Computer Vision based Near Miss Detection among Mixed Traffic Flows within Intersections

Jiahao Xia, Yaocheng Wu, Jie Gong

Department of Civil and Environmental Engineering, Rutgers, the State University of New Jersey.

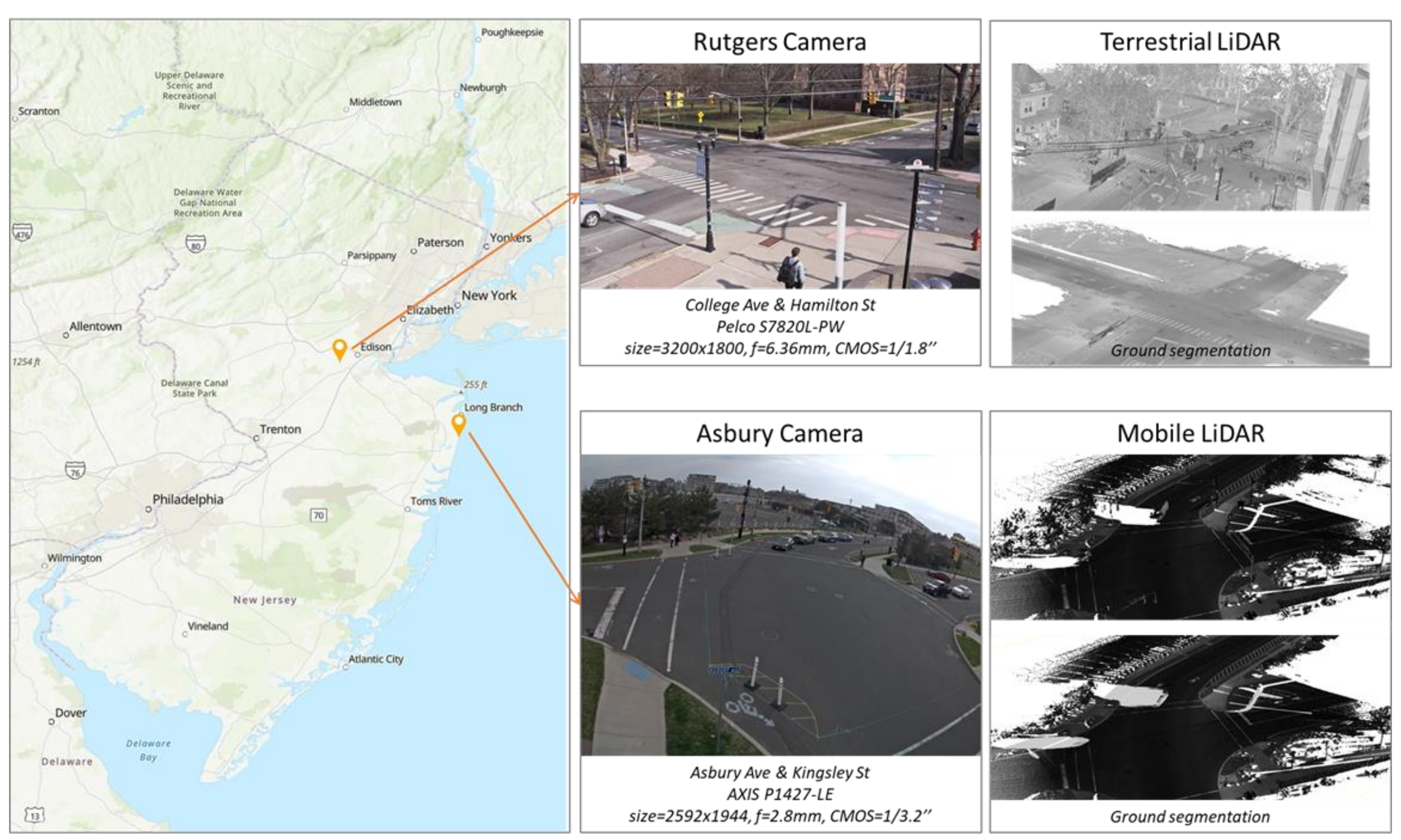
INTRODUCTION

Intersection safety stands as a paramount concern within the realm of transportation and urban planning. Statistical data underscores the significance of this issue, revealing that 20% of fatal traffic accidents and 50% of serious collisions transpire at road intersections. The quest for improving intersection safety is an ongoing challenge that has garnered increasing attention from traffic engineers, urban planners, and lawmakers. Addressing intersection safety encompasses a variety of measures, including intersection design and the implementation of effective traffic signals and signs. These measures all hinge on a meticulous analysis of traffic dynamics at intersections. To this end, traffic surveillance cameras play a pivotal role in continuously collecting valuable data for safety-related studies.

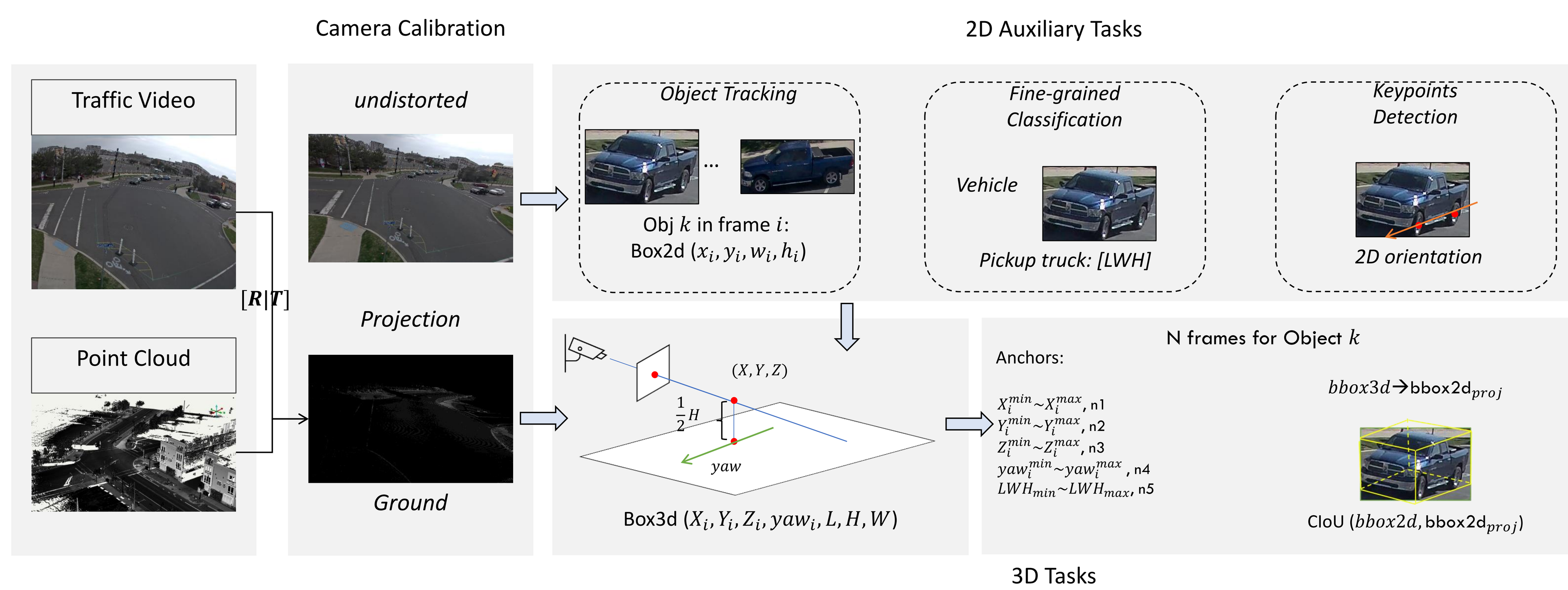
Nevertheless, a significant challenge persists in estimating the precise positions and orientations of vehicles in the real world based on traffic video footage. The task is intricate, primarily due to the inherent ambiguities of reasoning 3D information from monocular 2D images. Most contemporary vision-based traffic studies tend to focus on straight road scenarios, often overlooking the complexities presented by turning conditions. Deep learning-based 3D information estimation heavily relies on expensive 3D ground truth annotations or the integration of costly sensors capable of providing 3D ranging observations.

In this study, we propose a computer vision-based framework designed to estimate the 3D positions of vehicles and pedestrians from traffic videos. Notably, our approach does not necessitate the availability of ground truth 3D bounding boxes for training purposes. The derived 3D information enables the calculation of traffic flow velocities and directions, thus empowering us to identify potential near-miss incidents among road users

DATASET



METHODLOGY

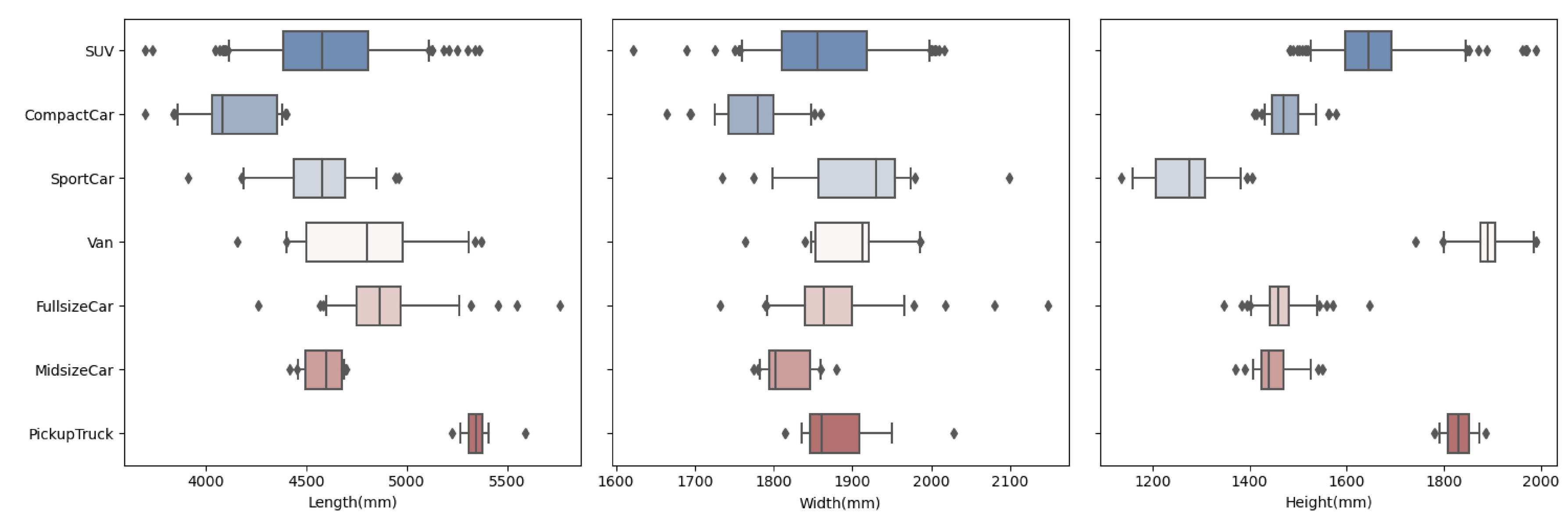


- Camera calibration---Radial distortion model

$$\hat{x}\ \hat{x} = x(1 + k1r^2 + k2r^4 + k3r^6)$$
$$\hat{y}\ \hat{y} = y(1 + k1r^2 + k2r^4 + k3r^6)$$
$$r = \sqrt{x^2 + y^2}$$

- Object Tracking: YOLOv5+StrongSort

- Fine-grained vehicle classification: CLIP ViT-bigG/14 - LAION-2B



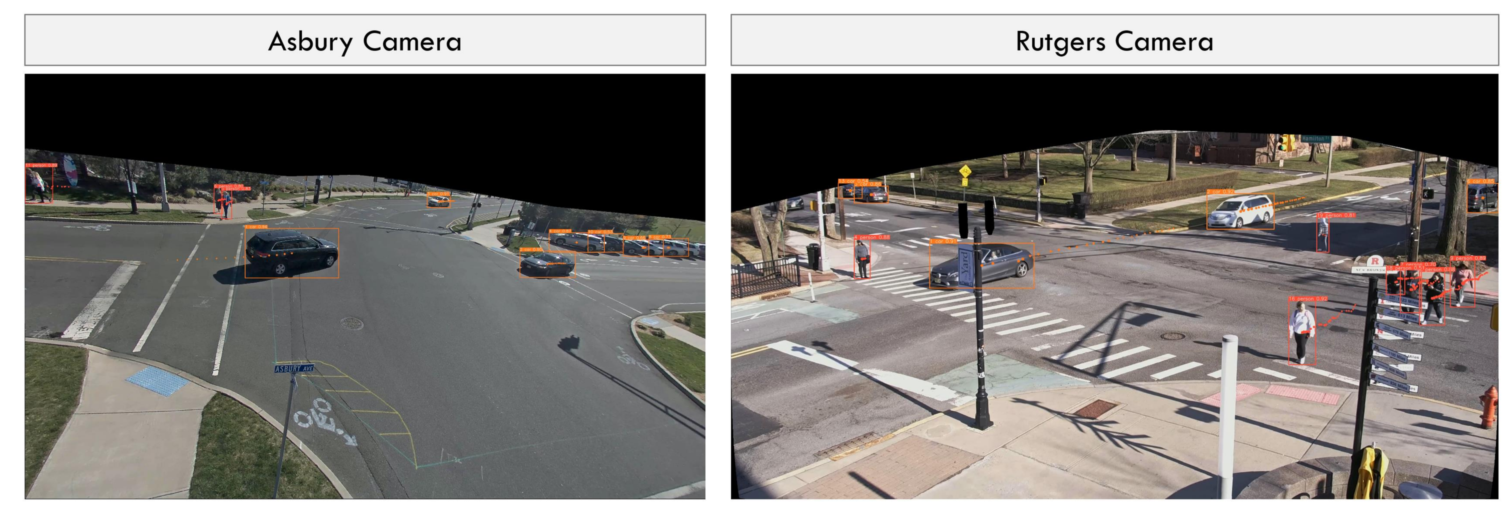
- Vehicle keypoint detection: HRNet

Index	Name	Index	Name
0	front_up_right	12	rear_light_left
1	front_up_left	13	rear_light_right
2	front_light_right	14	rear_low_left
3	front_light_left	15	rear_low_right
4	front_low_right	16	central_up_right
5	front_low_left	17	rear_corner_right
6	central_up_left	18	rear_wheel_right
7	front_wheel_left	19	front_wheel_right
8	rear_wheel_left	20	rear_plate_left
9	rear_corner_left	21	rear_plate_right
10	rear_up_left	22	mirror_edge_left
11	rear_up_right	23	mirror_edge_right

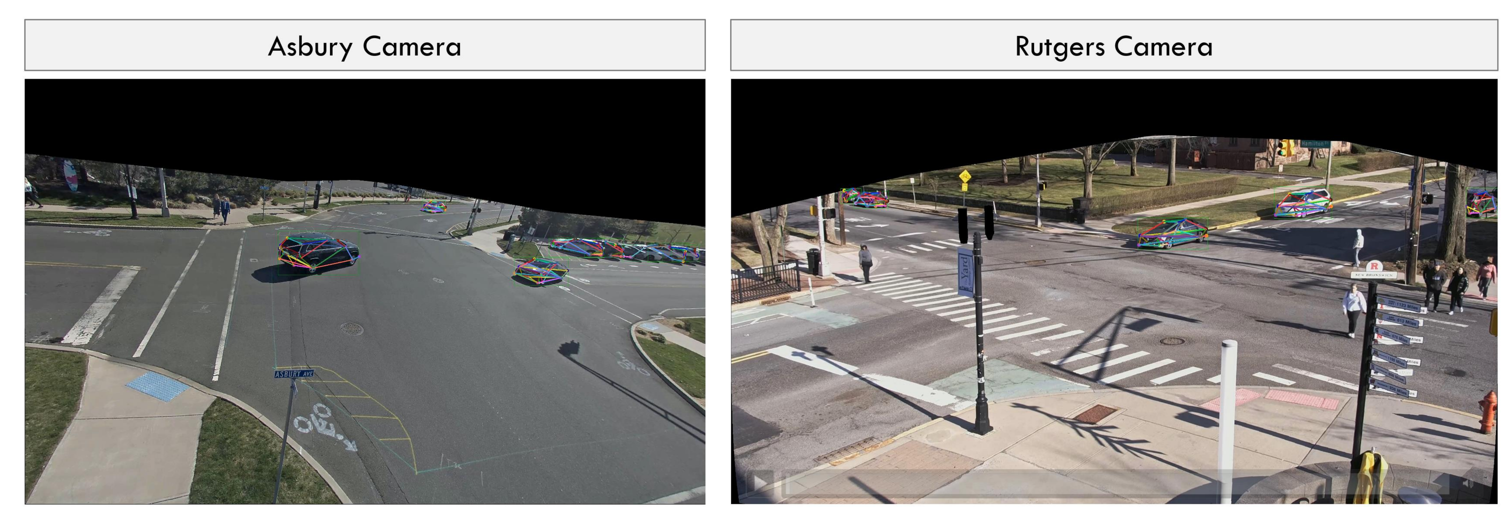
RESULTS AND DISCUSSIONS

- Camera calibration

- Tracking



- Keypoints



- Traffic analysis hub