CSC490 Module 2 Exploration: Improved Object Tracking

Presented by Koby Choy and Lily Yu
Group 10769 (choykwan, yuxiangn)

Motivation

We want to increase the accuracy of the tracking algorithm by implementing a more sophisticated loss function.

What problems are we encountering?

- The MOTA that we obtained in Part 1 is not that good, it is around 0.3, so we hope to improve it
- IOU might not perform very well when the two vehicles are very close to each other (particularly when the detections are noisy)
- IOU can’t handle cars that are occluded
    - Everytime a previously occluded vehicle reappears, it will be assigned to a new tracklet

Describing Methods

Part I: Implementing Geometric Loss (Proposed by Shi et al.)

- **Keypoint** $\hat{\mathbf{t}} = (\frac{\hat{p}_1 - x_1}{x_2 - x_1}, \frac{\hat{p}_2 - y_1}{y_2 - y_1})$.

  T_hat here refers to the keypoints of the bounding box 1, where (p1, p2) is the centroid and (x1, y1), (x2, y2) refers to the top left and bottom right coordinates of the box.
  - Measure the L1 loss for the bounding boxes' keypoints.
    Lkp=L1(t, t_hat)

- **Yaw**
  We obtain the L1 loss between the yaw of the first bounding box and that of the second bounding box.

- **Size**
  - We obtain the size of the bounding boxes by multiplying the length by the width. Measure the L1 loss between the sizes of the two bounding boxes.

Total Loss

$$L = \lambda_{kp} L_{kp} + \lambda_{size} L_{size} + \lambda_{yaw} L_{yaw}$$

The values of lambda that we tested were:

Group 1 Settings:
lam_t = 5
lam_size = 3
lam_yaw = 5

Group 2 Settings:
lam_t = 5
lam_size = 1
lam_yaw = 3

Describing Methods

Part II: Implementing Motion Feature Loss

- Let B_1 be the bounding box in the previous frame and B_2 is the bounding box in the current frame. The cost algorithm will follow the steps below.

    1. Find the associated tracklet of the actor detected in bbox B_1 and obtain the tracklet's second last item for B_0. Calculate the displacement D1 between bounding boxes B_1 and B_0

    2. Calculate the displacement D2 between bbox B_1 and bbox B_2

    3. Calculate np.linalg.norm(D1 - D2) as the motion cost

Experiments and Results

Experiments	**MOTA**	**MOTP**	**Mostly Tracked**	**Mostly Lost**
IoU Cost (baseline)	0.3243	0.6178	0.5865	0.3926
Geometry Cost	0.0871	0.6161	0.6074	0.0958
Motion Cost	0.3035	0.6159	0.5497	0.2376
Motion + IoU Cost	**0.3247**	**0.6176**	**0.5456**	**0.2739**

Results

The combined method achieved better results than what we have originally implemented in part 1.

Combining Motion Features Cost with IoU Cost

- The Mostly Lost tracklets for the combined approach is 0.2739, which is better than only IoU cost approach of 0.3926.
- This decrease in Mostly Lost suggests that the tracklets are less fragmented.

Bonus exploration: Occlusion handling for tracking

Motivation: Fragmentation of tracklets

Method: post-processing of tracklets

for each tracklet1 in the tracklets:
if tracklet1's tail bbox overlaps significantly with another tracklet2's head bbox (IoU thresholding):
connect the 2 tracklets: append tracklet2 to tracklet1 and delete tracklet2

Results:

- 0.002 increase in MOTA
- each actor has one colored tracklet in the visualization

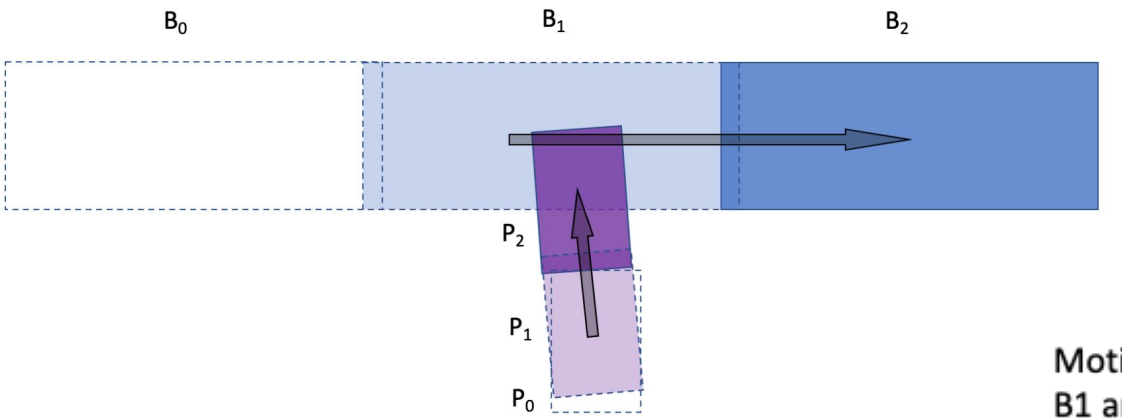
Limitations: motion assumption

Q & A

Appendix

Example Scenario

- Blue vehicle B moves at constant velocity (right of way)
- Purple vehicle P waits and then travels



IoU only:
B1 and P2 would be matched
⇒ bad!

Geometry only:
B1 and B2 would be matched because they have the same size, yaw and normalized keypoints so 0 geometry cost
⇒ good!

Motion only:
B1 and B2 would be matched because displacement(B0, B1) = displacement(B1, B2) so gives 0 motion cost
⇒ good!

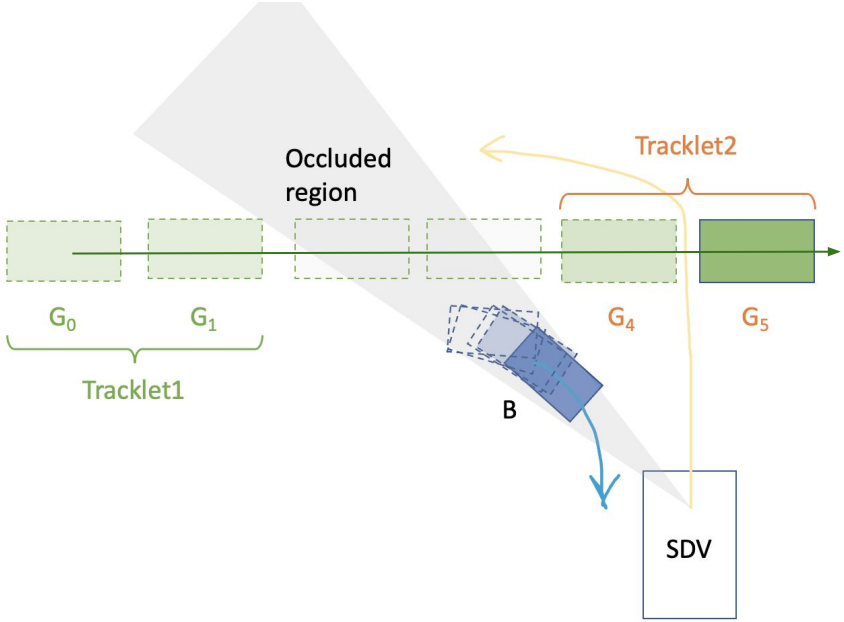
Appendix

Example Occlusion Scenario

- Our SDV stops to observe traffic (wants to turn left)
- Green vehicle G goes straight, occluded in some frames
- Blue vehicle B slows down to turn right

The tracker originally would mistakenly assign bounding boxes G4 and G5 to a new actor forming Tracklet2.

Use better motion assumption for the post-processing method to connect Tracklet1 and Tracklet2 together for the Green vehicle.



Appendix: example sequence ground truth