Progress Update: Motion Segmentation

16/02/2022

Introduction

This week we continued to think about self-supervised architectures and corresponding loss function.

We expanded our Carla data collection script to now also collect optical flow using the optical flow sensor new in Carla 0.9.13.

Furthermore, we plotted the static scene flow using our collected ground truth data of depths and poses.

Optical Flow Plotting/Verification

Using the transformation ground truths we collected in CARLA we wrote a python script to plot the 2D flow fields.

For each frame we collected one $\mathcal{SE}(3)$ matrix, representing the transformations from the previous frame at $t-1$ to the current frame at t.

To find the flow vector for a single pixel we do the following:

- take pixel coordinate $\mathbf{u} = [u, v]$ of image at $t-1$ and reproject them into 3D using camera intrinsics K and depth ground truth.
- The 3D point is then transformed from position at $t-1$ to t using the transformation matrix from the ground truth
- The transformed 3D point is then projected into the image plane at t and pixel coordinates $\mathbf{u'} = [u', v']$ are obtained
- The flow vector is now $\mathbf{u'} - \mathbf{u}$

We then performed the above for our own recorded ground truth. Figure1 shows the resulting plot. Our flow vectors are ~ 0 and all point in the same direction and thus are not what was expected.

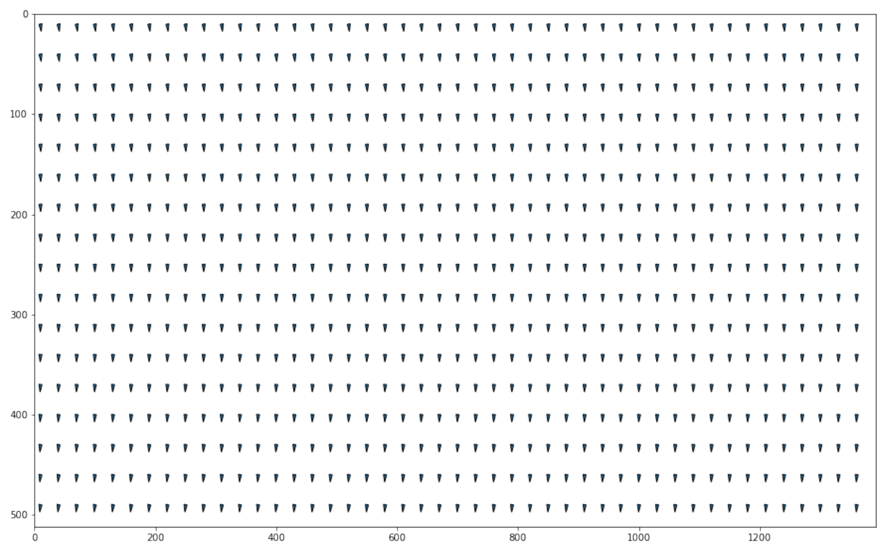


Figure 1: Flow Based on Own Ground Truth

In order to eliminate problems with our recorded transformation data, it was decided to use Week 4's dataset to test our script on. This is visualised in 2. The general structure of the flow looks correct but the magnitude of the vectors are > 2000 pixels.

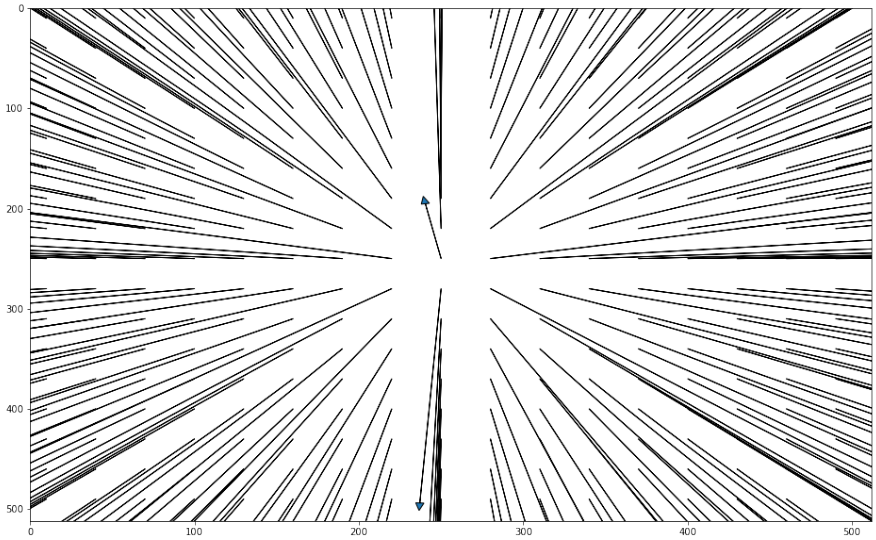


Figure 2: Flow Based on Week4 Ground Truth

The ground truth from Week 4 uses the position of the car in the world frame as the ground truth for the poses. Our ground truth uses the relative transformation from the previous frame to the current frame as the ground truth. Do we need to perform some other transformation on the reprojected 3D point before applying our ground truth transformation. What frame are the reprojected image points in?

Unsupervised Loss

We adapted the loss from [1] to now additionally include an error term which minimizes the geometric error between two images, see Equation 1.

$$E = \lambda_R E_R + \lambda_F E_F + \lambda_G E_G + \lambda_M E_M + \lambda_C E_C + \lambda_S E_S \tag{1}$$

where E_G is the geometric loss as

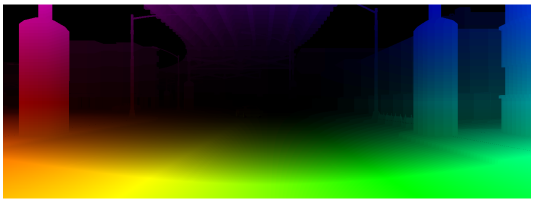
$$E_G = \sum_{s \in \{+,-\}} \sum_{\Omega} \rho(p_1, p_2) \tag{2}$$

We need to ensure that the points from image I_i that are reprojected, transformed and projected into image I_{i+1} are on the image plane. If they are not in the image bounds, these points will be disregarded and no loss will be accounted for. Similarly for points present in image I_{i+1} but not in image I_i, the geometric loss cannot be computed.

CARLA Optical Flow Sensor

We added the optical flow sensor to our data collection script. Visualizing the optical flow sensor results were straight forward, however obtaining the actual pixel motion v_x and v_y wasn't. Even though CARLA's docs specified the sensor data to be encoded in 64 bit bytes, using numpy bytes to array conversion returned zeros. We found out that in contrary to CARLA's docs, using a 32-bit conversion gave us the wanted optical flow vectors.

(a) RGB Image



(b) Optical flow

Figure 3: Optical flow sensor addition

Questions/Discussion Points for Self-Supervised Approach

- General questions on transformation regarding sensor data, time steps
- General discussion into how to proceed/finish up the project

References

[1] A. Ranjan, V. Jampani, L. Balles, K. Kim, D. Sun, J. Wulff, and M. J. Black, "Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019.* Computer Vision Foundation / IEEE, 2019, pp. 12 240–12 249. [Online]. Available: http://openaccess.thecvf.com/content_CVPR_2019/html/Ranjan_Competitive_Collaboration_Joint_Unsupervised_Learning_of_Depth_Camera_Motion_Optical_CVPR_2019_paper.html