A Vision Based Warning System for Safe Distance Driving with Respect to Cyclists

Raluca Brehar[a], Moldovan Flavia, Attila Füzes[b] and Radu Dănescu[c]
Computer Science Department, Technical University of Cluj-Napoca, Romania
fl

Keywords: Object Detection, Monocular Depth Estimation, Cyclist Detection, Collision Warning.

Abstract: Bicyclists are one category of vulnerable road users involved in many car accidents. In this context a framework for driver warning when safety distance with respect to bicyclists is low is developed. The approach realises on object detection, monocular distance estimation for developing the driver warning algorithm. The approach was tested on benchmark datasets and on real sequences in which a mobile phone camera was used for capturing the frames.

1 INTRODUCTION

Creating a warning system for vehicles that are driven very close to cyclists is an explored area of research (Ahmed et al., 2019), considering the high number of traffic incidents in which cyclists are severely injured due to drivers' inattention. Therefore, there is a significant need for a warning system in the hope that the number of incidents will decrease as much as possible.

In this regard, a warning system for drivers is proposed by this work. The proposed approach analyses RGB video sequences captured by either (i) a standard, mobile dashboard camera mounted in the car, facing the driving area or by (ii) a fixed camera mounted on the side part of the road. In the video sequences the bicycles and vehicles are identified by using state of the art object detectors, after which the distances between them are calculated based on monocular depth estimation techniques. Depending on the distance maintained by the drivers from the cyclists, there will be several overtaking scenarios. For each overtaking case, warnings will be issued to alert the driver, prompting them to take actions such as reducing speed. Thus, by reducing speed, the distance to the cyclist will increase, and the cyclist degree of safety will increase.

The main contribution of the paper reside in:

- The design and implementation of a driver warning system that combines classical deep learning based object detection and monocular depth estimation methods.
- The improvement of the depth estimation algorithm and reduction of computational resources by the computation of depth information only on the points belonging to the nearest cyclist with respect to the car.
- Analysis of driver behaviour with respect to safety distance of bicyclists in fixed camera scenarios (for surveillance applications).

[a] https://orcid.org/0000-0003-0978-7826
[b] https://orcid.org/0000-0002-9330-1819
[c] https://orcid.org/0000-0002-4515-8114

2 RELATED WORK

In response to the increasing number of traffic accidents that involve cyclists, a lot of articles have been published exploring the use of machine learning to enhance cyclist safety.

For example, the article (Teng, 2022) emphasizes the importance of prediction and prevention to ensure rider safety. The article uses machine learning techniques like Faster R-CNN in order to explore how these technologies can be used to detect vehicles and also to estimate the vehicle speed and it's distance from cyclists. Then provide warnings to riders, so that the cyclist can understand the vehicle intentions and be capable to avoid collisions. A two-stage Faster R-CNN detector is used, so that the cyclist will be warned ahead of time it the vehicle is speeding up or slowing down.

Brehar, R., Flavia, M., Füzes, A. and Dănescu, R.
A Vision Based Warning System for Safe Distance Driving with Respect to Cyclists.
DOI: 10.5220/0012984000003822

In *Proceedings of the 21st International Conference on Informatics in Control, Automation and Robotics (ICINCO 2024) - Volume 2*, pages 190-196
ISBN: 978-989-758-717-7; ISSN: 2184-2809

(Mohd Fauzi et al., 2018) present a warning system with the purpose to alert vehicle drivers on the existence of cyclists on road. Instead of using object detection models, RFID (Radio Frequency Identification) is used. With RFID, the presence of a cyclist on the road can be detected and the vehicle driver can be alerted.

The approach of (Yang et al., 2014) is based on bicyclist detection in naturalistic driving video. It proposes a two-stage multi-modal bicyclist detection scheme that can detect bicyclists with varied poses. A region of interest where cyclists may appear is generated and candidate windows are inferred using Adaboost object detector. Having the candidate windows, these are encoded into HOG representation and using a pre-trained ELM (Extreme Learning Machine) classifier, the candidate windows will be bicyclist or non bicyclist windows.

(Ahmed et al., 2019) presents a review of recent developments in cyclist detection and also distance estimation in order to increase safety of autonomous vehicle.

Advanced Driver Assistance Systems (ADAS) (Useche et al., 2024) are another solution for preventing car-rider crashes.

The article concentrates on two questions: if and how advanced driver assistance systems can contribute to reducing road fatalities among cyclists. These systems are designed to alert car drivers to the presence of cyclist in their surroundings. In this way, the number of crashes can be reduced. In order to detect the presence of cyclists, technologies like cameras, radar and proximity sensors are used. When the car driver receives a notification about the presence of a cyclist, he has the opportunity to slow down or to wait until overtaking is safe.

There are several type of ADAS like: Forward Collision Warning (FCW) (Dagan et al., 2004), which performs real-time analysis of the information from sensors and issues warnings to the car driver. The algorithms process data on the speed, position and direction of the cyclists. Emergency braking with cyclist detection (Cicchino, 2023), this is an extension of the FCW, and when a cyclist is detected, this system can automatically activate the breaks in case of an imminent collision. The sensors are critical in these systems, they analyze the presence and movement of the cyclists and it provides rapid responses in critical situations. Blind Spot Detection (BSD) (Hyun et al., 2017), uses sensors to detect cyclists in the car's blind spots. When a cyclist is detected in the blind spot and the car driver wants to change lane, the system issues an alert to prevent the collision. In this way the risk of collision is reduced when the visibility is limited. Adaptive Cruise Control (ACC) (Li et al., 2017) is used to adjust vehicle speed in order to maintain a safe distance from the cyclist detected ahead. When the vehicle speed is too dangerous when approaching a cyclist the speed is automatically adjusted to avoid risks.

Another aspect to be considered is related to cyclist datasets available for benchmarking the algorithms, we refer to Kitti dataset that contains less than 2000 cyclist instances and the Tsinghua-Daimler Cyclist Benchmark (Li et al., 2016).

3 PROPOSED APPROACH

3.1 Processing Pipeline

The proposed processing pipeline is shown in Figure 1. Its main components are the depth estimation module, the object detection module which detects bicyclists and cars in real-time videos, the 3D reconstruction module, and the final warning system based on the monocular depth estimation and distance computation between car and bicycles.

Before any processing on the frames, a calibration of the camera was needed. Camera calibration module is the first component addressed in the development of the warning system for car drivers. This step is needed in order to correct the distortions introduced by most cameras. It also improves the measurement accuracy.

The next component has the role of detecting the objects of interest, namely cars and bicyclists, and also to estimate depth from single image. Having the original image and the depth map, a mapping of 2D points to 3D points is done and 3D scene reconstruction data is obtained.

The final step computes the absolute distance in meters and emits corresponding warning messages for car drivers based on this distance. Suggestive messages are displayed on the interface of the application. All the time the distance will be displayed on the screen, even if the distance is respected or not, so that the car driver can see anytime the distance he is keeping from the cyclist.

3.2 Depth Estimation

For the depth estimation step, several algorithms were used and their results analysed in order to see which one gives the best distance approximation.

The first explored algorithm relies on (Birkl et al., 2023) (Multiple Depth Estimation Accuracy with Single Network). The method consists of an encoder-

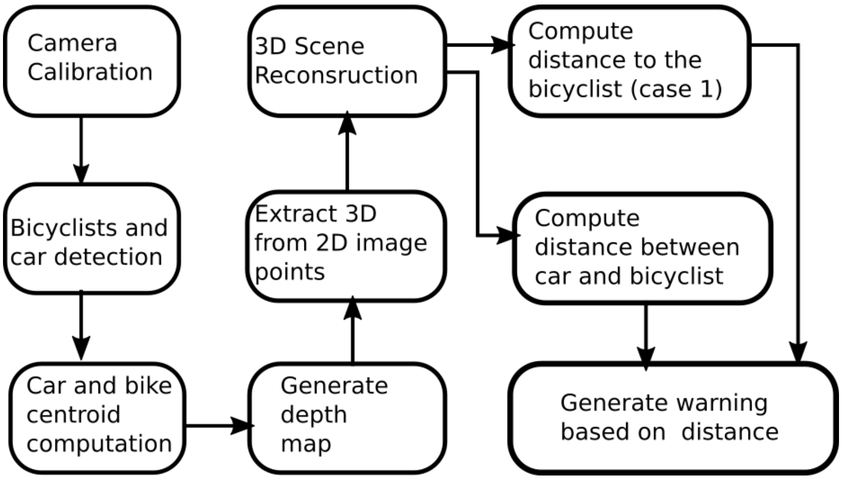

Figure 1: Proposed Processing Pipeline.

decoder architecture where the encoder does high level feature extraction and decoder does features upsampling (Paul and Godambe, 2021) and depth map generation. Unfortunately this algorithm gives only a relative depth estimation so it was not suitable for the proposed method, because of the need of absolute distance in meters as accurate as possible.

Another explored approach is based on Zoe Depth (Bhat et al., 2023) algorithm. Built on top of MiDaS, this method gives absolute distance so it is designed to make inferences in metric units. This algorithm brings improvements compared to MiDaS. Zoe Depth is based on a two-stage framework. First stage is responsible for relative depth estimation with an encoder-decoder architecture. The training of the model is done on different datasets using the same training strategy as MiDaS uses. The second stage is trained for metric depth estimation, and gives the absolute distance between objects.

A third monocular depth estimation approach that was considered is based on Depth Anything (Yang et al., 2024).

All three models are trained on indoor scenes and outdoor datasets like KITTI (Geiger et al., 2013) and NYUv2 (Nathan Silberman and Fergus, 2012).

3.3 Object Detection

In order to detect objects in real-time traffic video scenes, a single-stage detector was used, namely YOLO (Redmon and Farhadi, 2018) (You Only Look Once). Bikes and persons with high confidence scores, and and in proximity of each other, are considered good candidates for bicyclists.

3.4 3D Scene Reconstruction

This step is necessary to estimate the distance when a car is overtaking a cyclist. Unlike the previous case when a dashboard camera was used and the distance to the cyclist in front of the car was estimated, now we used a fixed camera mounted on the side of the road.

Camera calibration was mandatory before any processing on frames in order to remove distortions, as we need accurate measurements of real-world 3D coordinates from 2D traffic images. With the same fixed camera we used to take the traffic videos, we captured around 15 images of a chess board from different positions and angles and we have performed the calibration.

For this scenario, the object detection algorithm was used to detect the cars that overtakes a cyclist. The centroids of the object are considered as representative points.For each object of interest (car or cyclist) we have the coordinates of the car and the cyclist from 2D image, the distance estimation is done with Depth Anything for each point. Hence the last step was to generate 3D points using the intrisic parameters. The following formulas were used:

$$Z = Depth(u,v)X = (u - c_x) * Z/f_x \tag{1}$$

$$X = (u - c_x) * Z/f_x \tag{2}$$

$$Y = (u - c_y) * Z/f_y \tag{3}$$

where:

- (u, v) represent 2D coordinates of the pixel from original image
- Depth returns the estimated depth from the depth map for a specific point

- (c_x, c_y) represent the coordinates of the principal point (the optic center), which is usually in the center of the image
- (f_x, f_y) are the focal length on X axis and Y axis

Having two 3D points, one belonging to the center of the car and one belonging to the center of the cyclist, $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$, we compute the Euclidean distance between the two points.

3.5 Driver Warning Algorithm

After detecting the bicyclist in real-time traffic video captured by dash cam, the depth estimation algorithm is run and the car driver will be able to see on the screen the remaining distance to the cyclist in front of him. In order to display the distance in the most attractive way to capture the driver's attention, the driver warning algorithm presented listing 1. At this point we take into account the traffic code 2024 so that the driver will know at any moment if he is keeping the legal distance from the cyclist or not.

Also for the second perspective when the fixed camera is used, we determined the distance kept by the car driver when overtaking a cyclist. This distance too is passed to another driver warning algorithm which takes into account the traffic code and the legal distance that must be kept. Based on the laws (legal distance) and the actual distance, we will categorize the overtakings in legal or illegal.

3.5.1 Warning Algorithm - Approaching a Cyclist

A dashboard camera was used in order to capture real-time traffic video frames for this warning algorithm. It targets situations in which a car is approaching a cyclist in traffic. The algorithm is presented in Algorithm listing 1.

Before reaching this driver warning algorithm, the object detection algorithm was performed and only when a cyclist was detected in front of the car, the processing went further. Otherwise the process was stopped, and new frames analysed until a valid cyclist was detected on the road.

After generating the depth map, the measured distance was passed to this algorithm. Tree different levels of limits have been defined: namely safety limit when the cyclist is out of any danger, then warning limit when the car is getting a little closer, and the critical limit when the car is way too close.

3.5.2 Warning Algorithm - Overtaking a Cyclist

Overtaking a cyclist is a new perspective, at this point both car and bike need to be detected and depth map

Data: distance, safety limit, warning limit, critical limit, messages
Result: warning message according to the distance
safetyLimit = 10;
warningLimit = 5;
criticalLimit = 2;
messages["safety"] = "Distance is within the safety limits;
messages["warning"] = "Keep distance! Distance is within the warning limits;
messages["critical"] = "Critical! Distance is within the critical limits;
if *distance* $\geq$ *safetyLimit* **then**
    display message["safety"], color Green;
else
    if *distance* $\geq$ *warningLimit* & *distance* $<$ *safetyLimit* **then**
        display message["warning"], color Yellow;
    end
    display message["critical"], color Red;
end

Algorithm 1: Case 1 - approaching a cyclist.

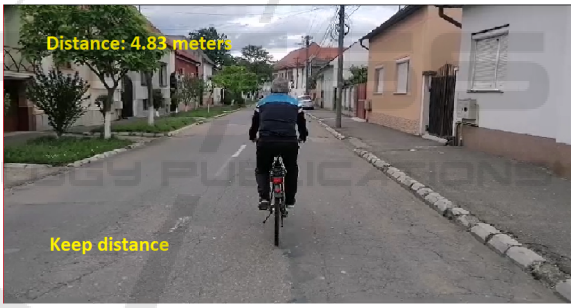

Figure 2: Displaying warning message - approaching a cyclist.

generated. After generating 3D points as presented in the previous chapter, the computed Euclidean distance is passed to this algorithm.

Two cases are considered: (i) a legal overtaking, or (2) an illegal overtaking. The traffic code for 2024 specifies the car drivers must leave at least 1.5 meters when overtaking cyclists. This distance increases as the speed increases. In this system the speed of the car was not taken into account as the focus was on other aspects, so the legal limit of at least 1.5 meters was considered. The proposed algorithm is described in Algorithm listing 2.

Data: distance, legal distance, messages
Result: warning message according to the distance
legalDistance = 1,5;
messages[”legal”] = ”Legal overtaking;
messages[”illegal”] = ”Illegal overtaking!!!;
if *distance* $\geq$ *legalDistance* **then**
  display message[”legal”], color Green;
else
  display message[”illegal”], color Red;
end

Algorithm 2: Case 2 - overtaking a cyclist.

3.6 Datasets

In order to evaluate the results, two datasets were used. KITTI dataset and a dataset that was captured for the use-cases envisioned in this paper. Kitti dataset (Geiger et al., 2013) is most commonly used in evaluating computer vision algorithms, for autonomous driving.

The dataset was collected with multiple sensors which include high-resolution color and grayscale cameras. The sensors were mounted on a vehicle. Also a Velodyne 3D laser scanner and GPS sensors were used. This dataset is divided in subsets. From these, we chose the raw data containing synchronized and calibrated data from all sensors.

The Depth Prediction dataset from category City which contains a video sequences with cyclists was also used. This dataset contains the ground truth depth map for each frame and it’s corresponding original image. It contains 93 thousand depth maps. Having the ground truth depth map, we have employed also Zoe Depth and Depth Anything to infer the depth and compared the predicted values with the real values.

For evaluating the accuracy of the distance measured when overtaking a cyclist, so the distance between the car and the cyclist, the 3D Velodyne point clouds (also provided by Kitti dataset) were used. There are over 100k points per frame and for each point we have the 3D coordinate. Having these, the Euclidean distance between the car centroid and bike centroid was computed, this is the real distance. Then these values were compared with the predicted distances computed with Depth Anything.

The proposed dataset is composed of several videos captured with the phone camera. For the first use case the phone camera was used and we have recoded a cyclist riding the bike, simulating a dashcam approaching a cyclist on the road. For each video we have measured the real distance with a roll meter. Then the real distances with the predicted distances were compared and some small errors were observed.

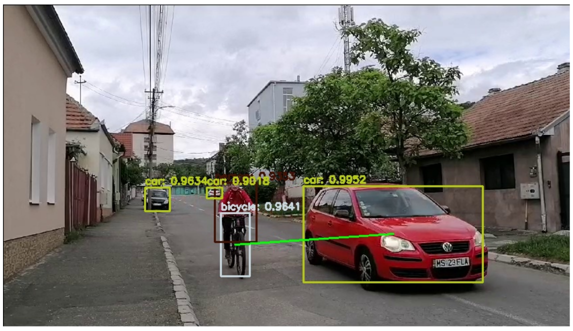

Figure 3: Object detection result.

For the second use case with the fixed camera on the side of the road, a cyclist being overtaken by a car at different distances from 1.4 meters to 3.30 meters was recorded.

Several legal overtaking and illegal overtaking sequences were recorded. When approaching a cyclist, we made sure that the cyclist was approached at different distances from 2 meters to 10 meters, so we can see all the warning messages from the warning system. We have frames where the car driver keeps the safe distance and a couple of frames when the safe distance is not respected.

3.7 Object Detection Results

In Figure 3 we can see the object detection results for the second use case when the cyclist is overtaken by a car. As we can see in this frame multiple cars are detected but only one is of interest for us, the one closest to the cyclist. We always choose the car with the highest confidence.

The video has a total of 341 frames. One of the frames is the one in Figure 3. The bike is detected successfully in 301 frames because when the bike is far away on the road, Yolo is not capable of detecting the bike, but as the bike approaches it is detected successfully. We computed the accuracy of object detection for car, person and bike.

In Table 1 we can see the precision metric computed for this specific video. Intersection over Union metric, the accuracy of the detection for the objects car, person and bicycle.

Table 1: Precision metrics for detecting objects - use case 2 (cyclist overtaken by a car).

Object	Bicycle	Person	Car
Total frames	341	341	341
Correct pedictions	301	335	341
IoU	0.90	0.99	0.99
Accuracy (%)	88	98	100

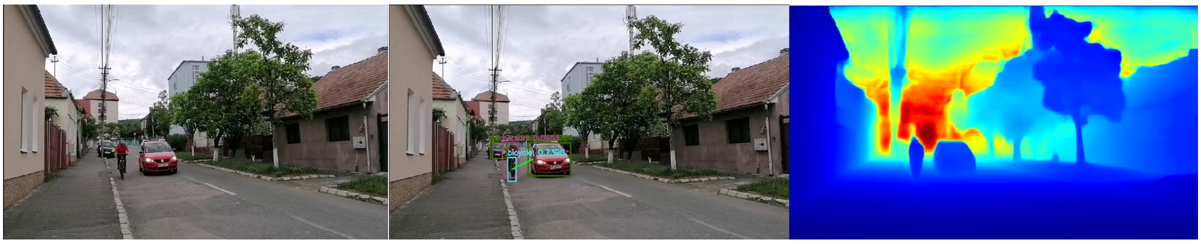

Figure 4: RGB image, detections and depth map - results on the proposed dataset.

3.8 Distance Estimation Results

The predicted distances were tested and validated on the KITTI dataset and on the proposed dataset. For distance estimation two methods were used: Zoe Depth and Depth Anything. First we generated the depth maps for each method for images from the proposed dataset and for images from KITTI dataset containing a cyclist.

Depth Anything is giving better results than Zoe Depth as we could see after comparing the predicted distances with the real distances. Using the dataset created for the specific use-cases of the paper we obtained an absolute average error of 1.84 meters. With Depth Anything we obtained an absolute average error of 0.37 meters.

The results were evaluated before and after the calibration of the camera. If before the calibration the absolute average error was of 0.33 meters when calculating the distance between the cyclist and the car, after calibration the absolute average error was only 0.05 meters.

Using the KITTI dataset and the 3D Velodyne point clouds the real Euclidean distance between two 3D points was computed and compared with the predicted distance. In this case the absolute error was of 0.14 meters. Three different scenarios were considered from the KITTI dataset containing situations in which a cyclist was approaching the car, at different distances. One point, the centroid of the cyclist was considered, and took from the ground truth depth map the corresponding depth - distance. Then run the Depth Anything algorithm on the same images and extracted the distance from the same three points. Depth Anything gave an average absolute error of 0.17 meters.

In Figure 4 some of the final results are presented. In the first picture from left to right we can see the original RGB image, then the image with the detected objects and the corresponding depth map.

4 CONCLUSIONS

Given that the safety situation for cyclists in traffic in some countries is precarious, the need for a warning system for drivers is very high. Data shows that the number of traffic accidents involving cyclists is significant, with many unfortunately suffering serious injuries and some even losing their lives. We propose a model based on state of the art object detection and monocular depth estimation methods, that warns the driver if he is approaching a cyclist at a dangerous distance. We obtained a prototype that further needs to be extended and tested for several users and real life urban traffic scenarios.

ACKNOWLEDGEMENTS

This research has been partly supported by the CLOUDUT Project, cofunded by the European Fund of Regional Development through the Competitiveness Operational Programme 2014-2020, contract no. 235/2020 and partly supported by a grant from the Ministry of Research and Innovation, CNCS—UEFISCDI, project number PN-III-P4-ID-PCE2020-1700.

REFERENCES

Ahmed, S., Huda, M. N., Rajbhandari, S., Saha, C., Elshaw, M., and Kanarachos, S. (2019). Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey. *Applied Sciences*, 9(11).

Bhat, S. F., Birkl, R., Wofk, D., Wonka, P., and Müller, M. (2023). Zoedepth: Zero-shot transfer by combining relative and metric depth.

Birkl, R., Wofk, D., and Müller, M. (2023). Midas v3.1 – a model zoo for robust monocular relative depth estimation.

Cicchino, J. B. (2023). Effects of forward collision warning and automatic emergency braking on rear-end crashes involving pickup trucks. *Traffic Injury Prevention*, 24(4):293–298. PMID: 36853168.

Dagan, E., Mano, O., Stein, G., and Shashua, A. (2004). Forward collision warning with a single camera. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 37–42.

Geiger, A., Lenz, P., Stiller, C., and Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*.

Hyun, E., Jin, Y. S., and Lee, J. H. (2017). Design and development of automotive blind spot detection radar system based on roi pre-processing scheme. *International Journal of Automotive Technology*, 18(1):165–177.

Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., Li, K., and Gavrila, D. M. (2016). A new benchmark for vision-based cyclist detection. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pages 1028–1033.

Li, Y., Li, Z., Wang, H., Wang, W., and Xing, L. (2017). Evaluating the safety impact of adaptive cruise control in traffic oscillations on freeways. *Accident Analysis & Prevention*, 104:137–145.

Mohd Fauzi, N. H., Aziz, I. A., and Mohd Jaafar, N. S. (2018). An early warning system for motorist to detect cyclist. In *2018 IEEE Conference on Wireless Sensors (ICWiSe)*, pages 7–11.

Nathan Silberman, Derek Hoiem, P. K. and Fergus, R. (2012). Indoor segmentation and support inference from rgbd images. In *ECCV*.

Paul, C. and Godambe, M. (2021). Image downsampling & upsampling.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*.

Teng, R. (2022). Prevention detection for cyclists based on faster r-cnn. In *2022 International Conference on Networks, Communications and Information Technology (CNCIT)*, pages 142–148.

Useche, S. A., Faus, M., and Alonso, F. (2024). “cyclist at 12 o’clock!”: a systematic review of in-vehicle advanced driver assistance systems (ADAS) for preventing car-rider crashes. *Front. Public Health*, 12:1335209.

Yang, K., Liu, C., Zheng, J. Y., Christopher, L., and Chen, Y. (2014). Bicyclist detection in large scale naturalistic driving video. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1638–1643.

Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., and Zhao, H. (2024). Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*.