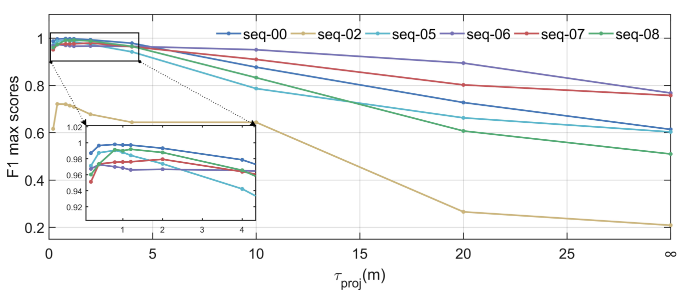
APPENDIX

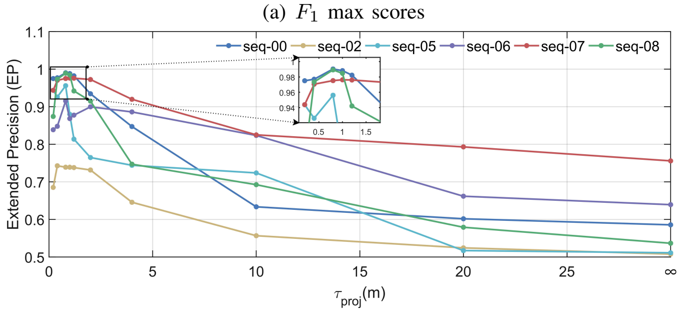
In this appendix, we investigate the effects of τ_{edge}, θ, and τ_{proj} on both place recognition and pose estimation. For each parameter, we change its value while keeping the others static, and then we run our method for all the sequences. To reduce the interference from semantic segmentation result, all the tuning experiments are performed based on the ground-truth semantic labels.

EFFECT OF τ_{proj}

Differently, τ_{proj} only affects place recognition because it is only used for global descriptor extraction. As shown in Fig. 7, both F_1 max scores and EP generally decrease when τ_{proj} becomes larger. This is because the vertex matching merely based on vertex descriptor is not perfect, there are still some incorrect matching result. The project operation can be considered as a geometry verification, which can effectively filter the incorrect matches. However, when such verification is too strict, i.e., a small value of τ_{proj}, both F_1 max scores and EP decrease. The reason could be the error from pose estimation T^* and the geometric centroid coordinate of the object, which will lead to the position inconsistency between the same object in two different frame after projection.



(a) F_1 max scores

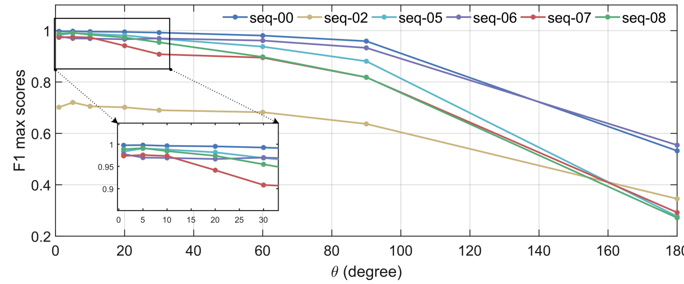


(b) Extended Precision (EP)

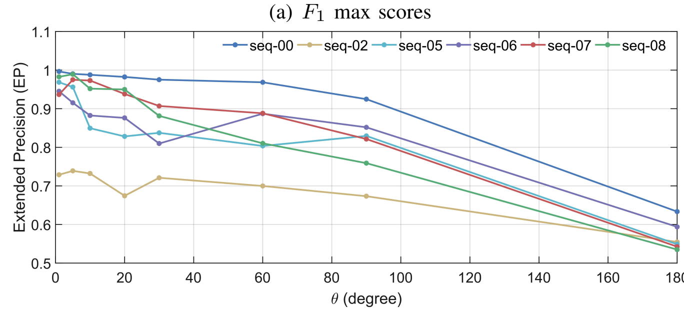
Fig. 7: F_1 max score and Extended Precision corresponding to different τ. ∞ represent that all of the vertex matching results from part B of section III are used.

EFFECT OF θ

In our method, θ directly affects the dimension of the vertex descriptor and the global descriptor. Fig. 8 and 9 show the F_1 max score, EP, average RTE, and average RRE corresponding to different θ. As θ increases, the performance of place recognition and pose estimation gradually decrease among most of the sequences. Obviously, when θ is larger, the dimension of the vertex and global descriptor become smaller. As a consequence, triplets with the same semantic combination while with different α contribute equally for the descriptor construction.

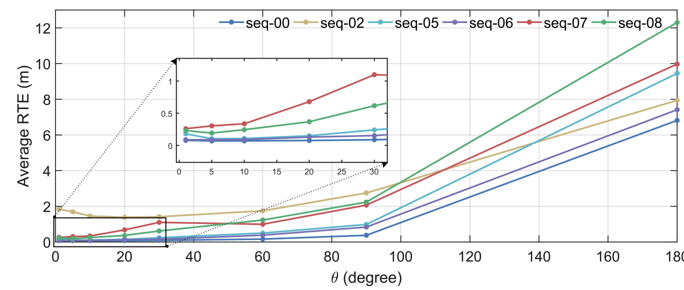


(a) F_1 max scores

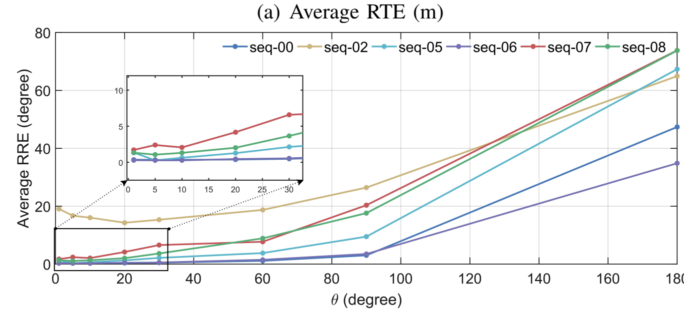


(b) Extended Precision (EP)

Fig. 8: F_1 max score and Extended Precision corresponding to different θ.



(a) Average RTE (m)



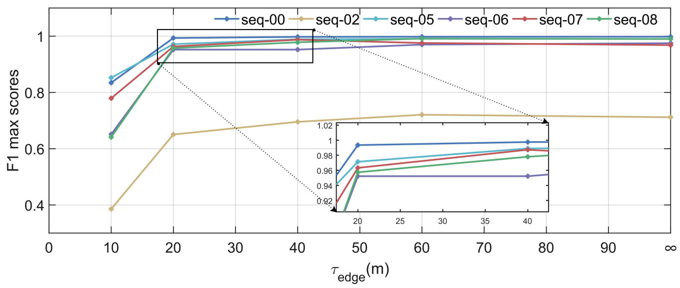
(b) Average RRE (degree)

Fig. 9: Average RTE, and average RRE corresponding to different θ.

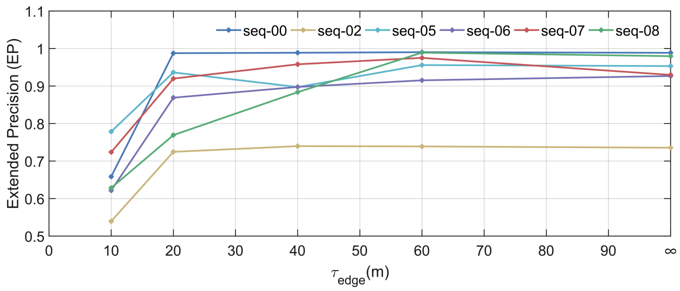
The descriptive power for both individual vertex and global descriptors degrades, leading to more incorrect vertex matches. Therefore, the pose estimation and place recognition become worse. Meanwhile, we also found that when θ is too small (i.e., $1°$), the general performance also degrades. The main reason is that we can not obtain the consistent geometric centriod of the same object from two different LiDAR point cloud. Viewpoint change and local obstruction can easily lead to such inconsistency.

EFFECT OF τ_{dege}

Different values of τ_{dege} will result in different constructed graphs. As τ_{dege} increases, the number of edges connecting

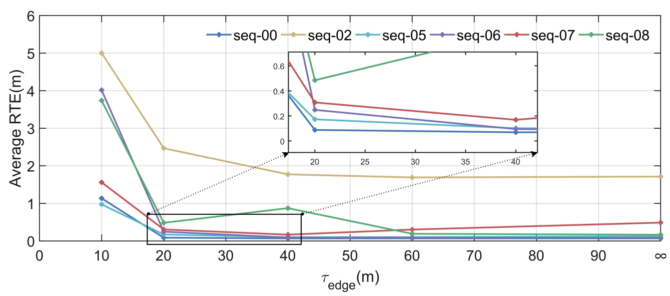


(a) F_1 max scores

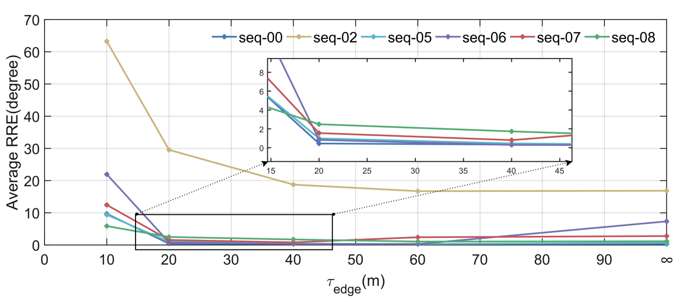


(b) Extended Precision (EP)

Fig. 10: F_1 max score and Extended Precision corresponding to different τ_{edge}.



(a) Average RTE (m)



(b) Average RRE (degree)

Fig. 11: Average RTE and average RRE corresponding to different τ_{edge}.

vertices increases, which means the graph becomes denser. And the overall performance becomes better, as shown in Fig. 10 and 11. Theoretically, a small τ_{dege} leads to less triplet for the same vertex, which means the vertex descriptor focuses on encoding local information. In addition, the same object in a certain frame might not be observed in another neighboring frame, which can lead to a huge difference between the local information in different frames. A denser graph can encode both local and global information for each vertex, which can greatly relieve the inconsistency between vertex descriptors caused by missing some common objects. However, we also found that the overall performance will slightly decrease when τ_{dege} is larger than a certain value (i.e., resulting in a fully-connected graph), as shown as the curves for Seq-02, 07, and 08 in Fig. 10 and 11. We conjecture the reason could be that the 6-DoF relative pose between the query and matched frames is larger, making a huger difference between the cluttered objects, especially those that are far away from the center. Empirically, a value of τ_{dege} that can connect most of the vertices except those that are too far away contributes to a better overall performance.