Interactive Relative Pose Estimation for 360° Indoor Panoramas through Wall-Wall Matching Selections

Bo-Sheng Chen
National Yang Ming Chiao Tung University
Tainan, Taiwan
laurence5181.ai10@nycu.edu.tw

Chi-Han Peng
National Yang Ming Chiao Tung University
Tainan, Taiwan
pengchihan@nycu.edu.tw

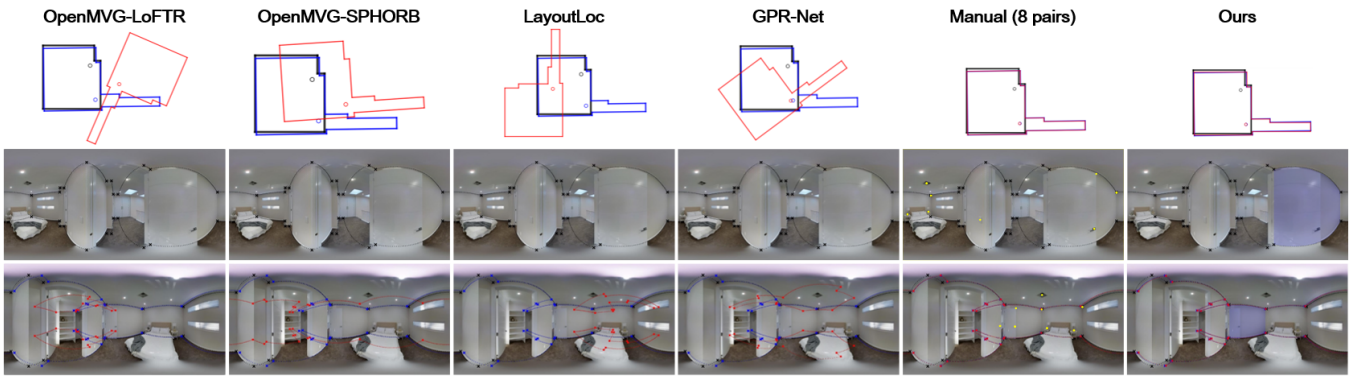

Figure 1: Left-to-right: Comparisons of camera pose estimations by OpenMVG with feature points collected by LoFTR and SPHORB, respectively, the optimization-based LayoutLoc algorithm proposed in ZInD, the state-of-the-art neural network-based method GPR-Net [Su et al. 2023], an interactive baseline that the user manually selected 8 feature point pairs, and our neural network-interaction hybrid approach. We show how the first (top) panoramas' estimated layouts (by HorizonNet [Sun et al. 2019]) and camera positions (small circles) are transformed to the spaces of the corresponding second panoramas (bottom) by the estimated camera poses (red) and ground truth poses (blue). We also show the manual selections as yellow dots and the wall-wall selection in our method in semi-transparent blue. In short, for this challenging case, only the interactive baseline and our approach were able to produce correct camera poses, while our method is significantly faster and easier / less error-prone.

ABSTRACT

We present an interactive approach to estimating the relative camera pose of two panoramas shot in the same indoor environment. Compared to the trivial interactive baseline, which would require the user to precisely select 8 or more pairs of matching points by mouse clicks, our method just needs the user to select a pair of matching walls with two mouse clicks or keyboard strokes. Our method is based on the key observation that, in most cases, there exist at least one or multiple pairs of roughly matched walls in the room layouts estimated by neural networks - which alone are sufficient to generate accurate relative camera poses. Tested on a real-world indoor panorama dataset, our method outperforms current state-of-the-art automatic methods by large margins, compensating the additional human efforts. Through user studies, we found that matched wall-wall pairs can be easily recognized and selected by humans in relatively short time, indicating that such an interactive approach is practical.

SA Posters '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0313-3/23/12.
https://doi.org/10.1145/3610542.3626114

CCS CONCEPTS

• **Computing methodologies** → **Matching**; • **Mathematics of computing** → **Solvers**.

KEYWORDS

Camera Pose Estimation, Panorama, Feature Point

ACM Reference Format:
Bo-Sheng Chen and Chi-Han Peng. 2023. Interactive Relative Pose Estimation for 360° Indoor Panoramas through Wall-Wall Matching Selections. In *SIGGRAPH Asia 2023 Posters (SA Posters '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3610542.3626114

1 INTRODUCTION

We address the challenging problem of estimating the relative camera poses between two sparse 360° panoramas captured in indoor environments, which is essential for efficient navigation and 3D reconstruction. Traditional methods based on feature point matching often fail in such scenarios due to the wide baseline, featureless regions, and visual ambiguity. Existing neural network-based methods are not robust to diverse and cluttered scenes, as they are only trained on a limited dataset of unfurnished homes. We propose a novel interactive approach that combines neural network predictions of room layouts with user inputs of matched walls. Compared to the trivial interactive baseline in which users are required to

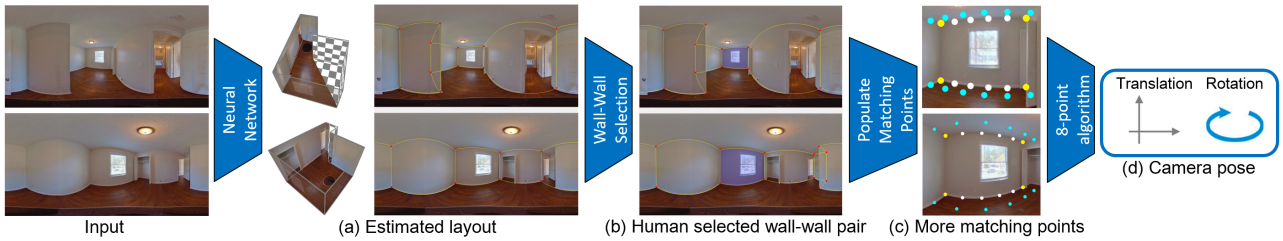

Figure 2: Overview of our pipeline. Given two panoramas shot in the same indoor environment, (a) we first use an existing method (such as HorizonNet [Sun et al. 2019]) to estimate a room layout for each of the panoramas. (b) We let the user to select a pair of matching walls. (c) A set of matched feature point pairs are generated based on the wall-wall pair using our novel approach. (d) Finally, an essential matrix / camera pose is calculated using the generated feature point pairs.

manually and precisely select 8 or more feature point pairs, our approach is much faster, easier, and less error prone, as we simply required users to select a pair of matching walls (can be done by simply two keystrokes). We demonstrate the state-of-the-art performance of our approach on two real-world datasets of indoor panoramas with ground truth camera poses: the Zillow Indoor Dataset and the Matterport 3D dataset. Our approach outperforms traditional by large margins and SOTA neural methods on their unseen (in-the-world) cases. Our contributions are:

- A novel interactive approach to estimate a camera pose between two wide-baseline indoor panoramas using neural network predictions of room layouts and user inputs.
- Compared to the trivial manual baseline, our approach is much faster, easier, and less error-prone.
- Our approach delivers state-of-the-art performances on two real-world indoor panoramas datasets, especially on the Matterport3D dataset with diverse and cluttered scenes.

2 METHOD

An overview of our method is shown in Figure 2. We first use a neural room layout estimation method (which takes a single 360° indoor panorama as input and output a 2D floor plan) to estimate walls for both panoramas. The user then simply selects a pair of matching walls from the two panoramas. Next, a dense set of feature point pairs is generated by our novel algorithm described as follows.

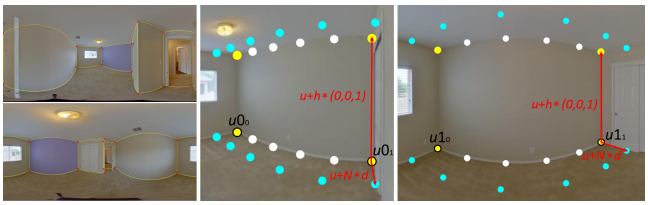

Figure 3: Feature point pairs generation process notations.

We denote the selected walls of the first and second panoramas as W_0 and W_1, respectively. The two "lower" (i.e., lying on the floor polygon) vertices of W_0 and W_1 are denoted as $u0_0$, $u0_1$, and $u1_0$, $u1_1$, respectively. Our algorithm then takes the following steps: First, we divide the "floor" edges from $u0_0$ to $u0_1$ and from $u1_0$ to $u1_1$ into s equal parts. We set s to 5 by default. Now, we have $s+1$ pairs of points that match on the floor edges. Second, for each point on the floor edge, we make a new feature point by *extruding along the wall normal* by a predefined distance d. The choice of d has no effect on the results and we set d to be $\frac{1}{5}$ of the floor edge's length for easy visualization. Now, we have $u + N * d$, N is the wall normal. Third, for every feature point we have obtained (they all lie on the floor polygon) on both room layouts, we induce a "ceiling" counterpart by extruding along the negative gravity direction (+z axis) by the height of the first panorama's estimated layout. In the end, we have generated $4(s+1)$ pairs of feature points for the subsequent camera pose calculation.

Finally, we use the matched feature point pairs to calculate an essential matrix that encodes the relative position and orientation of the two panoramas (i.e., a relative camera pose).

3 RESULTS AND CONCLUSION

We evaluated our method on two popular datasets of indoor panoramas (Zillow Indoor Dataset and Matterport3D dataset). Our method achieved significantly better accuracy than traditional 8-point algorithm methods (using feature point detection algorithms designed for planar domains (SIFT and LoFTR) and spherical domains (SPHORB)), and outperformed neural SOTA methods: CoVisPose [Hutchcroft et al. 2022] and GPR-Net [Su et al. 2023], which were trained on the ZInD dataset, in the unseen Matterport3D dataset. We also conducted a user study of 50 participants to use our interactive system to conduct wall-wall matching selections. Overall, we observed that the users could select correct wall-wall matching fairly accurately and quickly. Detailed results are in the supplementary materials. The results show that our interactive take to the challenging wide baseline pose estimation problem have advantages in speed (faster and less error-prone than the manual baseline approach) and accuracy (outperformed neural SOTA methods in in-the-wild cases).

ACKNOWLEDGMENTS

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

REFERENCES

Will Hutchcroft, Yuguang Li, Ivaylo Boyadzhiev, Zhiqiang Wan, Haiyan Wang, and Sing Bing Kang. 2022. CoVisPose: Co-Visibility Pose Transformer For Wide-Baseline Relative Pose Estimation In 360° Indoor Panoramas. In *ECCV*.

Jheng-Wei Su, Chi-Han Peng, Peter Wonka, and Hung-Kuo Chu. 2023. GPR-Net: Multi-View Layout Estimation via a Geometry-Aware Panorama Registration Network. In *CVPR*.

Cheng Sun, Chi-Wei Hsiao, Min Sun, and Hwann-Tzong Chen. 2019. HorizonNet: Learning Room Layout With 1D Representation and Pano Stretch Data Augmentation. In *CVPR*.