[CV21] Assignment 4 - Di Zhuang

Some Comments:

According to moodle discussion there is a small bug in the framework:
https://moodle-app2.let.ethz.ch/mod/forum/discuss.php?d=93017
This might cause the tensorboard figures of the test part to look a little bit weird.

In addition, I used **GPU** and **pytorch 1.7.1** for this assignment, but I failed at loading the model on my CPU only machine and I have no idea why..

2.1.3 RANSAC results

Estimated coefficients (true, linear regression, RANSAC):
(1, 10), (0.6159656578755456, 8.96172714144364), (0.9954905251271884, 10.01010661207374)

2.2.2 Differentiable warping

1. p_ij = K_i · (R_0i · (inverse(K_0) · p · d_j) + t_0i)

3.3 Training

See tensorboard.png in the zip file.

3.4 Test

1. Explain what geometric consistency filtering is doing in the report.

   Reproject_with_depth:

   First, given depth_ref, it projects the pixels in the reference image to the pixels in the source image.

   Next, based on the result of the projection in the first step, it remaps depth_src.

   Given the remapped depth_src, namely sampled_depth_src, it projects the pixels in the source image back to the reference image.

The reprojected depths, x’s and y’s are returned.

Check_geometric_consistency:

Compute the **distance** between the reprojected xy’s and the original estimated xy’s.

Compute the absolute **difference** between the reprojected depths and the original estimated depths.

Filter the estimations with too large distances/differences.

Essentially, what geometric consistency does is to check whether the estimation varies too much after reprojection. If yes, then the estimation is geometrically unreliable, and should be filtered out.

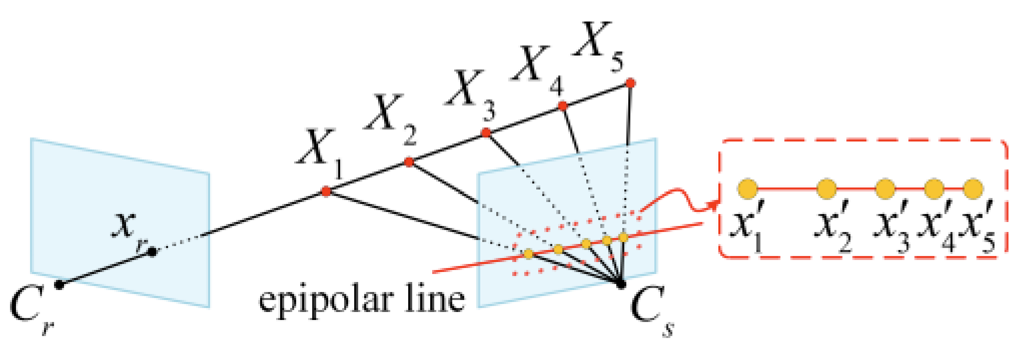
2. Test Visualize

See scan1.png and scan9.png in the zip file.

3.5 Questions

1. Inverse depth range

I think the inverse depth range is more suitable for large-scale scenes. Although the sampled depth hypotheses are uniformly distributed in depth space, their projected 2D points in a source image are not distributed uniformly along the epipolar line [1], as shown in the image below.



(image credit to https://arxiv.org/pdf/1912.11746.pdf)

[1] "Learning Inverse Depth Regression for Multi-View Stereo with ...." 26 Dec. 2019, https://arxiv.org/abs/1912.11746. Accessed 3 Dec. 2021.

If the true depth hypothesis is close to the camera center, then it might not be captured since the projected 2D points are sparse there; if the true depth hypothesis is far from the camera center, then it might be captured by multiple deep features in the source image, which is confusing, since the projected 2D points are dense there. As large-scale scenes tend to have large size and variability, this effect is especially obvious in large-scale scenes. Sampling in the inverse depth range makes our model more robust, detailed method see https://arxiv.org/pdf/1912.11746.pdf page 2.

2. Integrating matching similarity

No, I don't think it is robust to some challenging situations such as occlusions. Taking average does not consider the variance between the matching similarity from different source views, while this variance is usually important in challenging situations such as occlusions. For example, in case of occulations, some of the source views could have very bad similarity results, which gives an unreliable average value.