Vision-based Scene Analysis toward Dangerous Cycling Behavior Detection Using Smartphones

Hirotaka Hayashi, Anran Xu, Zhongyi Zhou, Koji Yatani
Interactive Intelligent Systems Lab., The University of Tokyo
Tokyo, Japan
{h.hayashi,anran,zhongyi,koji}@iis-lab.org

ABSTRACT

Cycling can contribute to improvements of user's physical and mental health, and can also bring positive impacts on the environment. However, many riders do not necessarily have a full understanding of safe bicycle riding. Existing sensing technology and mobile systems have demonstrated assistive features for car driving, and this motivates us to design a mobile application to feedback cyclist's bicycle riding. In this paper, we present the current progress of our mobile system to perform vision-based scene analysis tailored toward the detection of potentially dangerous cycling.

CCS CONCEPTS

• **Human-centered computing** → **Smartphones**; • **Computing methodologies** → *Scene understanding*.

KEYWORDS

Cycling, sensing with smartphone, computer vision, scene analysis.

ACM Reference Format:
Hirotaka Hayashi, Anran Xu, Zhongyi Zhou, Koji Yatani. 2021. Vision-based Scene Analysis toward Dangerous Cycling Behavior Detection Using Smartphones. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3460418.3479300

1 INTRODUCTION

While bikes are a sustainable and flexible mobility option, many people do not necessarily have a full understanding of safe cycling. As a result, they may not be able to notice dangerous situations during their ride, causing various dangerous cycling behaviors. To provide alerts of people's dangerous cycling behavior and conditions, prior research has built various interactive systems. BikeMate [4] is a detection system for dangerous cycling which utilizes inertia sensors to detect riders' dangerous behavior like standing pedalling. In addition to inertia sensing, Beecken et al. [1] incorporated a rear camera on a bike and used vision-based methods to estimate the road conditions (e.g., on an asphalt pavement, on a gravel road, or off road). Other similar systems are available for car drivers [6]. However, due to the differences in traffic regulations (e.g. position of ride and available device size) and the speed and the magnitude of vibration between cars and bicycles, these systems are not directly applicable to cycling. We envision that sensing capabilities available in modern smartphones can detect potentially dangerous cycling besides road conditions.

We present our preliminary explorations toward a mobile system to detect various dangerous cycling behavior by using the rear camera of a smartphone mounted in a bike. Such dangerous cycling behavior includes: ignoring a stop sign; overtaking cars without slowing down; not slowing down on a road where pedestrians exist; driving at a high speed in a place where there is a risk of pedestrians jumping out. In this paper, we explain our current prototype implementation which includes vision-based analysis of video streams recorded by a smartphone to understand the scenes of the user's ride. Our current implementation performs the recognition of important elements in the scene for further analysis of dangerous cycling behavior, such as stop signs, pedestrians, and other cars, to infer the dangerous behavior mentioned above.

2 PROTOTYPE

Our current prototype extract video streams from the rear camera of a smartphone (an Android device) mounted on a bike at the sampling rate of 5Hz. We choose this sampling rate so that the system can provide cyclists feedback within a few minutes after they finish riding. Note that our current prototype takes about 15 seconds to process each frame in our environment. The camera is facing in the direction the user is cycling, and the resolution of the video stream is 655×491. We envision that our system would offer users feedback after their ride instead of in real time because we do not want to draw users' attention to our system during riding. Thus, the video streams are processed in a post-hoc manner.

In our current implementation, we mainly perform computer vision approaches to recognize the conditions of the user's surroundings during the ride. Our future work will integrate sensor data and video streams from the rear camera into the analysis results to identify dangerous cycling behavior. We combine YOLO [5] and deeplab [2] for object recognition and image segmentation on the

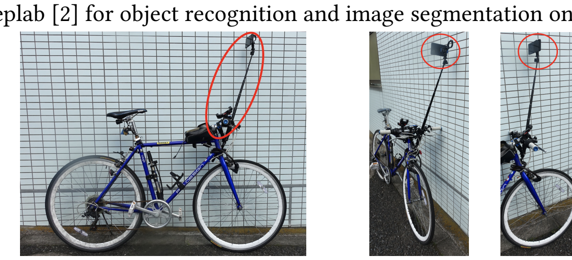


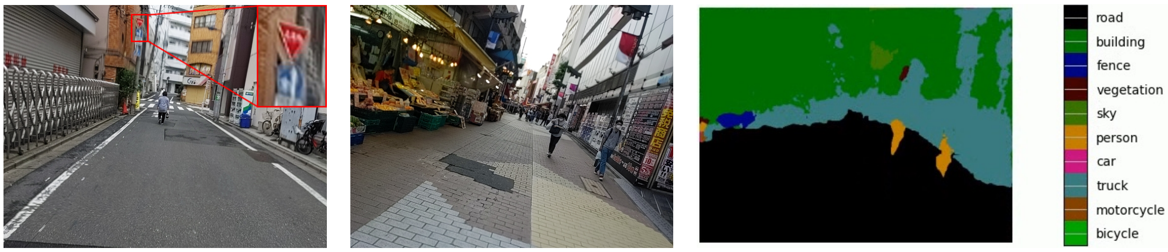
Figure 1: Smartphone installation in a bike.

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09.
https://doi.org/10.1145/3460418.3479300

(a) Stop sign **(b) Cars** **(c) Pedestrians** **(d) Fences on the left** **(e) Sidewalks on the left**

Figure 2: Examples of scenes where detection of objects and people was correct.



(a) A stop sign (red bounding box) was placed at a high location. **(b) A scene where our segmentation shows false existence of a truck.** **(c) The segmentation result of the scene shown in (b). The light blue area is the false segmentation of a truck.**

Figure 3: Examples of scenes where detection failed.

video stream. Our system utilizes our own datasets including traffic signs in Japan for YOLO and the Cityscapes dataset [3] for deeplab.

Our current implementation emphasizes on recognition of important elements to infer cycling situations which can potentially be at risk. More specifically, we have implemented the following recognition modules:

Stop Sign Recognition: We use the confidence score of the stop sign(s) detected by YOLO to recognize approaching to stop sign. We set the decay factor as 0.5 to calculate the sum of confidence scores of the target frame and all frames before it. If the sum exceeds 1.5, the system considers that the user needs to stop.

Pedestrian Recognition: Using the result of image segmentation, the system considers that pedestrians exist if the area of the segmentation of all pedestrians is over 10% of that of frame.

Vehicle Recognition: We use the result of image segmentation to recognize vehicles and their positions (right or left). In our current implementation, we use simple heuristics (e.g., whether large segmentation of a vehicle appears on the left side of the video frame[1]) to recognize whether the user is overtaking a vehicle, but future work will investigate more robust approaches.

Scene Analysis of the Cyclist's Sides: Using the segmentation result, our current implementation recognizes the presence of sidewalks, vegetation, terrains (e.g., grass, soil, or sand grounds), fences and walls on both right and left sides of the road. This information is used for estimating the risk of pedestrians suddenly jumping out.

3 PRELIMINARY RESULTS

We conducted a preliminary case study to understand the feasibility of the recognition of important elements in video frames which can be critical for the detection of potentially dangerous cycling behavior. One of the authors performed 17 bicycle rides around the campus of our university. The average video length was 76 seconds.

[1]Note that our current implementation follows the driving regulations in Japan, and the steering wheel of a vehicle is on the right side.

Figure 2 shows examples of scenes where our detection was correct. In general, our recognition performed well, and this is a promising result toward the detection of potentially dangerous cycling behavior. However, we also notice some cases where recognition failed. In Figure 3a, a stop sign was not correctly recognition. Our examination found that this stop sign was placed at a higher position than usual, and it disappeared from the camera view when the cyclist was approaching. In another example (Figure 3b), our segmentation results showed a false positive of a vehicle on the right side of the frame. The Cityscapes dataset does not include view captures from the center of wide sidewalks, and this result suggests that we will need further improvements on the analysis of scenes uniquely associated with cycling. Our future work will explore how we can reduce these failure cases.

4 CONCLUSION

We present our preliminary investigation toward a mobile system to identify potentially dangerous moments in cycling and provide feedback to users in a post-hoc manner. We will examine how we can exploit front camera images and other sensors (e.g., accelerometers, compasses, and GPS) to further improve the scene analysis and dangerous cycling behavior detection. We will also examine interface design to enable efficient review of user's cycling.

REFERENCES

[1] Janine Beecken and Andreas Reinhardt. 2019. How Smooth is my Ride? Detecting Bikeway Conditions from Smartphone Video Streams. In *Proc. PerCom 2019 Workshops*. 65–70.
[2] Liang-Chieh Chen et al. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proc. ECCV 2018*. 801–818.
[3] Marius Cordts et al. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proc. CVPR 2016*. 3213–3223.
[4] Weixi Gu et al. 2017. BikeMate: Bike Riding Behavior Monitoring with Smartphones. In *Proc. MobiQuitous 2017*. 313–322.
[5] Glenn Jocher et al. 2021. *ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations.*
[6] Fabius Steinberger et al. 2017. Designing Gamified Applications That Make Safe Driving More Engaging. In *Proc. CHI 2017*. 2826–2839.