Improving Registration-Based High-Resolution Panorama Depth Estimation with Enhanced Fine Details

Yu Sheng Yang
3D Computer Graphics and Vision Lab
National Yang Ming Chiao Tung University
Tainan City, Taiwan
john001225.ai12@nycu.edu.tw

Chi Han Peng
3D Computer Graphics and Vision Lab
National Yang Ming Chiao Tung University
Tainan City, Taiwan
pengchihan@nycu.edu.tw

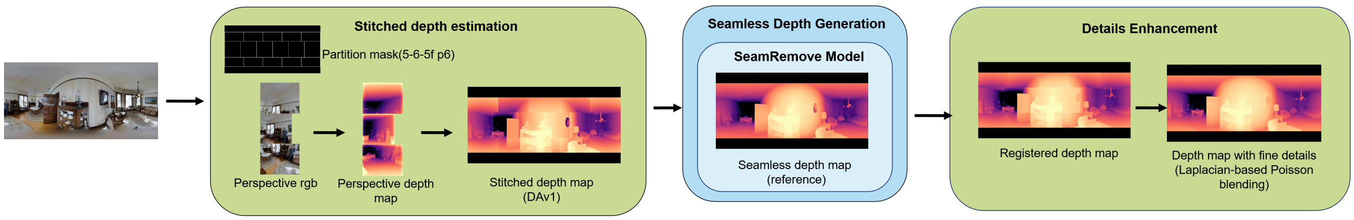

Figure 1: Our method consists of three stages: stitched depth estimation, seamless depth generation, and detail enhancement. First, we introduce a new partition strategy and apply DepthAnything V1 [Yang et al. 2024] for stitched depth estimation. Next, we train a model to refine the stitched depth. Finally, we adopt registration and Poisson blending, same as in prior work [Peng and Zhang 2023], to enhance image details.

Abstract

Current stitching- and registration-based panoramic depth estimation methods produce high-resolution depth maps (2048×1024) but remain limited in accuracy. Following [Peng and Zhang 2023], we adopt a three-step pipeline: improved stitched depth estimation, seamless depth generation, and detail enhancement (Fig 1). We introduce a new partition strategy and apply a SOTA perspective depth model to obtain stitched depth, refined by our SeamRemove model based on EfficientNet-B0 [Tan and Le 2019]. Finally, registration and Poisson blending further enhance fine details. Compared to [Cao and Wang 2024; Peng and Zhang 2023], our method achieves higher quantitative accuracy and better visual quality.

CCS Concepts

• **Computing methodologies → Computer vision**.

ACM Reference Format:
Yu Sheng Yang and Chi Han Peng. 2025. Improving Registration-Based High-Resolution Panorama Depth Estimation with Enhanced Fine Details. In *SIGGRAPH Asia 2025 Posters (SA Posters '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3757374.3771430

SA Posters '25, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2134-2/25/12
https://doi.org/10.1145/3757374.3771430

1 Introduction

Depth estimation from panoramas is challenging due to distortion and the limited field of view (≤180°), which make conventional monocular models ineffective. Panorama-specific methods adopt distortion-aware representations such as cube maps, tangent patches, or spherical transforms, but these often produce low resolution (typically 1024×512) and lack fine detail due to limited training data and distortion.

Stitching- and registration-based methods address these issues by projecting panoramas into perspective views, applying monocular depth models, and merging the results. Among them, the registration-based approach [Peng and Zhang 2023] achieves higher quality. Building on this, we propose SeamRemove, a model that refines stitched depth maps into seamless results. We further introduce an improved partition strategy and apply post-alignment to enhance detail (Fig. 1). Compared to prior methods, SeamRemove achieves better quantitative results without detail enhancement, and better visual quality with it.

More testing results can be found in the supplementary materials.

2 Method

2.1 Stitched Depth Estimation

We partition the panorama into three horizontal regions: upper, middle, and lower. Each region is divided into 3–6 perspective images, referred to as x-fold partitioning when the same number of images is used across regions. To handle spatial distortion near the poles, we reduce the number of images in the upper and lower regions (e.g., 5–6–5 fold) for better efficiency without loss of performance.

The number of partitions and their fields of view (FOVs) greatly influence depth estimation. Each image's FOV equals 360° divided by the number of folds (e.g., 60° for 6-fold). To study FOV effects, we

Table 1: Quantitative comparisons on Matterport3D at a resolution of 2048×1024. The best results are highlighted in red, the second-best in green, and the third-best in blue.

Method	RMSE	MAE	AbsRel	RMSE_{log}	$\delta 1$	$\delta 2$	$\delta 3$
CRF360D [Cao and Wang 2024]	0.4757	0.2611	0.0910	0.0582	91.62%	97.87%	99.39%
[Peng and Zhang 2023](HoHoNet(ref.))	0.4840	0.2691	0.1017	0.0669	89.94%	97.00%	98.91%
Our(SR model)(w/o detail enhancement)	0.4254	0.2208	0.0757	0.0511	93.51%	98.34%	99.60%
Our(SR model)(w/ detail enhancement)	0.4356	0.2277	0.0805	0.0554	92.83%	98.08%	99.46%

use fixed-fold setups with padding (e.g., 6-fold padding 6°), where padded regions are ignored during stitching.

After partitioning, the perspective images are fed into DepthAnything V1 [Yang et al. 2024] and stitched together. Inconsistent scale and shift among perspectives cause visible seams in the stitched depth.

2.2 Seamless Depth Generation

We train our SeamRemove model using stitched depth maps and corresponding RGB images, aiming to produce seamless and scale-consistent depth maps. The model adopts a U-Net architecture with EfficientNet-B0 [Tan and Le 2019] as the encoder.

The total loss combines pixel-wise L1 loss, gradient loss, and Laplacian loss to balance global accuracy and local detail:

- **L1 Loss:**

$$\mathcal{L}\text{L1} = \frac{1}{|M|} \sum i \in M \left| D_i - \hat{D}_i \right| \tag{1}$$

- **Gradient Loss:**

$$\mathcal{L}_{\text{grad}} = \frac{1}{|M_x|} \sum_{i \in M_x} \left| \nabla_x D_i - \nabla_x \hat{D}_i \right| + \frac{1}{|M_y|} \sum_{i \in M_y} \left| \nabla_y D_i - \nabla_y \hat{D}_i \right| \tag{2}$$

- **Laplacian Loss:**

$$\mathcal{L}_{\text{lap}} = \frac{1}{|M|} \sum_{i \in M} \left| \Delta D_i - \Delta \hat{D}_i \right| \tag{3}$$

The final objective is:

$$\mathcal{L}\text{total} = \mathcal{L}\text{L1} + \lambda_g \mathcal{L}\text{grad} + \lambda_l \mathcal{L}\text{lap} \tag{4}$$

with $\lambda_g = 0.5$, $\lambda_l = 0.2$.

Here, D and $\hat{D}$ denote ground truth and predicted depth; M is a mask for valid pixels ($D_i > 0$); ∇_x, ∇_y are gradient operators; and Δ is the Laplacian operator for high-frequency structures.

2.3 Details Enhancement

To enhance the details representation. We follow the previous registration-based method [Peng and Zhang 2023]. Registering the stitched depth to the seamless depth generated by our SeamRemove model, and perform the Laplacian-based Poisson blending.

3 Results

Quantitative results are presented in Table 1. The first one is a traditional method, the second one is the previous registration-based method. Compared to them, our methods yield the best and the second-best scores.

Figure 2 shows the results generated from SeamRemove model w/o (left) and w/ (right) details enhancement. The pendant light of the right result has clearer details representation than the left one.

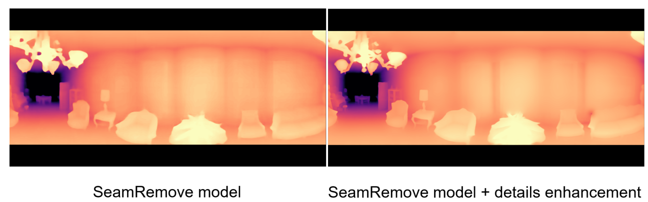

Figure 2: Qualitative results comparing SeamRemove without (left) and with (right) detail enhancement.

4 Conclusion

Our method consists of three stages: stitched depth estimation, seamless depth generation, and detail enhancement. In stitched depth estimation, we modify the partition strategy from 5-fold (72°) to 5–6–5 fold with 6° padding, and replace LeRes [Yin et al. 2021] with DepthAnything V1 [Yang et al. 2024] for better prediction. We train a U-Net with an EfficientNet-B0 [Tan and Le 2019] encoder, termed SeamRemove, to remove seams and ensure depth consistency. For detail enhancement, we follow [Peng and Zhang 2023] using registration and Laplacian-based Poisson blending. Compared to traditional panoramic methods (CRF360D [Cao and Wang 2024]) and the previous registration-based approach [Peng and Zhang 2023], SeamRemove achieves the best quantitative results and superior visual quality after enhancement.

Acknowledgments

This work is funded by the National Science and Technology Council of Taiwan (project number 111R10286C).

References

Zidong Cao and Lin Wang. 2024. CRF360D: Monocular 360 Depth Estimation via Spherical Fully-Connected CRFs. *arXiv preprint arXiv:2405.11564* (2024). https://arxiv.org/abs/2405.11564

Chi-Han Peng and Jiayao Zhang. 2023. High-Resolution Depth Estimation for 360° Panoramas through Perspective and Panoramic Depth Images Registration. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 3115–3124. doi:10.1109/WACV56688.2023.00313

Mingxing Tan and Quoc V. Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv preprint arXiv:1905.11946* (2019). https://arxiv.org/abs/1905.11946

Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10371–10381. doi:10.1109/CVPR52733.2024.00987

Wei Yin, Jianming Zhang, Oliver Wang, Simon Niklaus, Long Mai, Simon Chen, and Chunhua Shen. 2021. Learning to Recover 3D Scene Shape from a Single Image. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 204–213. doi:10.1109/CVPR46437.2021.00027