimation errors and need to be corrected. After correction, a simple depth-based clipping was applied to remove pixels with depth values outside the 225mm to 325mm range, as such pixels consistently degraded the quality of the reconstruction. This clipping step effectively eliminates approximately half of the valid and corrected pixels, representing a trade-off between maintaining high data quality and preserving the density of the final point cloud. Here, data reliability was prioritized to ensure robustness and to eliminate the otherwise necessary radial correction step. Consequently, the average point cloud fill rate, defined as the ratio of remaining pixels after correction and filtering to the total image pixels (848×480), was observed to be 9.4%, reflecting a relatively sparse yet sufficiently informative reconstruction suitable for the targeted mapping task, especially when the sparsity of individual frames is compensated by the high frame rate of the acquisition pipeline.

Feature Detection and Matching

To evaluate the performance of the feature detection and matching components within the proposed pipeline, a quantitative analysis was conducted across all frames of `sample_run`. In each frame, ORB features (Rublee et al. 2011), comprising keypoints and their corresponding

descriptors, were extracted along with a subset of 3D keypoints that included reconstructed spatial coordinates. For the reconstruction of local world coordinates corresponding to each keypoint, the camera intrinsics matrix, together with the raw depth image, was utilized. During the experiments, it was observed that the use of raw depth images provided superior performance compared to corrected depth images for the task of pose change estimation.

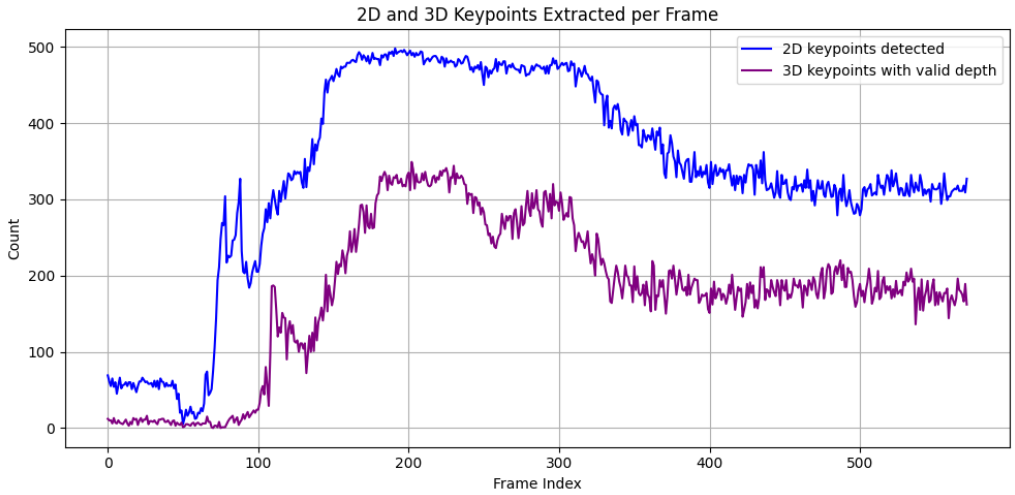

Figure 4.7: Number of 2D ORB keypoints detected per frame (blue) and the subset of 3D keypoints with valid depth information (purple) across the entire `sample_run` sequence.

Subsequently, a matching and registration process was implemented to identify correspondences across frames. Each detected feature with valid depth was either matched to an existing tracked feature based on descriptor similarity and 2D spatial proximity constraints or registered as a new feature if no match was found. Features were actively tracked across consecutive frames, and an age-based pruning strategy was applied to discard features older than 10 frames, maintaining computational efficiency while preserving the temporal continuity required for robust odometry estimation.

As can be seen in Figures 4.7 and 4.8, no significant features were detected before the weld section, reflecting the uniformity of the surface in the initial frames. Upon entering the weld section, there was a sharp increase in both the number of 2D keypoints with valid depth and the count of new and matched features, indicating the presence of distinct geometric and texture variations characteristic of the weld sections. After passing the weld section, the number of detected features decreased but remained considerably higher than in the

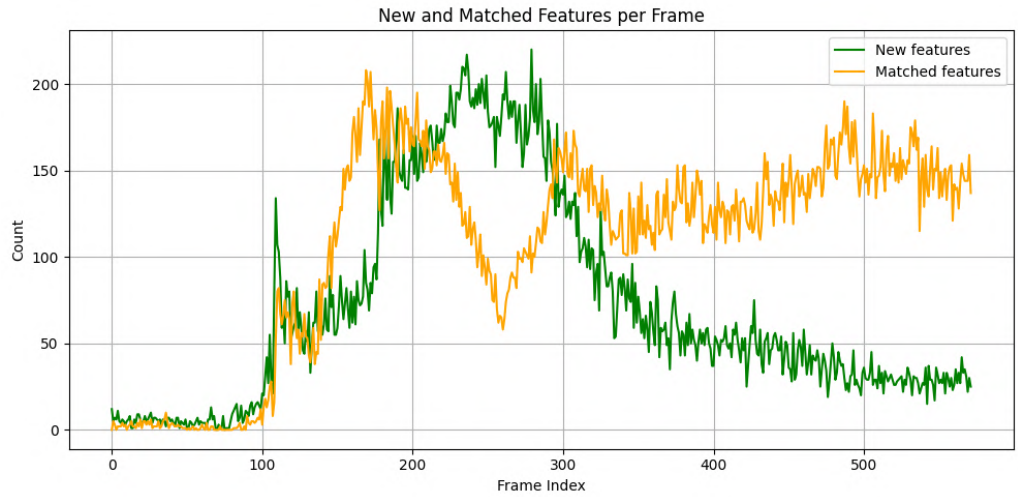

Figure 4.8: Count of new features detected per frame (green) and matched features tracked from previous frames (orange) throughout the entire `sample_run` sequence.

pre-weld region, as the weld section remained partially within the camera's field of view (background). Finally, Figure 4.9 shows the total number of actively tracked features over time, clearly highlighting the feature-rich weld sections compared to the featureless rolled sections. These plots collectively visualize the inherent challenges faced when performing feature-based odometry in featureless environments, where a lack of distinctive visual cues limits both detection and tracking performance.

Feature-Based Pose Change and Trajectory Estimation

To estimate the camera motion between consecutive frames, an ORB feature-based method was employed. For each incoming frame, ORB keypoints with valid (raw) depth values were extracted and matched—based on descriptor similarity—against previously tracked features. When a sufficient number of correspondences was found, the relative pose between frames was computed by solving the 3D–2D Perspective-n-Point (PnP) problem, using the 2D keypoints as image points and the corresponding 3D points from the previous frame as object points. This yielded a sequence of 4×4 transformation matrices describing the camera's trajectory through space. Under the assumption that the camera remains centered within the pipe, rotational components were omitted from the transformation, and only the translational vector was used. The trajectory was computed incrementally, with each new pose obtained by applying the estimated translation to the pose from the preceding frame. For this purpose, the dataset `sample_run_syn` was utilized, consisting of 773 frames.

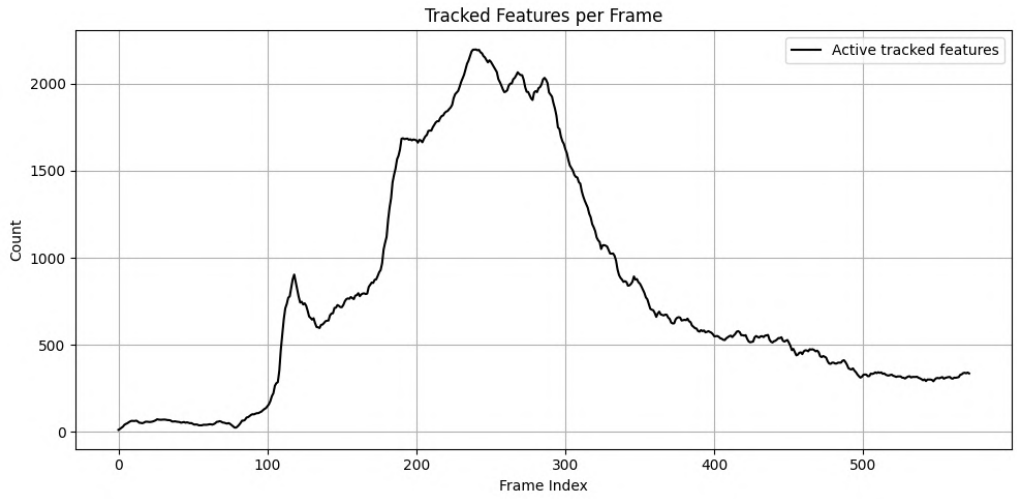

Figure 4.9: Total number of actively tracked features across the entire `sample_run` sequence. An age-based pruning strategy was applied to discard features that had not been matched for more than 10 frames. Each feature's age was incremented after every frame and reset upon a successful match, maintaining a robust yet computationally efficient tracking performance.

In the initial portion of the sequence, corresponding to the first rolled section, consistent pose estimation was hindered by a lack of distinctive visual features, leading to insufficient matches between frames. To evaluate performance, a representative segment starting at frame 146 was selected, and the trajectory was computed over 300 frames. The estimated trajectory exhibited a systematic bias arising from both the properties of the Intel RealSense sensor and the employed method. Since both color and depth were registered using the left monochrome imager, the stereo baseline introduced a consistent angular offset when the camera was centered inside the pipe, causing a cumulative lateral deviation. Furthermore, the trajectory length was underestimated by a factor of approximately 2.5 relative to the actuator-assumed motion, indicating significant compression over extended sequences.

Despite these biases, the use of 3D–2D correspondences improved frame-to-frame alignment and reduced visible gaps in the reconstructed point cloud. However, over more extended sequences, the underestimation of translational motion led to overlapping regions and reduced global consistency. In summary, the method provides accurate short-term alignment but is affected by sensor-specific biases and environmental constraints in long-term tracking.

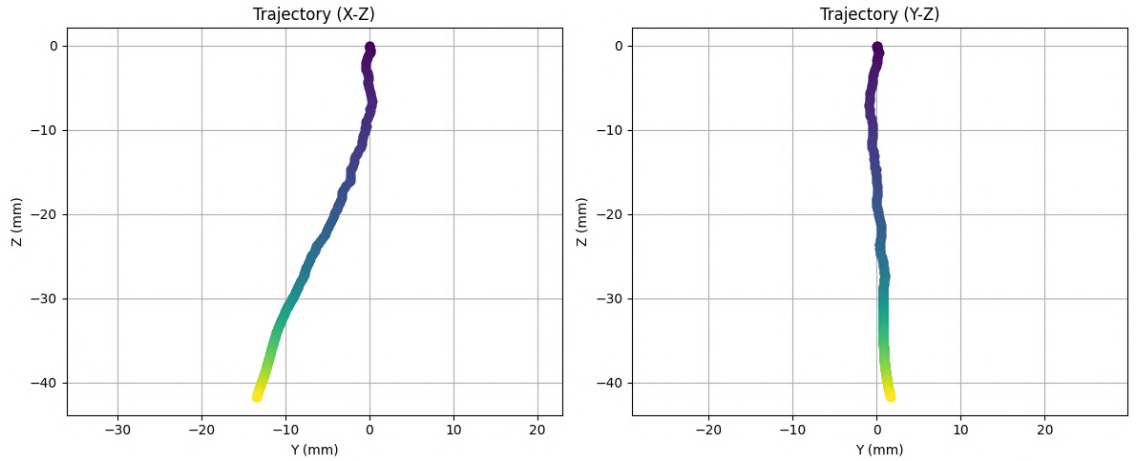

Figure 4.10: Estimated camera trajectory for the `sample_run_syn` dataset over 300 frames, starting at frame 146. The trajectory was computed using ORB feature correspondences and aggregated pose changes, illustrating short-term alignment accuracy and cumulative biases over the sequence.

Reconstruction Results

Reconstruction results obtained from corrected RGB-D data under varied experimental conditions are presented to evaluate the proposed processing algorithm and to compare feature-based pose estimation with actuator-based motion assumptions at constant velocity. The actuator-based approach assumes uniform linear motion of 12 mm/s along the negative Z-axis, corresponding to an inter-frame displacement of 0.4 mm at 30 FPS, without rotation. In contrast, the feature-based method estimates camera poses by extracting and matching visual features between consecutive RGB-D frames, solving a 3D–2D Perspective-n-Point (PnP) problem without additional global optimization such as bundle adjustment or loop closure.

Prior to reconstruction analysis, frame-wise geometry was characterized by fitting circles to point cloud slices segmented along the longitudinal axis. For each slice, circle parameters (center coordinates and radius) were extracted and averaged across corresponding slices over multiple frames to obtain representative cross-sections. Results reveal that fitted radii exceed nominal values by 15 to 27%, increasing with depth, accompanied by a consistent leftward shift of circle centers—attributable to the left imager's perspective. This parametrization reflects the average geometry over 720 frames in the dataset "`sample_run_syn`."

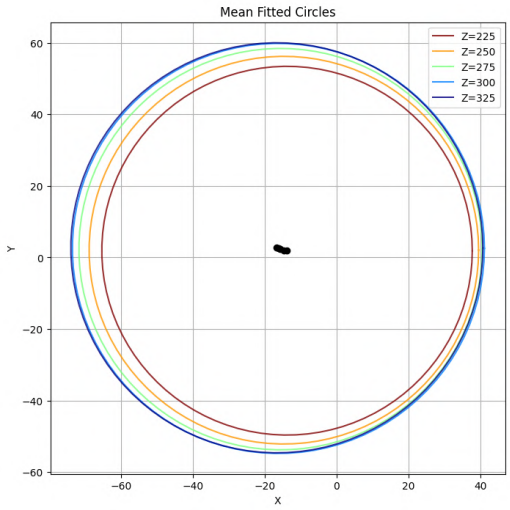

Figure 4.11: Average cross-sectional parameters obtained by fitting circles to point cloud slices along the longitudinal axis. Data represent an aggregate of 720 frames from the "sample_run_syn" dataset.

The second rolled pipe section (frames 345 to 704) was fully reconstructed using actuator-assumed poses, establishing a baseline for further evaluation. The reconstruction captures fine geometric detail under stable lighting conditions. Lens distortions lead to enlarged feature dimensions, and the effect of the left imager's field of view is apparent, with a higher density of points on the right side of the pipe. Despite these factors, the constrained, linear, and constant motion assumption—without external inputs—produces globally consistent reconstructions over extended sequences, confirming correct centering and processing.

Feature-based pose estimation improves alignment and detail fidelity over short intervals compared to actuator-based results. However, over longer durations, trajectory compression introduces overlapping, reducing global consistency. The weld section poses the most significant challenge due to diminished data quality and harsh lighting, resulting in increased distortion and sparser point clouds. Nonetheless, examination of the transition zones between rolled and weld sections demonstrates that feature-based visual odometry significantly enhances local alignment, mitigating gaps and improving structural coherence in these critical regions.

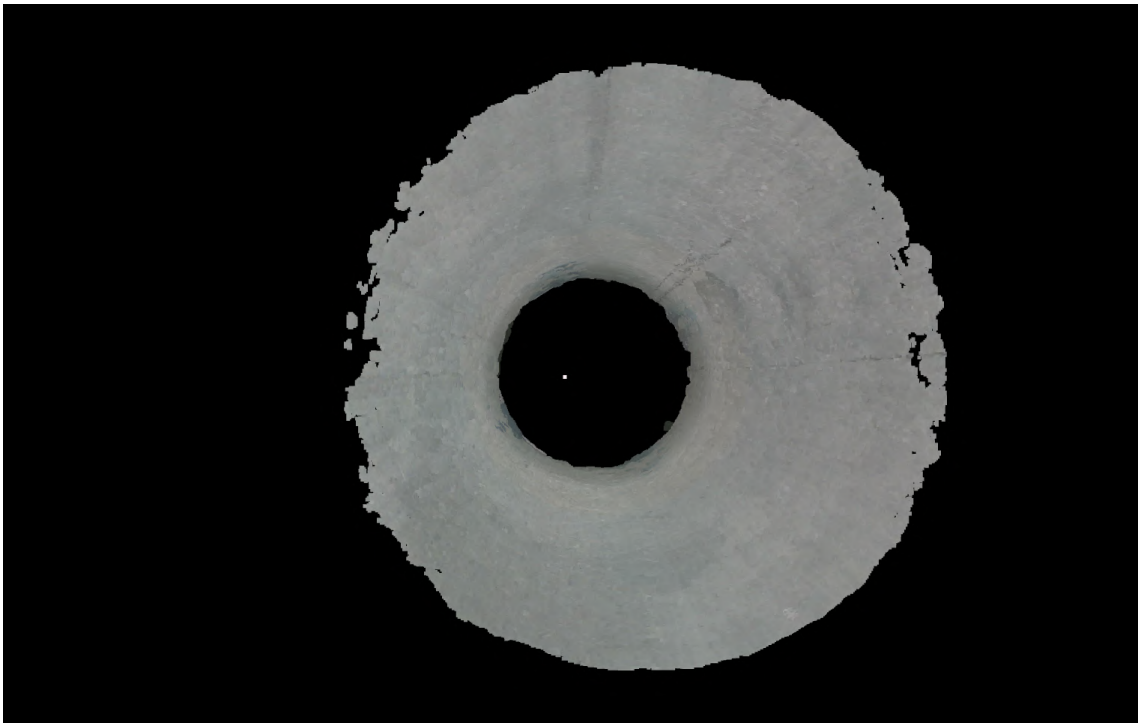

Figure 4.12: Reconstruction of the second rolled pipe section (frames 345–704) using actuator-assumed poses, showing detailed surface and effects of lens distortion and imager field of view.

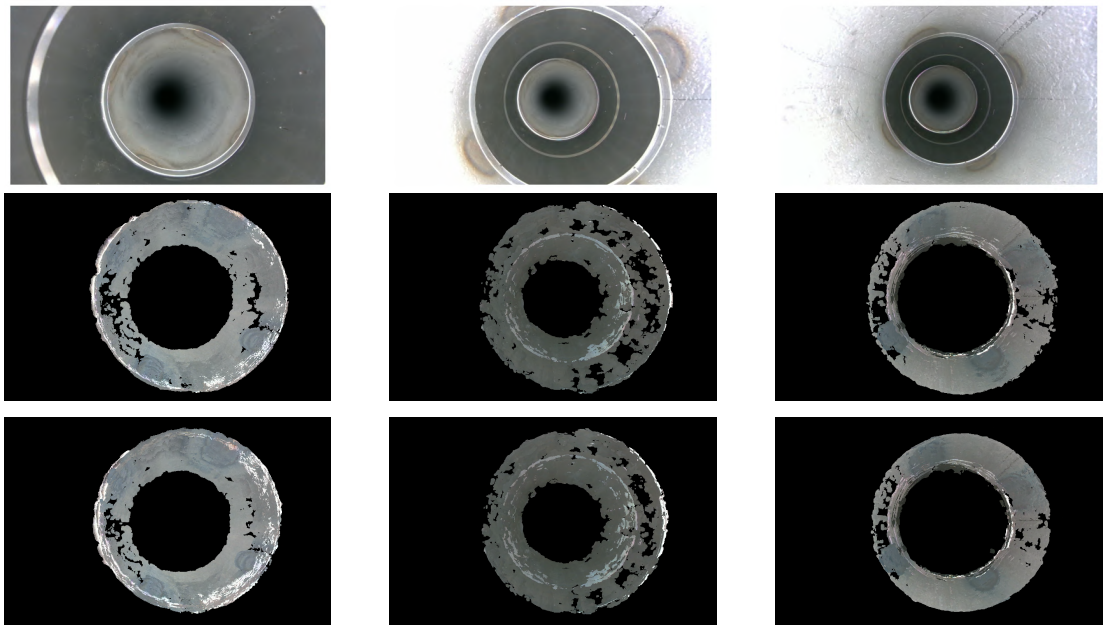

Figure 4.13: Comparison of reconstruction results at three locations along the pipe: entrance to the weld section (left), weld center (middle), and exit from the weld section (right). Top row: color images for reference. Middle row: point cloud reconstructions using actuator-assumed poses. Bottom row: reconstructions using feature-based pose estimation.

5 Discussion and Conclusion

Interpretation of Results

Developing a viable fusion power plant requires addressing significant challenges in plasma control, materials performance, and integrated plant engineering, while systematically reducing risks through targeted R&D activities (Federici, G. and Biel, Wolfgang and Gilbert, M.R. and Kemp, R. and Taylor, N. and Wenninger, R. 2017). To support this process, Technology Readiness Levels (TRLs) provide a structured scale—from 1, indicating basic research, to 9, signifying fully proven technologies in mission operations—that facilitates consistent assessment of technology maturity and comparison across different approaches (Mankins 1995).

Within the DEMO Service Joining Strategy, a key technical challenge is the limited maturity of non-destructive testing methods capable of reliably verifying weld integrity, identifying specific surface defects, while localizing both the robotic platform within the bore and the detected defects with high precision. Notably, RGB-D cameras offer distinct advantages by simultaneously capturing color and depth information, enabling the generation of detailed 3D point clouds. The availability of such data facilitates a variety of downstream applications, ranging from real-time robot localization and mapping to automated defect detection and classification. The wide availability and low cost of stereographic RGB-D cameras, combined with the growing number of open-source RGB-D data processing algorithms, make them particularly attractive for these applications.

The DEMO strategy currently relies on the evaluation of off-the-shelf solutions to assess technology maturity and identify R&D priorities. Within this framework, the Intel RealSense D405 camera was deployed to investigate the feasibility of in-bore RGB-D imaging. The experiments demonstrated both potential and limitations: while detailed point clouds were generated and feature-based visual odometry improved mapping accuracy over short sequences, depth estimates were degraded by systematic distortions induced by firmware assumptions, sensitivity to lighting variations. These challenges indicate that, although

promising for localization and mapping, the camera requires robust error mitigation strategies and further system integration prior to deployment in DEMO. On this basis, the work corresponds to Technology Readiness Level 4 , where fundamental components are integrated and validated under controlled laboratory conditions (Mankins 1995) , confirming the adaptability of RGB-D imaging to in-bore environments while acknowledging its current low-fidelity validation relative to the final system.

Suggested Improvements and Future Work

While the implemented approach proved effective within its defined scope, several limitations constrain its broader applicability and performance. The current method relies on black-box firmware depth estimation with limited access to algorithmic parameters, which prevents adaptation to the confined in-bore environment. The absence of bundle adjustment or loop closure restricts global map consistency, while the offline-only operation prevents real-time deployment. Computational demands remain high compared to typical SLAM pipelines, limiting the feasibility of using the system on embedded platforms.

Additional challenges were observed in relation to sensing and mapping quality. Performance degraded under sub-optimal lighting due to the absence of active illumination control, while the feature-poor pipe environment hindered reliable feature-based pose estimation. Systematic depth errors were only partially corrected using a simple mirroring approach, and random errors were not explicitly handled. A more rigorous treatment would involve analyzing error consistency across multiple frames—such as by adopting a sliding-window method that separates transient noise from persistent structural information. Moreover, the reconstructed maps have not yet been validated against ground-truth geometry, which limits their value for digital twin applications. Finally, radiation effects on sensing performance were not addressed, leaving an important gap for deployment in nuclear environments.

To overcome these limitations, several directions for future work are recommended. First, full system integration should be pursued, combining the IMU and RGB-D camera for improved localization, incorporating laser scanning for high-fidelity surface reconstruction, and fusing actuator data for coordinated inspection. Active lighting control should be introduced to ensure depth quality under varying illumination conditions. Real-time processing is also essential for operational use, requiring optimization of the pipeline for embedded platforms. In addition, the system's performance in bent pipe sections must be investigated to extend applicability beyond straight segments. A decision should also be made on whether to continue with the Intel RealSense D405—potentially reconfiguring it to exploit both left and

right IR imagers for custom stereo depth estimation with geometric priors—or to replace it with a fully custom stereo RGB-D solution, providing complete control over hardware and depth computation.

Conclusion

This study demonstrated the feasibility of using RGB-D cameras for in-bore localization and mapping in DEMO-like pipe environments, confirming that the Intel RealSense D405 can produce detailed point clouds and support short-term feature-based pose estimation. Key challenges were identified, including systematic depth distortions, sensitivity to lighting, and the feature-poor, repetitive geometry of pipe interiors, which limit accuracy and robustness. Overall, the work achieves Technology Readiness Level 4, providing a laboratory-validated foundation for further development toward integrated, real-time inspection systems suitable for deployment in nuclear fusion environments.

References

Azka, A. and M. Mittwollen (2024). Preliminary Investigation of Allowable Weld Defect Size and Orientation for DEMO Cooling Pipes. *IEEE Transactions on Plasma Science 52*(9), p. 3634 – 3637. Cited by: 0.

Bachmann, C., S. Ciattaglia, F. Cismondi, T. Eade, G. Federici, U. Fischer, T. Franke, C. Gliss, F. Hernandez, J. Keep, M. Loughlin, F. Maviglia, F. Moro, J. Morris, P. Pereslavtsev, N. Taylor, Z. Vizvary and R. Wenninger (2018). Overview over DEMO design integration challenges and their impact on component design concepts. *Fusion Engineering and Design 136*, p. 87–95. Special Issue: Proceedings of the 13th International Symposium on Fusion Nuclear Technology (ISFNT-13).

Batistoni, P., S. C. Lorenzo, K. Kurzydlowski, D. Maisonnier, G. Marbach, M. Noe, J. Paméla, D. Stork, J. Sanchez, M. Tran and H. Zohm (2010, March). Report of the Ad hoc Group on DEMO Activities. Chair: M.Q. Tran.

Crofts, O. and J. Harman (2014). Maintenance duration estimate for a DEMO fusion power plant, based on the EFDA WP12 pre-conceptual studies. *Fusion Engineering and Design 89*(9), p. 2383–2387. Proceedings of the 11th International Symposium on Fusion Nuclear Technology-11 (ISFNT-11) Barcelona, Spain, 15-20 September, 2013.

Crofts, O., A. Loving, D. Iglesias, M. Coleman, M. Siuko, M. Mittwollen, V. Queral, A. Vale and E. Villedieu (2016). Overview of progress on the European DEMO remote maintenance strategy. *Fusion Engineering and Design 109-111*, p. 1392–1398. Proceedings of the 12th International Symposium on Fusion Nuclear Technology-12 (ISFNT-12).

Federici, G., R. Kemp, D. Ward, C. Bachmann, T. Franke, S. Gonzalez, C. Lowry, M. Gadomska, J. Harman, B. Meszaros, C. Morlock, F. Romanelli and R. Wenninger (2014). Overview of EU DEMO design and R&D activities. *Fusion Engineering and Design 89*(7), p. 882–889. Proceedings of the 11th International Symposium on Fusion Nuclear Technology-11 (ISFNT-11) Barcelona, Spain, 15-20 September, 2013.

Federici, G. and Biel, Wolfgang and Gilbert, M.R. and Kemp, R. and Taylor, N. and Wen-

ninger, R. (2017). European DEMO design strategy and consequences for materials. *NUCLEAR FUSION 57*(9), p. 26.

Hirschmüller, H. (2005, June). Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, p. 807–814. IEEE.

Horaud, R., M. Hansard, G. Evangelidis and C. Ménier (2016). An Overview of Depth Cameras and Range Scanners Based on Time-of-Flight Technologies. *Machine Vision and Applications 27*(7), p. 1005–1020.

Intel Corporation (2025). Realsense™ D400 series product family datasheet. `https://dev.realsenseai.com/docs/intel-realsense-d400-series-product-family-datasheet`. Accessed: 18-Aug-2025.

International Atomic Energy Agency (2023). Demonstration Fusion Plants. `https://www.iaea.org/bulletin/demonstration-fusion-plants`.

Jang, W., C. Je, Y. Seo and S. W. Lee (2013). Structured-light stereo: Comparative analysis and integration of structured-light and active stereo for measuring dynamic shape. *Optics and Lasers in Engineering 51*(11), p. 1255–1264.

Kanade, T. and M. Okutomi (1994). A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*(9), p. 920–932.

Longuet-Higgins, H. C. (1981, September). A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature 293*(5828), p. 133–135.

Mankins, J. (1995, 01). Technology Readiness Level – A White Paper.

Mescheder, L. M., M. Oechsle, M. Niemeyer, S. Nowozin and A. Geiger (2018). Occupancy Networks: Learning 3D Reconstruction in Function Space. *CoRR abs/1812.03828*.

Mildenhall, B., P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi and R. Ng (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *CoRR abs/2003.08934*.

Pears, N., Y. Liu and P. Bunting (2012). *3D Imaging, Analysis and Applications*. SpringerLink : Bücher. Springer London.

Rublee, E., V. Rabaud, K. Konolige and G. Bradski (2011). ORB: An efficient alternative to SIFT or SURF. In: *2011 International Conference on Computer Vision*, p. 2564–2571.

Sarlin, P., D. DeTone, T. Malisiewicz and A. Rabinovich (2019). SuperGlue: Learning

Feature Matching with Graph Neural Networks. *CoRR abs/1911.11763*.

Smith, R., M. Self and P. C. Cheeseman (2013). Estimating Uncertain Spatial Relationships in Robotics. *CoRR abs/1304.3111*.

Tomasi, C. and T. Kanade (1992). Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision 9*(2), p. 137–154.

Tremethick, T., S. Kirk, K. Keogh, A. O'Hare, E. Harford and B. Quirk (2020). Service Joining Strategy for the EU DEMO. *Fusion Engineering and Design 158*, p. 111724.

Zollhöfer, M., P. Stotko, A. Görlitz, C. Theobalt, M. Nießner, R. Klein and A. Kolb (2018). State of the Art on 3D Reconstruction with RGB-D Cameras. *Computer Graphics Forum 37*(2), p. 625–652.

Intel RealSense D405

The Intel® RealSense™ Depth Camera D405 is a compact, short-range stereo RGB-D imaging system. The device is composed of two primary components: the Vision Processor D4 and the Depth Module D401.

The Depth Module D401 integrates a pair of wide-angle stereo imagers with a baseline of 18mm and operates without the use of an infrared projector. Both left and right imagers utilize the OmniVision OV9782 global shutter sensor. The field of view (FOV) of the depth system is approximately 84° (horizontal), 58° (vertical), and 92° (diagonal). Depth acquisition is based on disparity estimation between the left and right images. Correspondence points are identified between the two stereo inputs, and depth values are computed for each pixel based on the horizontal shift (disparity) between matched features. These per-pixel depth values are processed into sequential depth frames, which collectively form a real-time depth video stream. The D405 delivers optimal performance within a working distance of 7 cm to 50 cm, and can detect objects with features smaller than 1 mm when positioned at 7 cm.

The device supports output resolutions up to 1280×720 pixels at frame rates of up to 90 frames per second for both depth and color streams. Connectivity to a host system is facilitated via USB 2.0, USB 3.1 Gen 1, or MIPI interfaces. The unit is powered through the USB interface, with a nominal power consumption of approximately 1.55W during operation. The mechanical enclosure measures 42×42×23mm and supports industry-standard 1/4-20 UNC and M3 threaded mounting options.

Due to its sub-millimeter depth accuracy and compact form factor, the Intel RealSense D405 is particularly well-suited for RGB-D localization and mapping tasks in restricted or tubular environments, such as internal pipeline inspection.

Reference: Intel® RealSense™ D400 Series Product Family Datasheet. Available at: `https://dev.realsenseai.com/docs/intel-realsense-d400-series-product-family-datasheet`

Code Snippets

Project Structure and Setup

GitHub Repository: `https://github.com/yusufacikel/demo_pipe`
Python Requirement: Python 3.9, 3.10, 3.11, or 3.12 is required.

```
demo_pipe/
|-- demos/
|   |-- frame_visualization.ipynb
|   |-- baseline.ipynb
|   |-- orb_vo.ipynb
|-- sample_data/
|   |-- run_xx/
|       |-- color/
|       |-- depth/
|-- utils/
|   |-- config.py
|   |-- data.py
|   |-- depth.py
|   |-- pcd.py
|   |-- vis.py
|-- README.md
|-- requirements.txt
|-- setup.sh
|-- setup.bat
```

Contents

- `sample_data/` – sample dataset containing raw color and depth images captured from the devised test bench mimicking a short stretch of the DEMO piping system connected by mechanical joints.
- `demos/frame_visualization.ipynb` – visualizes raw color and depth data of a single frame, demonstrates depth correction, clipping, and radial filtering effects, and generates point clouds

from the processed data.

- `demos/baseline.ipynb` – baseline 3D reconstruction using assumed poses based on a constant 12 mm/s linear actuator movement along the negative z-axis of the pipe.
- `demos/orb_vo.ipynb` – ORB-based visual odometry and RGB-D reconstruction pipeline.
- `utils/` – utility modules providing core functionalities such as configuration, RGB-D dataset loading, image and point cloud processing, and visualization tools for both images and point cloud data.

Setup Instructions

```
git clone https://github.com/yusufacikel/demo_pipe.git
cd demo_pipe

# Linux setup
chmod +x setup.sh
./setup.sh

# Windows setup
.\setup.bat
```

Algorithmic Overview

Main 3D Reconstruction Loop (`baseline.ipynb` and `orb_vo.ipynb`)

```
Initialize dataset and visualizer
Initialize an empty global point cloud and a pose list (or Registry
    for orb_vo)

For each frame in sequence:
    Load color and raw depth images
    Apply depth correction and depth clipping
    Create point cloud from corrected & clipped depth and color
    Apply radial filtering to the point cloud
    If orb_vo:
        Register frame in Registry to estimate current pose
        If pose estimation fails, skip this frame
    Else (baseline):
        Assign pose assuming fixed translation along z-axis
```

```
    Transform point cloud by current pose
    Accumulate transformed point clouds into global point cloud

Visualize the global reconstructed point cloud
```

Depth Correction (`utils/depth.py`)

```
Function correct_depth(depth):
    reference_depth = max value in depth
    lower_bound = reference_depth - 5
    depth_clipped = clip_depth(depth, lower_bound, reference_depth)

    result = fit_2dcircle(depth_clipped)

    Extract (cx, cy), r from result (as integers)

    Create circle_mask for pixels inside radius r at (cx, cy)
    Set correction_mask = circle_mask AND (depth > 0)

    corrected_depth =
        for pixels in correction_mask:
            new depth = 2 * reference_depth - original depth
        else: keep original depth

    return corrected_depth
```

Feature-Based Odometry (`orb_vo.ipynb`)

```
Define class TrackedFeature:
    Properties:
        unique_id : string
        first_seen_frame_idx : integer
        age : integer = 0
        match_list : list = empty list

    Method add_feature(frame_idx, keypoint, descriptor, xyz):
        Append (frame_idx, keypoint, descriptor, xyz) to match_list

Define class Registry:
```

```
Properties:
    active_features : list = empty list
    discarded_features : list = empty list
    poses : list = empty list

Method register_frame(i, color, depth, corrected_depth):
    frame_idx = i
    extracted_features = detect_ORBs(corrected_depth, grayscale
       (color))
    matched_features = []

    For each feature in extracted_features:
        If matches any tracked_feature in active_features based
            on descriptor and proximity:
            Add feature to tracked_feature
            Reset tracked_feature age
            Add tracked_feature to matched_features
        Else:
            Create new TrackedFeature with unique ID
            Add feature to it
            Append to active_features

    For each feature in active_features:
        Increment feature.age by 1
        If feature.age > 10:
            Move feature from active_features to
               discarded_features

    If first frame:
        Append identity pose to poses
    Else:
        T = estimate_pose_change(matched_features)
        If T exists:
            new_pose = last pose * inverse(T)
            Append new_pose to poses
        Else:
            Return (skip update)

Method detect_ORBs(depth, gray):
```

```
        Create ORB detector
        Create mask where depth > 0
        Detect keypoints and descriptors in gray image using mask
        Convert depth map to 3D points using camera intrinsics

        For each keypoint:
            Get corresponding 3D point from depth

        Return list of (keypoint, descriptor, 3D point) tuples

    Method estimate_pose_change(matched_features):
        Initialize object_points and image_points as empty lists

        For each tracked_feature in matched_features:
            Get previous and current matches from match_list
            If frames differ and previous 3D point is valid:
                Append previous 3D point to object_points
                Append current keypoint position to image_points

        If fewer than 6 points, return None

        Use solvePnP with object_points and image_points to
           estimate pose
        If solvePnP fails, return None

        Construct transformation matrix T with R and t

        Return T
```

Image Gallery