CIS 5810, Computer Vision and Computational Photography, Fall 2022
Extra Credit Project

Ayush Goel, Akash Sundar

In this project, we intend to calculate the correspondences between two consecutive frames using the KLT tracking procedure. There is a displacement of pixels of the moving objects between frames.

We begin by creating bounding boxes across objects we want to track. We then proceed to finding features within this bounding box. Through the program, we are tracking the positions of these features as the scene moves forward.

We can think of the next frame to be shifted by some small distance δ with respect to the previous frame. Therefore, using a feature coordinate in a frame, we calculate flow that gives us the distance moved between the frame δ. This distance gives us an idea of how much each point needs to be translated to get the previous frame. This is done multiple times to get closer to the initial pixel.

3.1 Bounding box and features in the first frame

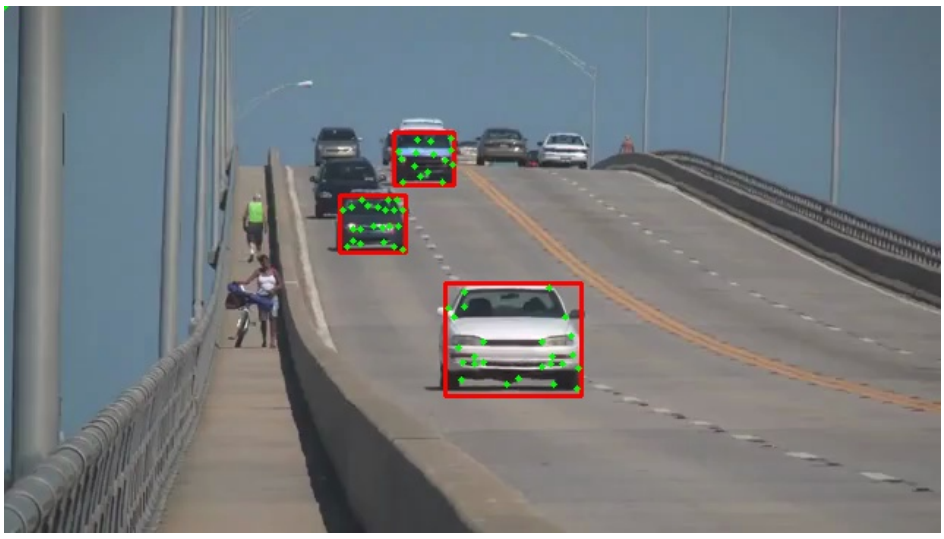

Figure 1: Bounding box and features in the first frame

3.2 Features in two frames with noticeable translation of tracked object

Figure 2: Features in Frame 11

Figure 3: Features in Frame 101

Figure 4: Features in Frame 201

3.3 Bounding boxes in two frames with noticeable translation of tracked object

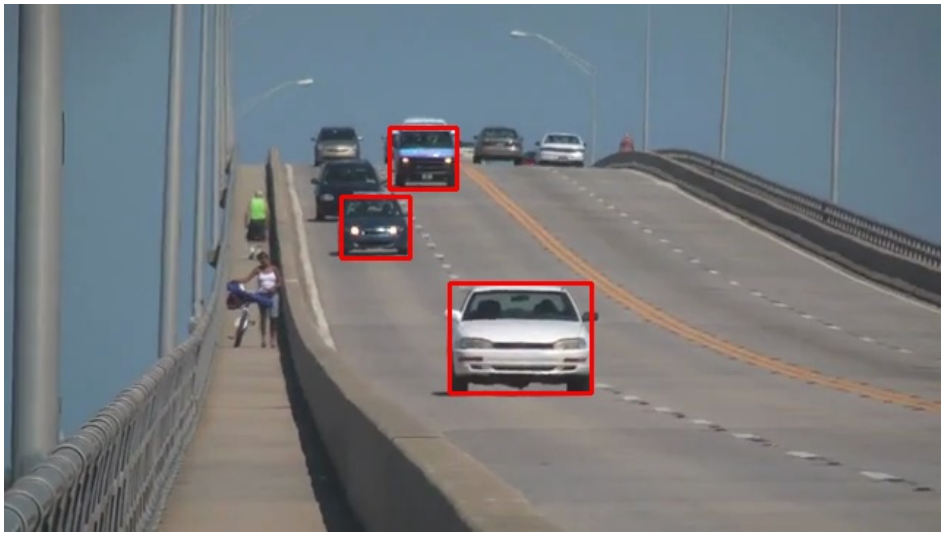

Figure 5: Features in Frame 11

Figure 6: Features in Frame 101

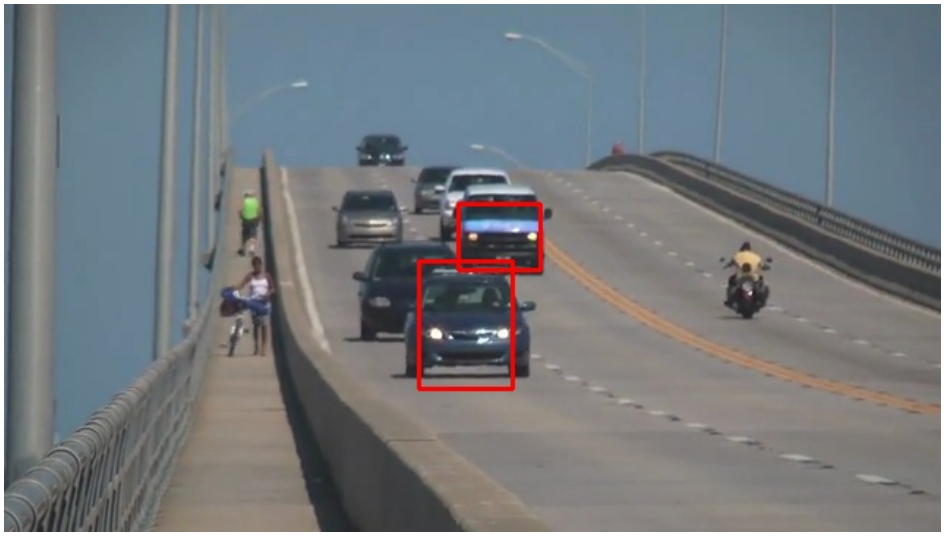

Figure 7: Features in Frame 201

3.4 Show the case when there are not enough feature points in one frame and you generate enough feature points in the next frame. List a few reasons why features are lost and discuss your strategy to detect and handle this situation

Figure 8: Frame with less features

As can be seen in the above image, the features detected on the white car are less. Sometimes, there might some features that might generate higher displacements than usual. Since optical flow assumes small displacements between frames, we should discard such features. Also, the pixels outside the bounding box must also be discarded since they are not the perfect features, which might later lead to noise.

We can tackle this by computing the displacement between the features of consecutive frames and discarding the ones that are above a certain threshold.