Depth Refinement in 3D Mapping of Construction Sites Using a Stereo Camera

Ishiguro, R. [1*], Susaki, J.[2] , and Ishii, Y.[3]

[1]*Graduate School of Engineering, Kyoto University, Japan. ishiguro.ryunosuke.62w@st.kyoto-u.ac.jp*

[2]*Graduate School of Engineering, Kyoto University, Japan. susaki.junichi.3r@kyoto-u.ac.jp*

[3]*Graduate School of Engineering, Kyoto University, Japan. ishii.yoshie.4k@kyoto-u.ac.jp*

Abstract: To address a severe labor shortage of skilled crane operators in Japan's construction industry, there is a pressing need for automated three-dimensional (3D) environmental mapping. This research proposes a practical and robust method for generating high-fidelity 3D maps, which circumvents the need for large-scale training data required by deep learning models and compensates for the weaknesses of conventional stereo matching methods like Semi-Global Block Matching (SGBM) that fail in texture-poor regions. The core of the method is a depth refinement process based on the PatchMatch Multi-View Stereo (PM-MVS) framework, featuring three key enhancements. First, it is initialized with a dense depth map derived from a high-quality stereo disparity map for a robust starting state. Second, it employs a priority-based propagation scheme that expands from pixels with the lowest initial depth error to suppress error propagation. Third, the random search range is adaptively adjusted for each pixel based on its estimated error, focusing computational resources where most needed. An evaluation using a synthetic dataset confirmed the method's effectiveness, showing a significant reduction in Mean Absolute Error (MAE) from 4.4 to 0.30 compared to standalone SGBM. This approach produces a high-fidelity dense point cloud without the heavy data requirements of machine learning, providing an essential foundation for future automated crane control systems.

Keywords: Depth Map; PatchMatch MVS; Stereo Camera; Construction Stites; Stereo Matching

1. Introduction

In recent years, the Japanese construction industry has been facing a severe labor shortage, particularly a deficit of skilled crane operators, due to an aging workforce and a decline in the number of young participants. Automation of crane operations is a promising solution. This technology involves a system that perceives the environment in three-dimensional (3D) and autonomously plans and executes a safe transport path to move a suspended load to its destination. Previous research (Kobayashi et al., 2023) has demonstrated the effectiveness of mounting a sensor vertically downward at the tip of the crane boom for environmental perception. However, these studies utilized monocular cameras, which introduce the problem of scale ambiguity, making it difficult to determine the real-world scale of the 3D information. Therefore, we employ a stereo camera mounted at the boom tip, slewing the boom to capture the surrounding environment. However, while existing related research (Sung and Kim, 2016) focuses on ≤5 m range, our work addresses a range of 5-30 m. At these greater distances, stereo cameras face a significant challenge: depth estimation errors increase substantially, making the accuracy insufficient for safe operational planning. To overcome this challenge, this study proposes a novel method that improves the existing stereo camera processing pipeline to enable highly accurate depth estimation, even at long distances.

2. Methodology

The proposed method first generates an initial depth map from stereo images using Semi-Global Block Matching (SGBM). Subsequently, the framework of PatchMatch Multi-View Stereo (PM-MVS) is applied as a post-processing step to refine this initial depth map. The core of the proposed method lies in the following three improvements made to the standard PM-MVS:

1. Initialization with SGBM: Instead of generating random initial values for optimization, the output from SGBM output—depth and derived surface normals—initializes plane hypotheses, accelerating convergence and stabilizing results. By starting with these informed plane hypotheses, the subsequent optimization process converges more rapidly and produces more stable results.
2. Priority-Based Propagation: This method prioritizes the dissemination of information from high-confidence regions to low-confidence regions based on the estimated error of the initial depth map. Specifically, all pixels are sorted into groups (bins) based on their initial matching cost, from lowest to highest error. The propagation process then proceeds sequentially, starting from the highest-confidence bin, ensuring that reliable depth information is spread first.
3. Adaptive Random Search: Based on the physical constraint that depth estimation error is proportional to the square of the distance to an object, the search range for both depth and surface normals is adaptively adjusted for each pixel—wider for distant views and narrower for near views—and is exponentially decayed over iterations to efficiently transition from a broad, global search to a fine-tuned, local optimization.

3. Results/Findings

To validate the effectiveness of our proposed method, we conducted a series of experiments using a dataset acquired from the simulator (Figure 1). This section presents the results from a single stereo image pair to demonstrate the method's performance. The evaluation was conducted from two perspectives: a qualitative assessment through visual comparison with the ground truth depth images and a quantitative assessment using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). First, we verified the improvement effect on the initial depth map generated by the SGBM method. Figure 2 qualitatively illustrates the progressive improvement of the depth map from a specific viewpoint. The initial depth map from SGBM (Figure 2b) shows significant estimation errors, particularly in the distant structures. In contrast, as our model's updates were applied one, two, and then three times (Figures 2c-e), the contours of the structures became sharper and the continuity of flat surfaces improved, showing the depth map gradually approaching the

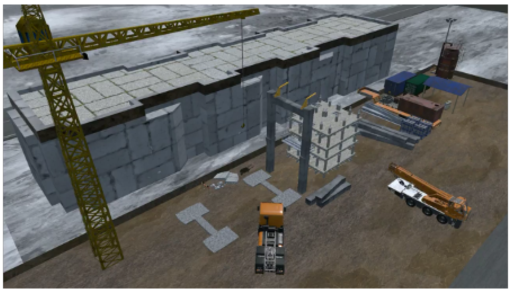

Figure 1: Top-down view of the simulated construction site.

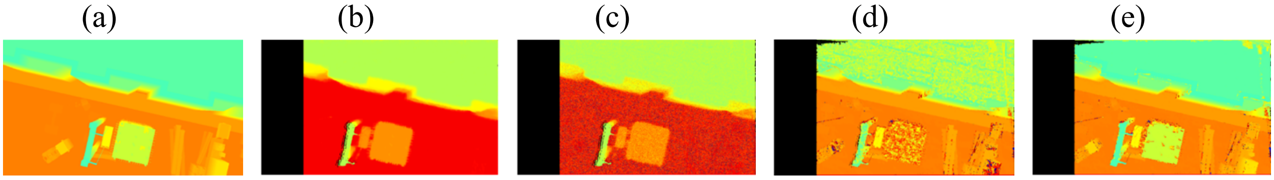

Figure 2: The process of improving the depth image with the proposed method. (a) The ground-truth depth image. (b) The initial depth map generated by SGBM. (c) through (e) show the results after one, two, and three applications of the proposed model update, respectively.

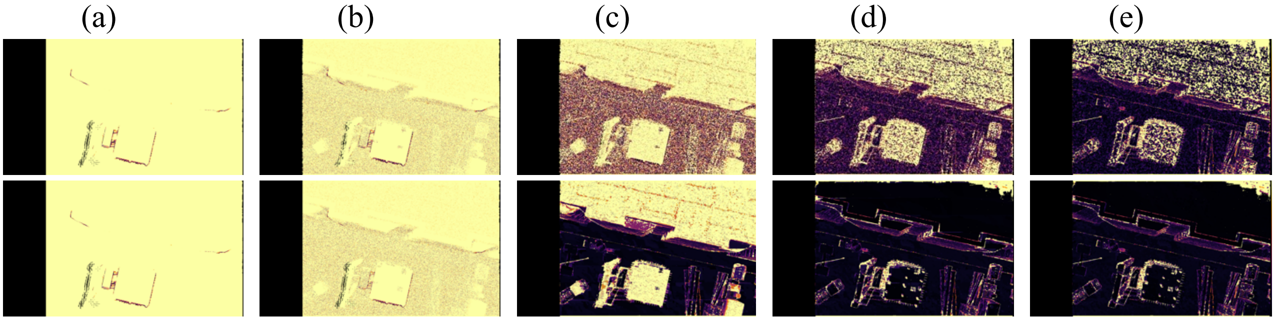

Figure 3: Comparison of error maps between the conventional and proposed propagation methods. (a) shows the initial error map, while (b)-(e) show the error maps after 1 to 4 model updates, respectively. The top row corresponds to the conventional method and the bottom row to the proposed method.

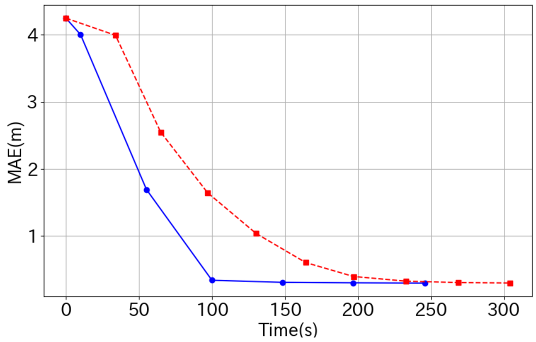

Figure 4: MAE value transition (blue: proposed propagation method, red: conventional propagation method).

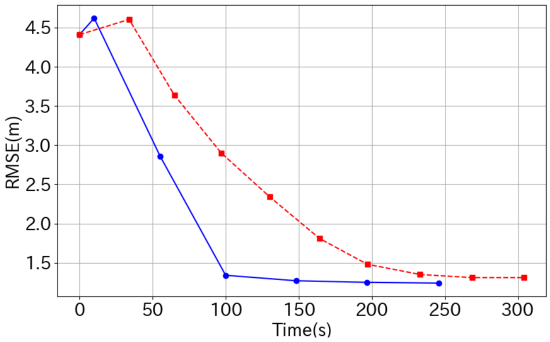

Figure 5: RMSE value transition (blue: proposed propagation method, red: conventional propagation method).

ground truth (Figure 2a).

Next, the effectiveness of the proposed reliability-based prioritized propagation was evaluated against the conventional checkerboard pattern propagation method. Figure 3 compares the error maps for both methods, showing the initial error and the results after one to four model updates. The top row corresponds to the conventional method, while the bottom row represents the proposed method. Visually, the top row shows that the conventional method struggles to reduce large areas of significant error (indicated by bright yellow regions), which persist even after several updates. In contrast, the bottom row clearly demonstrates that our proposed method substantially and rapidly reduces these errors, especially around the outlines of structures and in distant regions. This qualitative trend of faster error reduction is also confirmed by our quantitative evaluation.

Figures 4 and 5 show the change in MAE and RMSE values over time, respectively. For both metrics, the proposed method (blue solid line) shows the error converging much more rapidly than the conventional method (red dashed line). While the conventional method required approximately 240 seconds to reach a stable error level, our method achieved a comparable level of accuracy in about 100 seconds, demonstrating its efficiency.

4. Conclusion

This paper presented a depth refinement framework designed to improve long-range stereo reconstruction for crane-mounted cameras in construction environments. By leveraging a robust SGBM algorithm for initialization and introducing key improvements to the PM-MVS framework, our method achieves high-accuracy depth estimation. The primary contribution is a practical and effective solution that addresses the inherent limitations of stereo vision at long distances. The experimental results confirm that our approach outperforms standard algorithms, producing more accurate depth maps promising for safety-critical planning under the evaluated conditions. Future work will focus on validating the proposed method with real-world data captured from a crane on an active construction site and optimizing the implementation for real-time performance. Ultimately, the proposed method successfully improved the initial MAE from approximately 4.4 m to about 0.3 m. This demonstrates the significant impact of our approach in refining depth estimation for distant objects.

References

Kobayashi, T., Susaki, J., Shigemori, H., Yoneda, T., & Ososinski, M. (2023). High speed 3D-mapping around crane from video images of monocular camera. *Japanese Journal of JSCE*, *79*, Article 22. https://doi.org/10.2208/jscejj.22-22002

Sung, C., & Kim, P.Y. (2016). 3D terrain reconstruction of construction sites using a stereo camera. *Automation in Construction*, *64*, 65-77. https://doi.org/10.1016/j.autcon.2015.12.022