

Inline Photometrically Calibrated Hybrid Visual SLAM

Nicolas Abboud, Malak Sayour, Imad H. Elhadj, John Zelek, Daniel Asmar

Abstract—This paper presents an integrated approach to Visual SLAM, merging online sequential photometric calibration within a Hybrid direct-indirect visual SLAM (H-SLAM). Photometric calibration helps normalize pixel intensity values under different lighting conditions, and thereby improves the direct component of our H-SLAM. A tangential benefit also results to the indirect component of H-SLAM given that the detected features are more stable across variable lighting conditions. Our proposed photometrically calibrated H-SLAM is tested on several datasets, including the TUM monoVO as well as on a dataset we created. Calibrated H-SLAM outperforms other state of the art direct, indirect, and hybrid Visual SLAM systems in all the experiments. Furthermore, in online SLAM tested at our site, it also significantly outperformed the other SLAM Systems.

I. INTRODUCTION

The advent of Visual Simultaneous Localization and Mapping (V-SLAM) has been a cornerstone in realizing autonomous navigation and perception systems. Advancements in V-SLAM have been realized with the development of hybrid V-SLAM systems that merge direct and indirect methods, and leverage the strengths of both methods to overcome their inherent limitations. Direct methods [1], which estimate motion and structure directly from pixel intensity variations are highly susceptible to variations in illumination, often resulting in decreased robustness and accuracy in dynamically lit environments [2]. The integration of indirect methods [3] (which rely on feature extraction and matching) with direct methods offers a balanced solution, enhancing the system’s adaptability and reliability across a wide range of scenarios by ensuring stable feature tracking even under challenging lighting conditions.

Photometric calibration plays a crucial role in further bolstering the capabilities of the direct component within hybrid SLAM systems (Fig.1). By accurately mapping the scene’s radiance to the camera’s intensity values, photometric calibration addresses critical aspects such as radiometric calibration and vignetting compensation. Radiometric calibration, is required to estimate the Camera Response Function (CRF), thereby ensuring the authenticity of intensity values relative to real-world irradiance [4]. This is particularly vital in enhancing system robustness under fluctuating illumination

N. Abboud, M. Sayour, I. H. Elhadj, and D. Asmar are with the Vision and Robotics Lab, Maroun Semaan Faculty of Engineering and Architecture, American University of Beirut, 1107 2020, Riad El Solh, Beirut, Lebanon; email: nfa53@mail.aub.edu, ms423@aub.edu.lb, ie05@aub.edu.lb, john email, da20@aub.edu.lb.

J. Zelek is with the department of System Design Engineering at the University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada; email: jzelek@uwaterloo.ca.

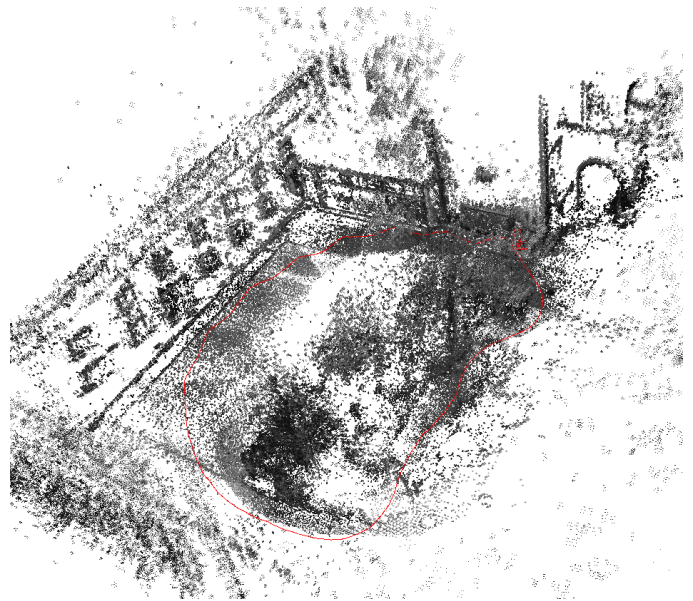


Fig. 1. Sample output of the photometrically calibrated hybrid SLAM system run on Sequence 30 of the TUM monoVO dataset.

conditions [2]. Vignette compensation, conversely, rectifies the radial intensity reduction towards the image periphery, a common lens-induced aberration, thereby achieving uniform sensitivity across the camera’s field of view. This uniformity is critical for improving the quality and reliability of feature detection and tracking in visual odometry.

Our proposed system integrates our Online Sequential Photometric Calibration (OSPC) technique [5] within our Hybrid direct-indirect visual SLAM (H-SLAM) system [6], utilizing exposure values from frame metadata to sequentially estimate the camera response function and vignette. The estimated photometric parameters are used to rectify the input frames to the V-SLAM system; thus improving the visual odometry module’s adaptability within the hybrid SLAM architecture, and providing a robust and consistent basis for feature tracking and depth estimation. This integration enriches the OSPC component’s effectiveness, which depends on precise feature extraction to identify corresponding points across frames. As a result, the enhanced feature extraction feeds back into the OSPC, leading to more accurate photometric calibration. This synergy creates a cycle within the system, where improved photometric calibration further bolsters the visual odometry component, thereby strengthening the entire SLAM framework’s performance and reliability.

The practicality and robustness of this advancement have been validated through both dataset and real-world testing, diverging from the conventional reliance solely on dataset evaluations. The real-world focus offers a comprehensive understanding of the system’s performance in dynamic and unpredictable environments.

This paper presents a real-time Visual SLAM system, that merges OSPC’s photometric estimation with H-SLAM’s efficient methodology. The source code is available at: https://github.com/AUBVRL/HSLAM_docker/tree/hslam_ospc. The key contributions are as follows:

- 1) **Unified Hybrid Direct-Indirect SLAM:** we leverage the strengths of both direct and indirect approaches to achieve robust camera pose estimation, demonstrating accuracy on par with or surpassing state-of-the-art monocular SLAM/Odometry system. This system efficiently computes both local and global map representations in a joint representation, enhancing the reutilization of global map points and reducing memory consumption significantly compared to traditional methods.
- 2) **Adaptive in-Line photometric calibration:** building upon OSPC’s approach to photometric calibration, our system implements this technique adaptively and in-line within a SLAM pipeline, directly contributing to the overall improvement of hybrid V-SLAM performance.
- 3) **Enhanced accuracy and robustness in feature-deprived environments:** our approach achieves enhancements in localization and mapping precision by uniquely keeping track of both pose-pose and co-visibility constraints, advancing robustness in feature deprived environments, a common challenge in real-world applications.
- 4) **Comprehensive experimental validation:** experiments and evaluations conducted in real-world scenarios validate the effectiveness and robustness of our proposed system. These tests highlight the system’s adaptability to various conditions and its practical applicability.

II. RELATED WORK

In this section we review the existing literature pertinent to our study, highlighting advancements and identifying gaps that our research aims to address.

A. Hybrid Visual SLAM: Bridging Direct and Indirect Methods

Feature-based methods involve the extraction and matching of a limited set of keypoints, aiming to minimize reprojection errors for pose estimation. This approach, as exemplified by the works of [7], [8], [3] and [9] relies on the robustness of modern feature extraction algorithms to handle variations in image intensity and geometric noise. However, it should be noted that their performance tends to deteriorate in environments lacking distinctive textures or when images

are affected by photometric noise, as discussed by [1]. Direct methods as introduced by [1], [10] and [11] do not rely on feature extraction. Instead, they estimate motion and structure by minimizing photometric errors directly, which represent discrepancies in image intensity. Nevertheless, it is important to acknowledge that these methods are highly sensitive to changes in image intensity caused by factors such as varying lighting conditions and camera exposure times [1]. Hybrid methods attempt to combine the strengths of both feature-based and direct methods. In the work of Lee and Civera [12], a loosely-coupled approach integrates a feature-based method and a direct method, allowing one to support the other when faced with challenges. However, this architecture, demands additional computational resources as it entails maintaining both methods simultaneously. Semi-direct Visual Odometry (SVO) [13], represents an efficient hybrid system that combines direct image alignment using initially detected features with Bundle Adjustment (BA) based on reprojection error for pose estimation refinement but lacks back-end optimization and re-localization. Other researchers opt to first optimize the reprojection error to obtain an initial coarse camera pose and then refine the result using a direct image alignment algorithm, as demonstrated by [14], [15], and [16]. This approach can be unified, as shown by Younes *et al.* [17], to yield a feature-assisted visual odometry model.

Recent research into this unified approach addresses the intricate challenges encountered by V-SLAM systems in representing local and global maps, primarily stemming from parametrization disparities between direct and indirect data sources [8] [3]. Indirect methods rely on multi-view geometry to construct point clouds (X, Y, Z) , while direct methods hinge on small frame-to-frame motion for inverse depth parametrization [18]. Additionally, this unified approach confronts the significant challenge faced by hybrid methods [12], [16], and [17] which is the non-interchangeability of triangulation techniques. H-SLAM [6] effectively surmounts these issues by adopting a versatile map representation that seamlessly adapts to function as either an (X, Y, Z) point cloud or an inverse depth parametrization, offering a comprehensive solution to the complexities faced by V-SLAM systems and hybrid approaches.

B. Photometric Calibration: Enhancing Image Understanding

In the domain of radiometric calibration for visual algorithms, a central concern revolves around the consistent maintenance of intensity across frames. Conventional approaches, as outlined by [19] and [20], typically rely on assumptions about the camera’s response function. Unfortunately, these assumptions can lead to inaccuracies due to non-monotonic response functions [4]. To tackle this problem effectively, Grossberg and Nayar introduced the Empirical Model of Response [21], which operates on single images but does not recover the vignette or the exposure times. Moreover, authors of [22] and [23] developed CRF recovery methods using color and grayscale imagery, but they do not

account for vignetting and exposure time variations. Multi-image static scene methods, like those developed by [24], [20], and [25], are not suitable for dynamic video sequences due to their reliance on multiple exposures of a static scene. Zheng *et al.* [26] proposed a vignette recovery approach, but they also neglect exposure time estimation and assume a pre-calculated CRF. Building on this foundation, Kim *et al.* [27] enhanced the process by optimizing exposure differences and optical flow simultaneously, thereby improving robustness in the face of changing image intensity. However, their approach demands a significant number of center-symmetric samples and utilizes a joint estimation method, limiting its real-time feasibility. More recently, Bergmann *et al.* [2] proposed an online photometric calibration system to recover radiometric parameters from auto-exposure videos. However, their method also employs joint estimation, requiring a point from the camera response function's ground truth data to resolve ambiguities. An ambiguity-free estimation for the camera response function was proposed by [28], but it ignores vignetting effects. We introduce [5] a novel photometric calibration method using sequential optimization to independently estimate the camera response function, vignette, and exposure values, thereby avoiding the ambiguities and computational complexities typical of joint optimization methods.

C. Photometrically Calibrated Hybrid Visual SLAM

Numerous recent endeavors have aimed to integrate photometric calibration techniques into V-SLAM systems to optimize the visual odometry module. In their work, Liu *et al.* [29] introduce a real-time photometrically calibrated monocular direct SLAM system, showcasing significant advancements; however, it does exhibit certain limitations. Specifically, this system solely relies on a direct sparse method, foregoing the versatility offered by hybrid approaches. Furthermore, the real-time photometric calibration technique proposed introduces computational complexity through a joint optimization approach, which potentially affects real-time performance. The system's effectiveness is significantly influenced by lighting conditions, possibly leading to error accumulation and impacting its long-term reliability in various operational scenarios.

Conversely, Luo *et al.* [16] introduce a Hybrid Sparse Odometry system, integrating feature-based and direct methods V-SLAM. Their system aims to enhance robustness against variations in image intensity and motion blur. It incorporates photometric calibration into the visual odometry component, but this process is also performed as a joint optimization, introducing complexity and ambiguities in exposure estimation that may not be resolved without ground truth data. This poses potential challenges to the system's accuracy and reliability, especially in real-world applications where processing speed is paramount. Furthermore, the study lacks an extensive comparison of their approach with existing methods, particularly in real-world diverse scenarios, leaving its practical effectiveness and adaptability to various conditions and environments relatively

unproven. In response to these limitations, we propose an innovative solution that marries a sequential photometric calibration method with a hybrid direct-indirect SLAM system, effectively addressing the shortcomings identified in prior research. Our approach simplifies the calibration process, reducing computational complexity while preserving accuracy. Rigorous testing across a range of operational scenarios underscores the system's reliability and precision in real-world environments. In summary, our hybrid SLAM system, enriched with streamlined photometric calibration, presents an enhancement in performance, accuracy, and adaptability in V-SLAM systems.

III. PROPOSED SYSTEM

This section presents a framework integrating the OSPC within the H-SLAM architecture, designed to enhance the robustness, accuracy, and adaptability in SLAM applications. The integrated system coalesces the strengths of OSPC's sequential photometric calibration with H-SLAM's advanced feature tracking, pose estimation, and map maintenance capabilities ensuring consistency and coherency of the spatial mapping.

Here, we detail the technical nuances of the proposed integrated framework, emphasizing the synergistic interplay between OSPC and H-SLAM components.

A. Descriptor sharing and feature extraction interplay

The ORB feature extraction utilized in H-SLAM plays an important role in ensuring feature consistency, which is foundational for OSPC's process. The descriptor sharing inherent in H-SLAM enhances OSPC by providing a rich, multi-faceted representation of each feature, which is necessary for the tracking and photometric rectification processes employed.

The robustness of ORB features, combined with the multi-descriptor approach, ensures that OSPC tracks features consistently across frames, a vital prerequisite for accurate photometric rectification. The rich data provided by descriptor sharing empowers OSPC's calibration algorithms, enabling them to accurately estimate photometric parameters, thereby refining the photometric model used across the system.

B. Sequential in-line photometric calibration

The OSPC's sequential photometric calibration operates as an in-line process within the H-SLAM architecture. The calibration is sequential and operates until a pre-defined validation threshold is reached, after which the estimated photometric parameters are set. Subsequent frames undergo photometric rectification based on these set parameters, maintaining consistency in the visual information processed by H-SLAM.

Integration into H-SLAM's multi-threaded architecture: OSPC is seamlessly integrated as a fourth thread in H-SLAM's multi-threaded architecture, originally consisting of dedicated threads for mapping, pose estimation, and loop closure (Fig. 2). This integration ensures that the sequential photometric calibration operates concurrently with H-SLAM's other processes without impeding performance. The

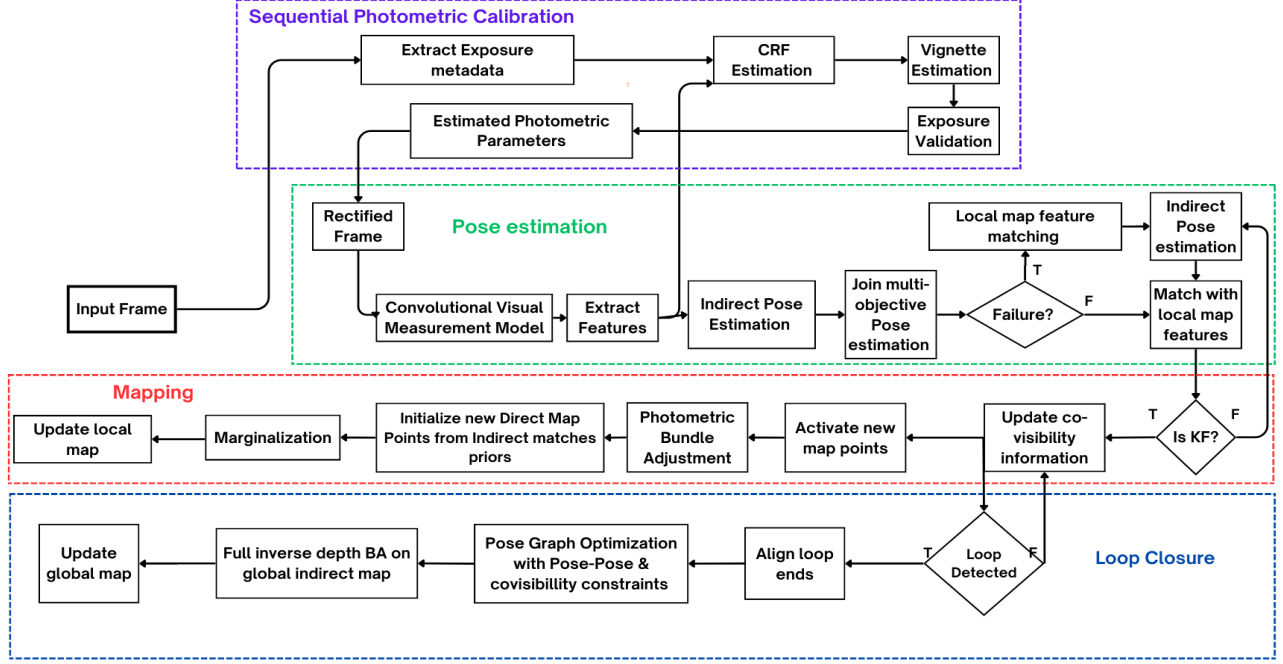


Fig. 2. Diagram of the proposed system showing the integration of sequential photometric calibration with the multi-threaded architecture of Hybrid SLAM

decoupled nature of this architecture allows OSPC to perform photometric rectification on each frame independently, contributing to the overall robustness and adaptability of the system without imposing additional computational burden on the core SLAM processes.

Mathematical formulations: the photometric calibration in OSPC optimizes photometric parameters as follows:

$$\frac{f^{-1}(M_1)}{f^{-1}(M_2)} = \frac{e_1}{e_2}, \quad (1)$$

representing the irradiance ratio between pairs of frames. f^{-1} represents the inverse camera response function, M_1 and M_2 are the corresponding intensity values, and e_1 and e_2 are the exposure value pairs. Similarly, the vignetting effect (V) is estimated by incorporating radial movement (R_1, R_2) alongside the CRF formulation:

$$\frac{f^{-1}(M_1)}{f^{-1}(M_2)} = \frac{e_1 V(R_1)}{e_2 V(R_2)}, \quad (2)$$

Exposure validation and CRF estimation are as follows:

$$k = \frac{1}{N} \sum \frac{f^{-1}(M_1) V(R_2)}{f^{-1}(M_2) V(R_1)}, \quad (3)$$

in which N stands for the total number of corresponding points per image pair. These equations are fundamental to OSPC and enable sequential optimization and validation of photometric parameters, ensuring accurate calibration within the multi-threaded H-SLAM framework without overburdening its core processes.

C. Harmonized integration: joint optimization, map maintenance and loop closure enhancement

The convergence of OSPC's photometric rectification with H-SLAM's joint optimization, hybrid connectivity graphs, and loop closure synergistically enhances the system's performance. It ensures not only geometric coherency and photometric accuracy but also computational efficiency in the loop closure process.

Optimized joint multi-Objective pose estimation: photometric rectification, facilitated by OSPC, plays an important role in minimizing residuals in H-SLAM's joint multi-objective pose estimation. By rectifying each frame, OSPC ensures that the input to H-SLAM's pose estimation module is void of photometric distortions, leading to a significant reduction in both geometric and photometric residuals. The photometric residuals are calculated using the same method as DSO [1], and the geometric residuals are the difference between the predicted and perceived key-point positions. The joint multi-objective pose optimization in H-SLAM is defined as an optimization problem:

$$\operatorname{argmin} e(\xi) = \operatorname{argmin} \left(\frac{\|e_p(\xi)\|_{H_\delta}}{n_p \sigma_p^2} + K \frac{\|e_g(\xi)\|_{H_\delta}}{n_g \sigma_g^2} \right), \quad (4)$$

where $\|\cdot\|_{H_\delta}$ denotes the Huber norm, $e_p(\xi)$ and $e_g(\xi)$ are the photometric and geometric residuals respectively, balanced by the number of features n_p, n_g , their variances σ_p^2, σ_g^2 , and a utility function K defined as:

$$K = \frac{5e^{-2l}}{1 + e^{\frac{30-N_g}{4}}}, \quad (5)$$

where l is the pyramid level and N_g is the number of current inlier geometric matches, dynamically adjusts the optimization's focus, prioritizing geometric residuals in early stages and shifting towards photometric residuals as optimization progresses.

This approach ensures that the input to H-SLAM's pose estimation module is rigorously optimized for both geometric coherence and photometric consistency, translating into more coherent and precise map updates, enhancing the overall quality and reliability of the spatial map maintained by H-SLAM.

Accurate hybrid connectivity graphs and loop closure enhancement: the reduced residuals and subsequent enhancements in mapping data ensure that the co-visibility and pose-pose constraints used in the hybrid connectivity graphs are based on precise data, enhancing the overall quality and utility of these graphs in the SLAM process. Precise co-visibility and pose-pose constraints in the hybrid connectivity graphs, facilitate accurate loop alignment and coherent loop closure. The accurate loop closure, aided by the consistent photometric information and the robust connectivity graphs, effectively minimizes drift over time. This contributes to the overall integrity and accuracy of the global map, ensuring that the system maintains a high level of performance even in extensive operational environments.

Impact on bundle adjustment: with enhanced feature matching and tracking, the input to the bundle adjustment process (*i.e.*, the correspondences between 2D features and 3D points) is more accurate, improving the precision of the 3D structure reconstructed by the inverse depth bundle adjustment. This makes the global mapping more robust and the pose estimation more reliable, even in dynamically changing environments.

By leveraging the synergy between photometric rectification, feature extraction, and joint multi-objective pose estimation the system achieves high levels of accuracy and robustness, while avoiding high computational costs.

IV. EXPERIMENTS

We validate our system through experiments conducted on both publicly available datasets and real-life scenarios. The experiments are designed to assess our system's performance under various conditions, comparing outcomes with and without our suggested photometric calibration approach, and benchmarking against established SLAM systems.

A. Dataset Evaluation

The first set of experiments focuses on evaluating H-SLAM performance on the TUM monoVO dataset [25], which originally comprises 50 sequences with photometric calibration, recorded across diverse environments, including both indoor and outdoor settings, as well as transitions from indoor to outdoor environments. However, due to the limited configuration of photometric calibration, we utilized only 28 sequences captured by a narrow FOV camera. This dataset demonstrates the impact of our proposed photometric approach in response to changes in lighting conditions.

H-SLAM undergoes a comparative analysis with the following SLAM systems:

- Direct Sparse Odometry (DSO) [1]: DSO employs a fully direct probabilistic model, minimizing photometric errors while jointly optimizing model parameters without the need for keypoint detectors or descriptors.
- ORB-SLAM3 [30]: ORB-SLAM3 is a feature-based SLAM system designed to support wide baseline loop closing and relocalization, with the added feature of full automatic initialization.
- Hybrid Sparse Monocular Visual Odometry (HSO) [16]: HSO utilizes direct image alignment with adaptive mode selection and employs ratio factors for image photometric description, enhancing robustness against substantial changes in image intensity and motion blur.

All experiments were performed on an Intel i7-9750H CPU with 16 GB RAM, without GPU parallelization. Real-time execution was enforced, and the GUI was disabled to enhance performance. To assess our photometrically calibrated H-SLAM, we utilized the TUM calibration sequence with the OSPC to derive the camera response function, vignetting, and exposure time for each frame.

B. Evaluation in diverse operational settings

Our approach to sequential photometric calibration requires the availability of exposure values from frame metadata. This information is not provided in most publicly available SLAM benchmarking datasets. Hence, to further evaluate the performance of H-SLAM we conducted a second set of experiments on campus at the American University of Beirut.

The experiments were conducted in an outdoor areas spanning approximately 30mx40m, and ground truth GNSS trajectory data were obtained using a Real-Time Kinematic (RTK) GPS setup, ensuring centimeter-level accuracy for positional data. This RTK-GPS system utilizes both a stationary and a rover GNSS receiver, employing differential corrections to achieve its high precision. The ground truth trajectory serves as a benchmark for evaluating the trajectories generated by each of the tested SLAM systems.

HSO, ORB-SLAM, DSO, and H-SLAM underwent testing in the recorded environment to assess their performance. Recognizing the probabilistic nature of SLAM systems, each system was run five times, and the best-performing result was considered for analysis. This approach ensures a comprehensive evaluation, capturing the variability inherent in SLAM system performance.

C. Live real-time evaluation

The third set of experiments took place indoors on Level 4 of the Irazy Oxy Engineering Complex (IOEC) at the American University of Beirut (Fig. 6). These experiments aimed to qualitatively assess the real-time performance of the SLAM systems. Utilizing the ROS architecture, all systems were simultaneously executed on the same machine. A single image frame captured by the camera was broadcasted across the ROS network, ensuring that each system received

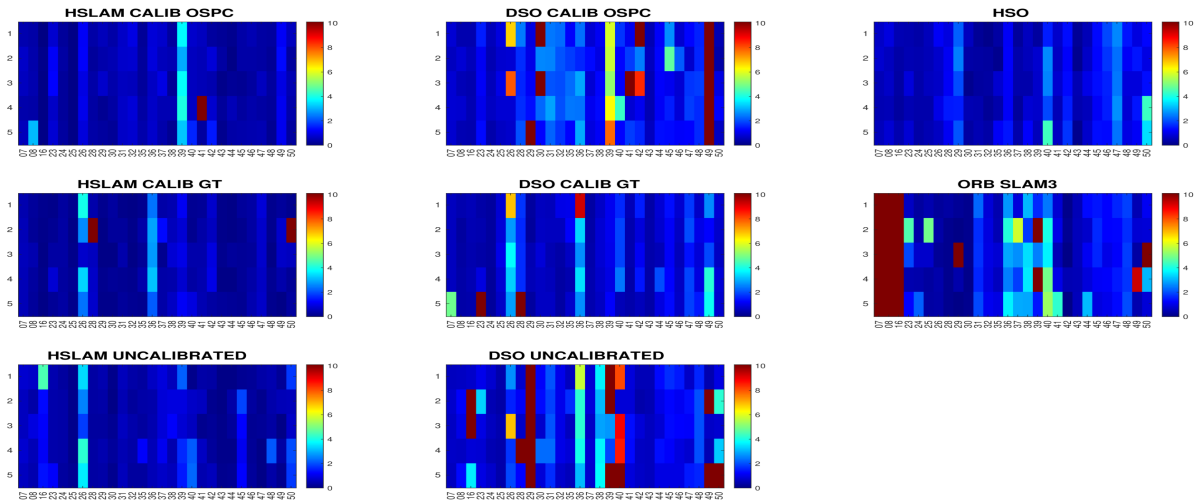


Fig. 3. Full evaluation results, showing cumulative alignment error for all tested sequences in the TUM-monoVO dataset. Each square corresponds to the color-coded alignment error, as defined in [25]. We run each of the tested sequence (horizontal axis) 5 times each (vertical axis).

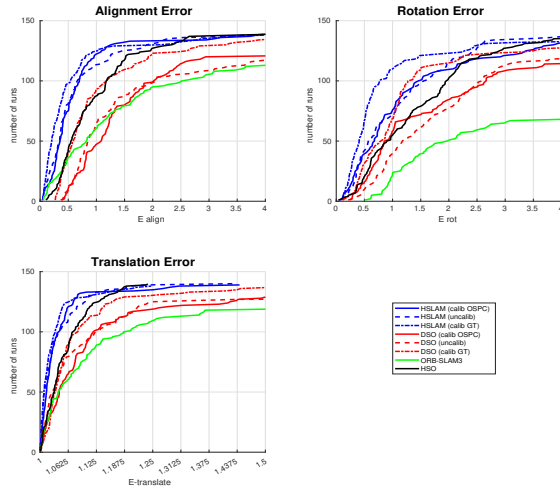


Fig. 4. Accumulated alignment error, rotational drift error and translational error as defined in [25] for each system over all runs

identical real-time input data. Key observations were focused on evaluating the quality of live performance and identifying any instances of failure among the SLAM systems. In all experiments, H-SLAM’s loop closure feature was disabled to ensure a fair comparison with DSO, preventing any potential bias that would skew the results in favor of H-SLAM.

V. RESULTS AND DISCUSSION

A. Dataset observations

The detailed evaluation outcomes for each sequence is present in Fig. 3. ORB-SLAM3 encounters challenges in Sequences 35 to 40, all of which involve indoor settings characterized by featureless hallways and plain classrooms. Conversely, the algorithm demonstrates strong performance in outdoor sequences (45-48 and 29), where the environment

was rich in distinctive features. Similarly, in the absence of photometric calibration, DSO faces difficulties mapping indoor sequences that lack distinctive features. However, its performance improves in such scenarios when combined with OSPC, reaching its optimal capability when utilizing the provided DSO calibration. Additionally, DSO encounters challenges in mapping Sequences 26 and 27, both of which involve staircases. It is important to highlight that the DSO variant incorporating loop closure, LDSO, exhibited poor performance along all sequences and thus was not reported in the results. This poor performance can be attributed to the TUM-VO dataset’s inherent nature, characterized by huge closed loops which makes it difficult to associate the current data with the historical data. On the other hand, both HSO and H-SLAM consistently exhibit good performance across the sequence variations. Both algorithms effectively handle mapping tasks in indoor featureless scenarios and outdoor settings which proves the robustness of the proposed hybrid systems. To facilitate a comprehensive comparison of the performance across the four algorithms, we calculated the cumulative alignment, rotation, and translation errors over the course of multiple runs, as defined in [1]. The resulting outcomes are presented in Fig. 4. The graphs demonstrate H-SLAM consistently outperforming other algorithms by most metrics, regardless of the applied photometric calibration. Remarkably, even a photometrically uncalibrated system utilizing H-SLAM exhibits better performance than photometrically calibrated DSO, online photometrically calibrated HSO, and ORB-SLAM3, underscoring the robustness of the proposed system. HSO performance closely follows in our evaluation, exhibiting approximately similar performance to the calibrated DSO when considering ground truth photometric values.

Conversely, the graphs reveal the sensitivity of DSO to the photometric component. An uncalibrated DSO system proves to be the least performing, with gradual performance

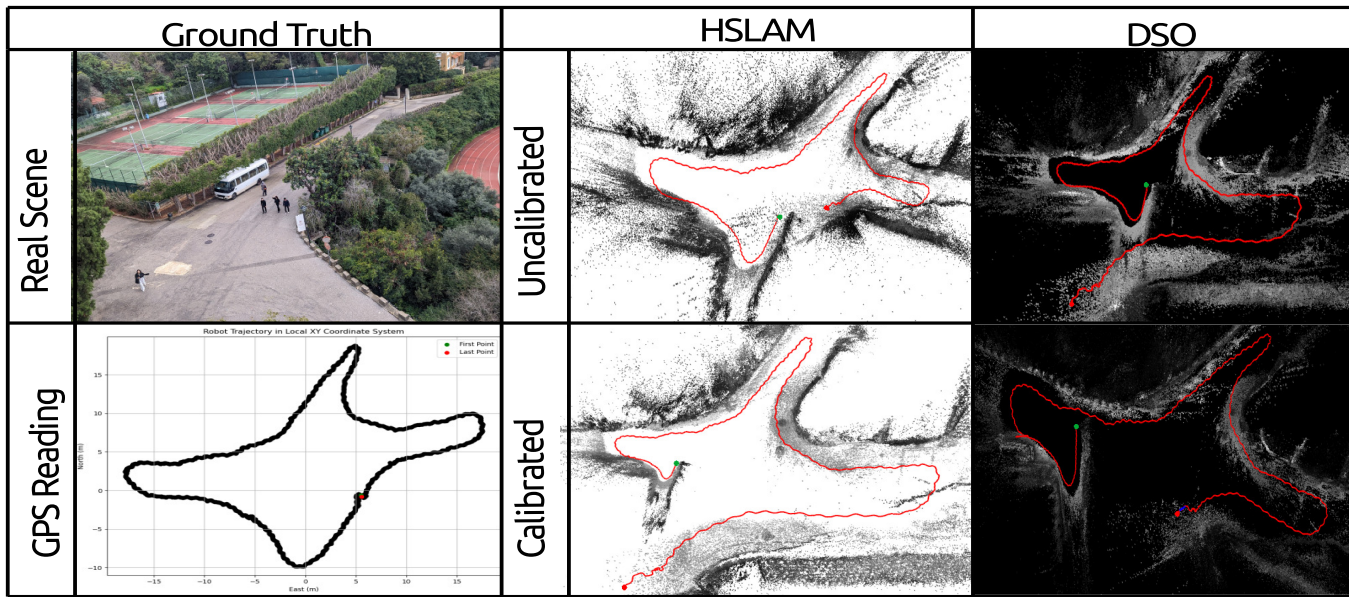


Fig. 5. Tabulated results of the live experiment conducted on campus at the American University of Beirut. The table shows the trajectories of HSLAM and DSO under two calibration modes alongside the "ground truth" obtained from RTK-GPS

improvements when OSPC calibration is applied, reaching peak performance when ground truth photometric values are utilized.

B. Outdoor Experiments Observations

When applied to the dataset captured at AUB, both HSO and ORB-SLAM did not perform well. While both systems initialized properly, they struggled to maintain tracking as the scene transitioned into frames featuring the asphalt ground with sparse features. In contrast, DSO and H-SLAM, with their direct modules, demonstrated robustness in handling frame sequences with sparse features. The results produced by DSO and H-SLAM are presented in Table 5. In the absence of photometric calibration, both DSO and H-SLAM exhibit significant error drift. However, upon the application of photometric calibration (OSPC) both DSO and H-SLAM showcased a notable reduction in geometric and photometric residuals, thereby mitigating the overall drift error.

A photometrically calibrated H-SLAM demonstrates more accurate pose estimation compared to its uncalibrated counterpart, resulting in a final trajectory that is closer to the groundtruth GNSS data

C. Live Real-Time Experiments

The results of the live real-time experiments are depicted in Fig. 6. In Trial 1, conducted in Region 1, the trajectory formed a loop within the lab, extending to the lab door leading to the corridor (Region 2). DSO initially tracks the trajectory but loses track when encountering a bright glare from the lab window. ORB loses tracking at the door and exhibits erratic behavior, and H-SLAM is stopped when ORB loses track. Trial 2 commences from Region 1 through the door and into the corridor (Region 2). ORB fails to initialize, while DSO initially runs but loses track at the corridor

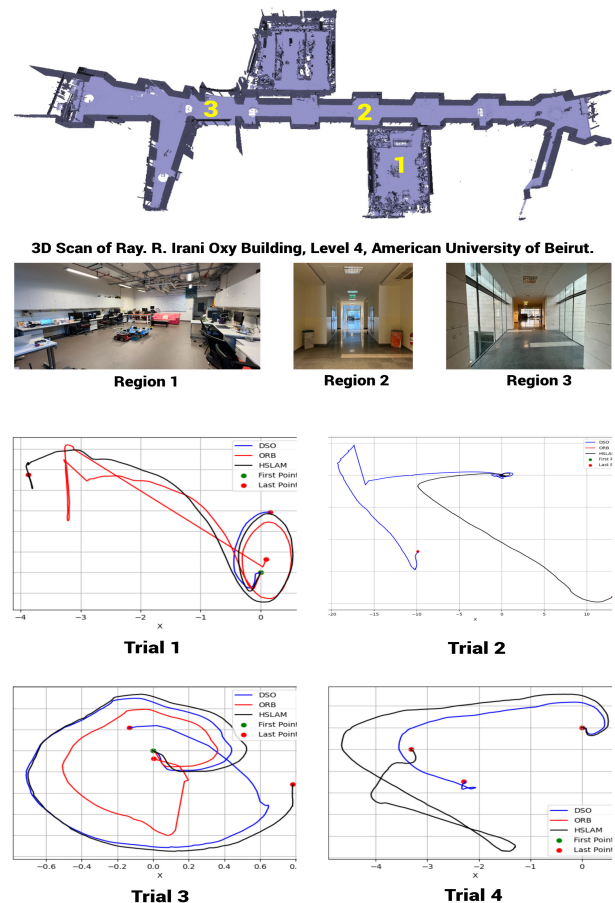


Fig. 6. Trajectory generated by H-SLAM, ORB-SLAM and DSO during the live experiments, above which is a 3D scan and images of the area used in testing

door before reinitializing. However, H-SLAM encounters no issues throughout the trial. Trial 3 involves a simple loop within Region 1, showcasing trajectories drawn by the systems in a confined space. ORB and DSO produce trajectories deviating from the actual motion, with DSO briefly losing track and reinitializing, while H-SLAM exhibits the most accurate trajectory. Trial 4, conducted in Region 2 with a turn at the glass walls (Region 3), sees ORB failing to initialize. DSO successfully tracks until encountering the turn at the glass walls, where it loses track entirely. H-SLAM operates without issues during this trial.

Overall, H-SLAM demonstrates greater stability and lower sensitivity to light changes compared to the other two systems, exhibiting a lower rate of tracking loss and re-initialization.

VI. CONCLUSION

This paper presented an enhanced performance of V-SLAM systems in real-world applications, integrating OSPC for improved photometric calibration and feature extraction within the H-SLAM framework. Experiments on two datasets demonstrated superior performance for our proposed calibrated H-SLAM versus state of the art in V-SLAM systems. We are currently working on integrating Inertial Measurement Units (IMU) into H-SLAM system to provide more robust solutions for navigation and mapping in complex environments.

ACKNOWLEDGMENT

This work was supported by the DIDYMOS-XR Horizon Europe project (grant number 101092875–DIDYMOS-XR, www.didymos-xr.eu), the Dean’s Office and the Vertically Integrated Projects Program at Maroun Semaan Faculty of Engineering and Architecture at AUB.

REFERENCES

- [1] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.
- [2] P. Bergmann, R. Wang, and D. Cremers, “Online photometric calibration of auto exposure video for realtime visual odometry and slam,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 627–634, 2017.
- [3] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “Orb-slam: a versatile and accurate monocular slam system,” *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [4] M. D. Grossberg and S. K. Nayar, “Determining the camera response from images: What is knowable?” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 11, pp. 1455–1467, 2003.
- [5] J. Haidar, D. Khalil, and D. Asmar, “Ospc: Online sequential photometric calibration,” *arXiv preprint arXiv:2305.17673*, 2023.
- [6] G. Younes, D. Khalil, J. Zelek, and D. Asmar, “H-slam: Hybrid direct-indirect visual slam,” *arXiv preprint arXiv:2306.07363*, 2023.
- [7] A. Geiger, J. Ziegler, and C. Stiller, “Stereoscan: Dense 3d reconstruction in real-time,” in *2011 IEEE intelligent vehicles symposium (IV)*. Ieee, 2011, pp. 963–968.
- [8] G. Klein and D. Murray, “Parallel tracking and mapping for small ar workspaces,” in *2007 6th IEEE and ACM international symposium on mixed and augmented reality*. IEEE, 2007, pp. 225–234.
- [9] T. Pire, T. Fischer, J. Civera, P. De Cristóforis, and J. J. Berles, “Stereo parallel tracking and mapping for robot localization,” in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2015, pp. 1373–1378.
- [10] J. Engel, J. Sturm, and D. Cremers, “Semi-dense visual odometry for a monocular camera,” in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1449–1456.
- [11] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, “Dtm: Dense tracking and mapping in real-time,” in *2011 international conference on computer vision*. IEEE, 2011, pp. 2320–2327.
- [12] S. H. Lee and J. Civera, “Loosely-coupled semi-direct monocular slam,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 399–406, 2018.
- [13] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, “Svo: Semidirect visual odometry for monocular and multicamera systems,” *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2016.
- [14] P. Kim, H. Lee, and H. J. Kim, “Autonomous flight with robust visual odometry under dynamic lighting conditions,” *Autonomous Robots*, vol. 43, no. 6, pp. 1605–1622, 2019.
- [15] N. Krombach, D. Droschel, and S. Behnke, “Combining feature-based and direct methods for semi-dense real-time stereo visual odometry,” in *Intelligent Autonomous Systems 14: Proceedings of the 14th International Conference IAS-14 14*. Springer, 2017, pp. 855–868.
- [16] D. Luo, Y. Zhuang, and S. Wang, “Hybrid sparse monocular visual odometry with online photometric calibration,” *The International Journal of Robotics Research*, vol. 41, no. 11-12, pp. 993–1021, 2022.
- [17] G. Younes, D. Asmar, and J. Zelek, “A unified formulation for visual odometry,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 6237–6244.
- [18] J. Civera, A. J. Davison, and J. M. Montiel, “Inverse depth parametrization for monocular slam,” *IEEE transactions on robotics*, vol. 24, no. 5, pp. 932–945, 2008.
- [19] S. Mann and R. Picard, “On being” undigital” with digital cameras: Extending dynamic range by combining differently exposed pictures, roceedings of is&t 48th annual conference society for imaging science and technology annual conference,” 1995.
- [20] T. Mitsunaga and S. K. Nayar, “Radiometric self calibration,” in *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*, vol. 1. IEEE, 1999, pp. 374–380.
- [21] M. D. Grossberg and S. K. Nayar, “What is the space of camera response functions?” in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2. IEEE, 2003, pp. II–602.
- [22] J. Gu and H.-Y. Shum, “Radiometric calibration from a single image,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2. IEEE, 2004, pp. II–II.
- [23] S. Lin and L. Zhang, “Determining the radiometric response function from a single grayscale image,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 2. IEEE, 2005, pp. 66–73.
- [24] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 643–652.
- [25] J. Engel, V. Usenko, and D. Cremers, “A photometrically calibrated benchmark for monocular visual odometry,” *arXiv preprint arXiv:1607.02555*, 2016.
- [26] Y. Zheng, S. Lin, C. Kambhamettu, J. Yu, and S. B. Kang, “Single-image vignetting correction,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 12, pp. 2243–2256, 2008.
- [27] S. J. Kim and M. Pollefeys, “Robust radiometric calibration and vignetting correction,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 4, pp. 562–576, 2008.
- [28] Z. Mo, B. Shi, S.-K. Yeung, and Y. Matsushita, “Ambiguity-free radiometric calibration for internet photo collections,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1670–1684, 2019.
- [29] Q. Liu, Z. Wang, and H. Wang, “Optical flow monocular visual-inertial odometry with online photometric calibration,” in *Journal of Physics: Conference Series*, vol. 1828, no. 1. IOP Publishing, 2021, p. 012163.
- [30] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, “Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam,” *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.