Abstract:

- Image-to-Point Cloud registration between 2D images and 3D LiDAR point clouds is a significant task in computer vision. The traditional registration pipeline first establishes correspondences between images and point clouds and then performs pose estimation based on the generated matches. However, 2D-3D correspondences are inherently difficult to be established due to the large modality gap between images and LiDAR point clouds. To this end, we build a bridge to alleviate the 2D-3D modality gap, which aligns LiDAR point clouds to the virtual points generated by images. In this way, the modality gap can be alleviated to the domain gap of different types of point clouds. Concretely, our framework conducts feature fusion from the LiDAR and virtual point cloud by utilizing the Transformer layer. To relieve the domain gap, a frustum points retrieval module and a combined correspondences retrieval module are proposed based on the consistency of the feature and position descriptor to select the correct correspondences among the candidates, which are generated from the simultaneous retrieval of features and position descriptors. In the implementation procedure, we design a frustum retrieval loss and a combined correspondence retrieval loss for cross-modality correspondence retrieval. Experimental results and comparison with state-of-the-art Image-to-Point Cloud methods on KITTI and nuScenes datasets demonstrate our proposed method has achieved superior performance.

Contribution:

- We propose to utilize the virtual point cloud as bridge for Image-to-Point Cloud registration, which alleviates the 2D/3D modality gap to domain gap between LiDAR point clouds and virtual point clouds generated from the 2D images.
- We propose a frustum points and a combined correspondence retrieval module to relieve the domain gap between the LiDAR and virtual point cloud by introducing frustum retrieval loss and combined correspondence retrieval based on the consistency of feature and position descriptor.
- Experiments on the KITTI Odometry and nuScenes datasets demonstrate the effectiveness of our proposed method, especially when confronting complicated scenarios with 2Dotation and transformation.

Experiments:

- Quantitative Comparison. Analysis of the registration recall of different Image-to-Point Cloud registration methods with various RTE and RRE thresholds on KITTI and nuScenes datasets.
- Qualitative Comparison. For visual comparison, the point cloud is projected into images with the extrinsic parameters predicted by compared methods on KITTI and nuScenes datasets.

Method:

- Virtual Point Cloud Generation. To perform accurate virtual point cloud generation, we can resort to supervised monocular depth estimation methods to obtain depth maps. Then, virtual points can be converted from depth maps using camera intrinsic parameters.
- Frustum Points Retrieve. As for $N \times C$ point-wise LiDAR point cloud $L \times C$ fusion feature, we utilize the pair-wise hue-distance embedding to construct the $N \times L$ distance matrix $D$ for retrieve as:

$$d_{ij} = F_i \left( \frac{|f_i - f_j|}{\sigma_{ij}} \right) \times W^{HD}, i \in N, j \in L.$$