Detection and Tracking from Multiple Cameras:
Recognition of Helmet Usage by Cyclists on Public Roads

Maikel Boonstra, Computer Vision and Data science
Supervisors: Maya Aghaei, Klaas Dijkstra

Summer 2023

Introduction

- Due to rising numbers in serious accidents among cyclists, the need for accurate number of cyclists wearing helmets is needed.
- Computer vision can be used to detect and track cyclists going over bikeroads.
- In this study, we aim to generate a dataset to test the accuracy of such algorithms.
- Additionally, we propose a pipeline to detect and track cyclists in order to keep track of the number of passing cyclists.

Materials and Methods

- For the object detector we use the small version of Yolov5[1].
- To track the cyclists over a series of frames we apply tracking by Norfair[2].
- The tracker uses a Hungarian algorithm to make predictions about where the currently detected object will be in the next frame.
- To connect the predicted location and the detection location a global minimizing distance function is used.

Abstract

This study introduces a new dataset and pipeline for tracking cyclists to address the increasing number of cycling accidents. Our evaluation reveals that cameras with a direct side view of cyclists perform best in detection and tracking. Future research can explore combining object detection from multiple cameras for improved predictions. This work contributes to the development of surveillance systems for cyclist safety.

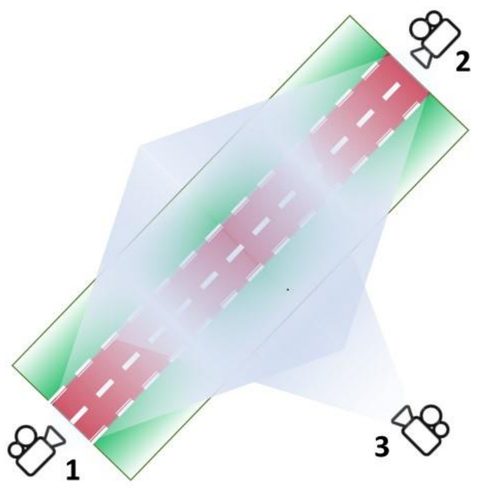

Figure 1. The above image shows the approximated location of the cameras as they have been setup by the dataset acquisition.

Materials and Methods

Camera 1 perspective | Camera 2 perspective | Camera 3 perspective

- Bikeroad

- Roundabout

- Campus

Experiments

- The first experiment revolves around the detection of helmets and cyclists. The detector is trained for the two categories on different cameras and locations to determine the best one.
- In the second experiment, the best dataset is used to evaluate tracking of cyclists from different cameras.
- The number of helmets detected throughout each track is known.

Object detection Results

Yolov5s	*Campus Dataset*			
Camera	Accuracy	Precision	Recall	F1 score
cam1	5,9%	45,4%	6,3%	11,1%
cam2	45,3%	71.3%	55.3%	62.3%
cam3	76.0%	97.9%	77.3%	**86.3%**

Yolov5s	*Roundabout Dataset*			
Camera	Accuracy	Precision	Recall	F1 score
cam1	5.6%	16.1%	7.8%	10.5%
cam2	18.3%	93.9%	18.5%	30.9%
cam3	37.8%	96.2%	38.4%	**54.9%**

Yolov5s	*Bikeroad Dataset*			
Camera	Accuracy	Precision	Recall	F1 score
cam1	3.3%	97.9%	3.3%	6.3%
cam2	17.5%	90.7%	17.8%	29.8%
cam3	26,0%	98,4%	26,0%	**41,1%**

- For all the locations, the camera perspective for the 3rd camera perspective is performing the best.
- Among most cameras, the precision is very high, meaning the predictions made are well but a lot of possible predictions are missing leading to low recall.
- The best predictions are made with training with the Campus Dataset, hence we will use this dataset to evaluate tracking.

Tracking Results

Yolov5s	IDF1	IDR	IDP	Idsw	MOTA	MOTP
cam 1	2,2%	1,1%	100,0%	1	10,9%	49,0%
cam 2	34,5%	34,3%	34,7%	39	18,5%	67,9%
cam 3	56,4%	45,1%	75,7%	5	54,9%	80,2%

- All metrics are calculated using trajectories by matching the overlapping predictions with the ground truth in such a way that they minimize the false positives and misses.
- IDF1 is the ratio of correctly identified detections over the average number of ground-truth after a global min cost function.
- MOTA shows how many errors the tracker system has made in terms of Misses, False Positives and Mismatch errors. And is based on the single frame detection.
- MOTP is the average overlap of correctly matched predictions and their ground truth.

Conclusions

- The performance of the tracker depends greatly on the performance of the object detector.
- Models have a better performance when the view is perpendicular to the objects motion.
- Cameras should be setup in such a way that they can focus on 1 part of the road and minimize the background clutter.
- Datasets should be kept as simple as possible and get more difficult after careful consideration.
- Motorcycles and passengers are currently still detected as cyclists. Due to a low appearance in the training set.

Acknowledgements

- This project is a collaboration between the Computer Vision & Data Science professorship and Provinsje Fryslân within the ROF (Regionaal Orgaan verkeersveiligheid Fryslân) programme.

NHL STENDEN
computer vision & data science

provinsje fryslân
provincie fryslân

References

([1]) https://docs.ultralytics.com/yolov5/
([2]) https://github.com/tryolabs/norfair