

Supplementary Material for: GroundSLAM: A Robust Visual SLAM System for Warehouse Robots Using Ground Textures

In this supplementary material, we provide detailed proofs of the theorems presented in the main paper (Section 1), a comprehensive description of the newly collected dataset (Section 2), and additional experimental results (Section 3). Specifically, the proof of Theorem 2 is provided in Section 1.1, the proof of Theorem 3 in Section 1.2, and the proof of Theorem 4 in Section 1.3. In Section 2, we describe the motivation for collecting the dataset, the data acquisition setup, and comparisons with existing datasets. In Section 3.1, we present a more detailed efficiency analysis. In Section 3.2, we present experiments conducted on a real warehouse robot. In Section 3.3, we provide a failure case analysis to further discuss the limitations of our method.

1. Proof of Theorem

1.1 Equivariance of KCC

In Theorem 2 of the paper, we show that KCC is equivariant to affine transformations. The proof is provided in this section.

Lemma 1 (Equivariance [1]). *Consider a function $\mathbf{f}$ and a transformation $\mathcal{T}$, the function $\mathbf{f}$ is equivariant to $\mathcal{T}$ if*

$$\mathbf{f}(\mathcal{T}(\mathbf{x})) = \mathcal{T}'(\mathbf{f}(\mathbf{x})) \quad (1)$$

for all input $\mathbf{x}$, where $\mathcal{T}'$ might be the same as $\mathcal{T}$ or another related transformation. In simple terms, a function is said to be equivariant if the input changes in a certain way, the output changes in a predictable and corresponding manner.

Theorem 2 (KCC Equivariance). *Kernel cross-correlator is equivariant to transformation $\mathcal{T}_i$, if $\mathcal{T}_i$ is periodical: $\mathcal{T}_i(\mathbf{x}) = \mathcal{T}_{i+m}(\mathbf{x})$, where m is the period. In other words,*

$$\mathbf{g}(\mathcal{T}_i(\mathbf{x})) = \kappa_{\mathbf{z}}(\mathcal{T}_i(\mathbf{x})) \otimes \mathbf{h} = (\kappa_{\mathbf{z}}(\mathbf{x}) \otimes \mathbf{h})_{(i)} = \mathbf{g}_{(i)}(\mathbf{x}), \quad (2)$$

where $\cdot_{(i)}$ denotes the vector circularly shifted by i elements.

Proof. Since circular translation is a periodical transform function and cross-correlation is equivariant to circular translation [1], we have $(\mathbf{a} \otimes \mathbf{b})_{(i)} = \mathbf{a}_{(i)} \otimes \mathbf{b}$, where $\mathbf{a}, \mathbf{b}$ are two random vectors. Assume $\mathbf{g}' = \kappa_{\mathbf{z}}(\mathcal{T}_j(\mathbf{x})) \otimes \mathbf{h}$, then

$$\mathbf{g}' = [\kappa(\mathcal{T}_j(\mathbf{x}), \mathcal{T}_0(\mathbf{z})), \dots, \kappa(\mathcal{T}_j(\mathbf{x}), \mathcal{T}_{m-1}(\mathbf{z}))] \otimes \mathbf{h}. \quad (3)$$

Since $\mathcal{T}_j$ is periodical, we have

$$\kappa_{\mathbf{z}}(\mathcal{T}_j(\mathbf{x})) = (\kappa_{\mathbf{z}}(\mathbf{x}))_{(j)}. \quad (4)$$

This means that

$$\kappa_{\mathbf{z}}(\mathcal{T}_j(\mathbf{z})) \otimes \mathbf{h} = (\kappa_{\mathbf{z}}(\mathbf{z}))_{(j)} \otimes \mathbf{h} = (\kappa_{\mathbf{z}}(\mathbf{z}) \otimes \mathbf{h})_{(j)}, \quad (5)$$

which completes the proof. $\square$

1.2 Closed-Form Solution to KCC

In Theorem 3 of the paper, we show that the following objective function

$$\min_{\hat{\mathbf{h}}^*} \|\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{h}}^* - \hat{\mathbf{g}}^*\|^2 + \lambda \|\hat{\mathbf{h}}^*\|^2, \quad (6)$$

admits a closed-form solution. In this section, we provide the proof of Theorem 3.

Theorem 3 (KCC). *There is a closed-form solution to the objective function (6) for kernel cross-correlator,*

$$\hat{\mathbf{h}}^* = \frac{\hat{\mathbf{g}}^* \odot \hat{\kappa}_{\mathbf{z}}^*(\mathbf{z})}{\hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) \odot \hat{\kappa}_{\mathbf{z}}(\mathbf{z}) + \lambda}, \quad (7)$$

where the operator $\div$ denotes the element-wise division and $\hat{\mathbf{h}}^*$ is the conjugate of Fourier transform of $\mathbf{h}^*$.

Proof. Without ambiguity, we will denote $\mathbf{g}^*$ as $\mathbf{g}$ in this proof for simplicity. To solve the optimization problem (6), we set its first derivative with respect to $\hat{\mathbf{h}}^*$ to zero, i.e.,

$$\frac{\partial}{\partial \hat{\mathbf{h}}^*} \left(\|\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{h}}^* - \hat{\mathbf{g}}\|^2 + \lambda \|\hat{\mathbf{h}}^*\|^2 \right) = 0. \quad (8)$$

Then by calculating the square ($\|a\|^2 = aa^*$), we have

$$\frac{\partial}{\partial \hat{\mathbf{h}}^*} \left(\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{h}}^* \odot \hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) \odot \hat{\mathbf{h}} - \hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{h}}^* \odot \hat{\mathbf{g}}^* - \hat{\mathbf{g}} \odot \hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) \odot \hat{\mathbf{h}} + \hat{\mathbf{g}} \odot \hat{\mathbf{g}}^* + \lambda \hat{\mathbf{h}}^* \odot \hat{\mathbf{h}} \right) = 0. \quad (9)$$

We next obtain the derivative w.r.t. $\hat{\mathbf{h}}^*$ by taking $\hat{\mathbf{h}}$ as an independent variable of $\hat{\mathbf{h}}^*$ [2], thus

$$0 = \hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) \odot \hat{\mathbf{h}} - \hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{g}}^* + \lambda \hat{\mathbf{h}}. \quad (10)$$

Therefore, we can obtain $\hat{\mathbf{h}}$ as

$$\hat{\mathbf{h}} = \frac{\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{g}}^*}{\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) + \lambda}. \quad (11)$$

Due to the properties of the complex conjugate, i.e., $a^* \cdot b^* = (a \cdot b)^*$, $a^*/b^* = (a/b)^*$, and $a^* + b^* = (a + b)^*$, the solution (7) can be obtained from (11) by taking the conjugate on both sides, which completes the proof. $\square$

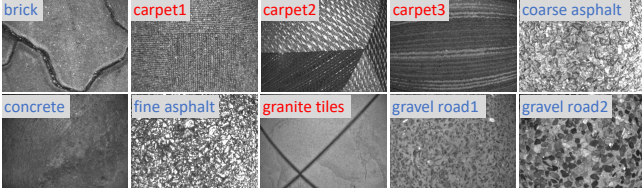


Fig. 1: Example ground texture images in our dataset. We collect 10 kinds of ground textures, including 6 kinds of outdoor textures and 4 kinds of indoor textures, which are marked in blue and red, respectively.

1.3 Closed-Form Solution to Weighted Regularized KCC

In Theorem 4 of the paper, we show that the weighted regularized KCC with objective function

$$\min_{\hat{\mathbf{h}}^*} \|\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\mathbf{h}}^* - \hat{\mathbf{g}}\|^2 + \lambda \|\sqrt{\hat{\kappa}_{\mathbf{z}}(\mathbf{z})} \odot \mathbf{h}\|^2. \quad (12)$$

admits a closed-form solution. In this section, we provide the proof of Theorem 4.

Theorem 4 (Weighted Regularized KCC). *There is a closed-form solution to the objective (12) for KCC:*

$$\hat{\mathbf{h}}^* = \frac{\hat{\mathbf{g}}^*}{\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) + \lambda}. \quad (13)$$

Proof. Since only the regularization of the objective function is changed, the solution to (12) can be obtained by updating the regularization term of (11), hence we have

$$\begin{aligned} \hat{\mathbf{h}} &= \frac{\hat{\mathbf{g}}^* \odot \hat{\kappa}_{\mathbf{z}}(\mathbf{z})}{\hat{\kappa}_{\mathbf{z}}(\mathbf{z}) \odot \hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) + \lambda \hat{\kappa}_{\mathbf{z}}(\mathbf{z})}, \\ &= \frac{\hat{\mathbf{g}}^*}{\hat{\kappa}_{\mathbf{z}}^*(\mathbf{z}) + \lambda}. \end{aligned} \quad (14)$$

Similarly, solution (13) can be obtained by taking the conjugate on both sides, which completes the proof. $\square$

2. PathTex Dataset

In this section, we introduce our newly collected dataset. To the best of our knowledge, only two ground-texture datasets are publicly available for evaluating localization systems: the Micro-GPS [3] and the HD Ground [4]. Both are primarily designed for evaluating relocalization systems based on prior maps, which introduces two limitations that make them less suitable for assessing VO and SLAM methods. (1) Limited overlap: These datasets aim for wide area coverage with fewer images, emphasizing visual aliasing effects [4]. Therefore, they downsample original sequences to reduce frame redundancy. This results in minimal overlap between adjacent frames, posing challenges for VO and SLAM evaluation. (2) Ground truth accuracy: Both datasets derive ground truth poses from image stitching rather than precise pose sensors. As a result, their accuracy depends heavily on the quality of feature detection and matching in the stitching process. However, we found that our KCC-based image matching achieves even higher accuracy than traditional feature-based methods (this is further validated as

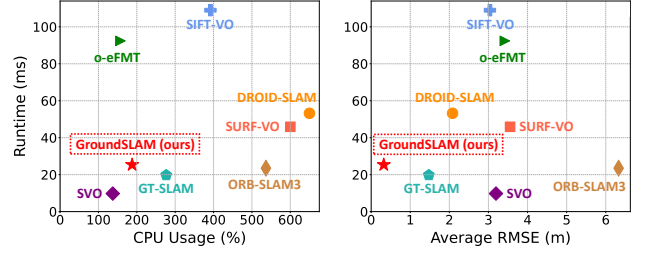


Fig. 2: Efficiency analysis for different systems on the GeoTracking dataset. The vertical axis is the average runtime of each frame. The horizontal axes are the CPU usage (left) and average RMSE (right), respectively.

demonstrated by the experimental results in Section V.B and Section V.C of the main paper.).

To address these limitations, we introduce the PathTex dataset. Data are collected using a warehouse robot equipped with a downward-facing camera mounted 0.1m above the ground. To ensure consistent illumination, the lighting system is mounted around the camera. For accurate ground-truth poses, a prism was mounted on the robot and tracked continuously by a Leica Nova MS60 MultiStation laser tracker. As shown in Fig. 1, PathTex includes 6 outdoor ground textures and 4 indoor ground textures. Compared to existing datasets, our dataset offers two key advantages: higher ground-truth accuracy and significantly greater frame-to-frame overlap. Measured by intersection-over-union (IoU), the average overlap in our dataset (93.69%) is more than twice that of the HD Ground dataset (47.74%), making it better suited for evaluating frame-by-frame visual odometry and SLAM systems.

3. Additional Experimental Results

3.1 Efficiency Analysis

This section presents the efficiency analysis of our GroundSLAM system. The evaluation is conducted on a computer with an Intel Core i9 CPU and the image resolution is 640×480 . We compare the running speed, computing resource consumption, and the accuracy of different systems. The metrics are per-frame runtime, CPU usage, and average RMSE. For sequences where tracking failures occur, we assign an error of 10 meters when computing the average RMSE. The results are presented in Fig. 2. Overall, our system achieves the highest accuracy while maintaining real-time performance with minimal computational resource consumption. GroundSLAM demonstrates a runtime comparable to GT-SLAM [5] and ORB-SLAM3 [6] while utilizing fewer CPU resources. Although o-eFMT [7] is also a correction-based ground-texture localization system, it exhibits significantly lower efficiency and accuracy compared to our approach, further validating the effectiveness of the proposed methods. It is also worth noting that DROID-SLAM [8] actually utilizes more than 6 CPU cores, and its CPU usage in Fig. 2 is only for a compact presentation.

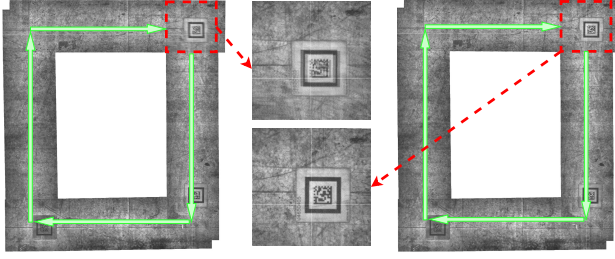


Fig. 3: A live mapping demo of our system without (left) and with (right) loop correction. The robot moves in a rectangular path (the green line) with the starting and ending being at the same location (the red square). The drift error of our VO module is only **0.2%** of the trajectory, which is significantly lower than other SOTA systems (about 1%). The drift error causes the blurring in the stitched map (left), and it is eliminated by our loop correction (right).

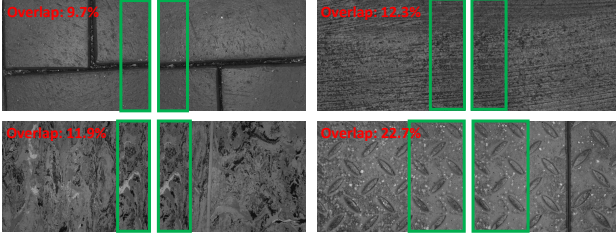


Fig. 4: Some failure cases of our system on the HD Ground dataset. As the publicly available sequences are sampled from the original captured sequences, the overlaps of some adjacent images in this dataset are very small.

3.2 Live Mapping Demo

We present a real-time mapping demonstration incorporating loop detection and correction. The data was collected in a warehouse environment using an Automated Guided Vehicle (AGV). As shown in Fig. 3, the robot followed a rectangular trajectory (highlighted in green), starting and stopping at the same location (marked by the red square), thereby forming a loop. Images along this path were then stitched using poses estimated by GroundSLAM, both without (left) and with (right) loop correction. Note that the tag was not utilized for pose estimation but served solely as a reference ground truth. Due to drift errors, the stitched map without loop correction exhibits noticeable blurring, which is effectively eliminated on the right. This clarity underscores the effective elimination of drift errors and the precise pose corrections. Additionally, we have tested GroundSLAM with this AGV in a real industrial environment spanning over 10,000 m^2 . The system exhibited consistent robustness and accuracy. Refer to <https://youtu.be/2sI76SDptDA> or <https://www.bilibili.com/video/BV1vWh8zTE77/> for more details.

3.3 Failure Case Analysis

We observed that many failure cases of our method stem from the low image overlap in the HD Ground dataset, as shown in Fig. 4. To further analyze the effect of overlap rate on performance, we uniformly sample the original PathTex sequences to create new sequences with varying overlaps. Fig. 5 shows the average overlap rates and localization errors

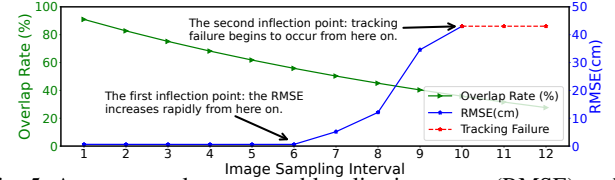


Fig. 5: Average overlap rates and localization errors (RMSE) under different sampling intervals on the “Coarse_asphalt” texture. When the overlap rate is less than 50%, the localization error begins to increase rapidly. When the overlap rate is less than 32%, the tracking failure begins to appear.

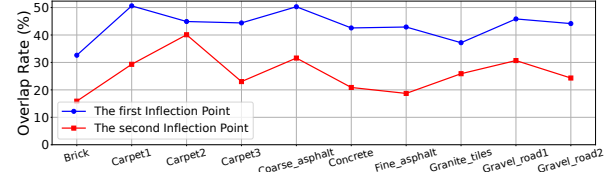


Fig. 6: The overlap rates of two inflection points for different ground textures. The average overlap rates of the first inflection point and the second inflection point are about 43% and 25%, respectively.

on the “Coarse_asphalt” texture under different sampling intervals. Two inflection points are observed: when the overlap rate drops below 50%, localization error increases rapidly, and when it falls below 32%, tracking failures begin to occur. Similar trends are observed across other textures. Fig. 6 summarizes the overlap rates at these inflection points for different ground textures, showing slight variations across textures. On average, the first and second inflection points occur at 43% and 25% overlap, respectively. Based on these findings, we conclude that for high-precision performance, GroundSLAM requires an overlap rate greater than 43% between adjacent images, while maintaining an overlap above 25% is necessary to prevent tracking failures.

In our real-world application, our warehouse robot operates at a maximum speed of 2.5m/s, ensuring a minimum image overlap of 53%. As a result, our system can provide reliable localization for our warehouse robots. To deploy our system on a faster-moving robot, a higher-frame-rate camera or a higher-mounted camera with a wider field of view can be used to ensure sufficient overlap between adjacent frames.

References

- [1] L. Gaunce Jr, J. P. May, M. Steinberger, *et al.*, *Equivariant stable homotopy theory*. Springer, 2006, vol. 1213.
- [2] D. Messerschmitt *et al.*, “Stationary points of a real-valued function of a complex variable,” *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2006-93*, 2006.
- [3] L. Zhang, A. Finkelstein, and S. Rusinkiewicz, “High-precision localization using ground texture,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6381–6387.
- [4] J. F. Schmid, S. F. Simon, R. Radhakrishnan, S. Frintrop, and R. Mester, “Hd ground-a database for ground texture based localization,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7628–7634.
- [5] K. M. Hart, B. Englot, R. P. O’Shea, J. D. Kelly, and D. Martinez, “Monocular simultaneous localization and mapping using ground textures,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 2032–2038.
- [6] 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.

- [7] W. Jiang, C. Li, J. Cao, and S. Schwertfeger, "Optimizing the extended fourier mellin transformation algorithm," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 3453–3460.
- [8] Z. Teed and J. Deng, "Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras," *Advances in neural information processing systems*, vol. 34, pp. 16 558–16 569, 2021.