Camera-based measurement of cyclist motion

Journal Title
XX(X):1–17

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Chris Eddy[1], Christopher de Saxe[1,2] and David Cebon[1]

Abstract

Heavy Goods Vehicles (HGVs) are overrepresented in cyclist fatality statistics in the UK relative to their proportion of total traffic volume. In particular, the statistics highlight a problem for vehicles turning left across the path of a cyclist on their inside. In this paper we present a camera-based system to detect and track cyclists in the blind spot. The system uses boosted classifiers and geometric constraints to detect cyclist wheels, and Canny edge detection to locate the ground contact point. The locations of these points are mapped into physical coordinates using a calibration system based on the ground plane. A Kalman Filter is used to track and predict the future motion of the cyclist. Full-scale tests were conducted using a construction vehicle fitted with two cameras, and the results compared with measurements from an ultrasonic-sensor system. Errors were comparable to the ultrasonic system, with average error standard deviation of 4.3 cm when the cyclist was 1.5 m from the HGV, and 7.1 cm at a distance of 1 m. When results were compared to manually extracted cyclist position data, errors were less than 4 cm at separations of 1.5 m and 1 m. Compared to the ultrasonic system, the camera system requires simpler hardware and can easily differentiate cyclists from stationary or moving background objects such as parked cars or roadside furniture. However, the cameras suffer from reduced robustness and accuracy at close range, and cannot operate in low-light conditions.

Keywords

Active safety systems, Cyclist detection, heavy goods vehicles, computer vision, object detection.

Introduction

In Britain in 2013 there were more than 19,000 road accidents involving cyclists, including more than 100 fatalities[1]. This represents 11% of all road casualties, despite cyclists only accounting for 1% of total traffic. Heavy Goods Vehicles (HGVs) accounted for 23% of cyclist deaths, despite representing only 5% of total road traffic. There is a clear need to address safety issues of cyclist-HGV interactions on UK roads. Figure 1 shows a breakdown of cyclist-HGV accidents by configuration. 43% of accidents occur when the HGV turns left across the path of the cyclist.

[1]Transportation Research Group, Cambridge University Engineering Department, UK

[2]Council for Scientific and Industrial Research (CSIR), South Africa

Corresponding author:

Chris Eddy, Transportation Research Group, Cambridge University Engineering Department Cambridge, CB2 1PZ, UK.

Email: ce302@cam.ac.uk

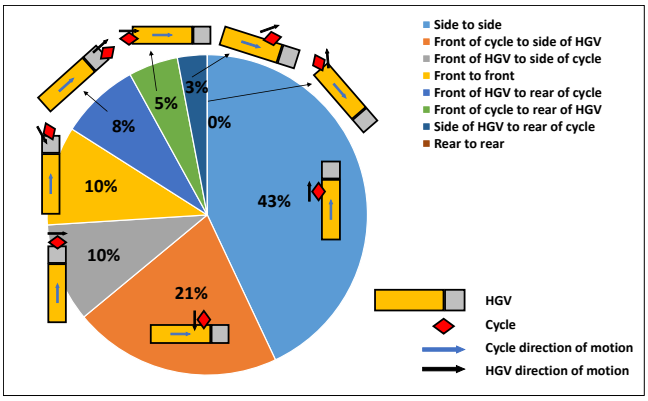


Figure 1. Breakdown of cyclist-HGV collisions by configuration. Data from Robinson and Chislett[2], graphic adapted from Jia[3]

This can be attributed to two primary causes: the large blind-spot in this area next to the HGV, and the cut-in behaviour exhibited by HGVs by virtue of their long wheelbase.

The relevance of this particular scenario is further supported by an analysis of 19 fatal cycling accidents involving left-turning HGVs in the UK by Jia[3]. Two of the accidents occurred at roundabouts and 17 at road junctions, mostly with traffic lights. In 15 of the 19 accidents the cyclist's intention was to travel straight ahead. Only four intended to turn (left) at the roundabout/junction. All the accidents occurred at speeds of less than 15 km/h. Further, 15 of the HGVs were rigid vehicles (not articulated), and most of these were construction vehicles.

The objective of this work is to develop a system that can detect and accurately locate a cyclist in the left-side blind-spot of an HGV. The system should run in real-time, have a field-of-view which covers the entire length of the vehicle, and be suitably accurate to perform relative motion predictions. The focus is on rigid HGVs in low-speed manoeuvres.

Related work

Cyclist detection systems

A number of commercial systems exist to detect and prevent low-speed collisions with vulnerable road users. These range from non-discriminating range sensors to high-end combinations of radar and cameras. Simple ultrasonic proximity systems are low cost, but do not discriminate between cyclists or pedestrians and inanimate objects such as roadside furniture. This can cause false alarms giving rise to a risk that drivers may become desensitized to alerts.

'Cycle Safety Shield', developed by Safety Shield Systems and Mobileye[4], is a camera-based system which warns the driver of the presence of cyclists. Different versions cover different fields-of-view around the vehicle, however only moving objects trigger an alert.

Cycle Eye® is a high-end system that uses a combination of image processing and radar to detect and locate cyclists[5]. The use of radar improves the accuracy in poor light conditions but adds cost. The manufacturer claims a 98.5% success rate in detecting cyclists over three days testing in London, including during rush hour.

A system based entirely on an array of ultrasonic sensors was developed by Jia and Cebon in the Cambridge Vehicle Dynamics Consortium (CVDC)[6]. The system is intended to provide an accurate but low-cost alternative to existing commercial systems, making predictions about future cyclist motion and actuating the vehicle brakes to execute an emergency stop in the event of a predicted collision. This strategy was shown to be effective at preventing reconstructed accidents in simulation and was successfully proven in low-speed field trials on a prototype system[6] (Figure 2).

Vision technologies for vehicle and cyclist detection

The use of wheel detection techniques has been popular in vehicle and cyclist detection applications, owing to the ubiquity and consistency of the features. Ardeshiri et al.[7] investigated the use of ellipse-fitting methods to detect bicycle wheels, using reflective wheel-rims and dark backgrounds to limit the number of pixels to process. The

system used the Hough transform[8,9] to detect ellipses (and hence wheels), though it was noted that this approach is very computationally expensive unless steps are taken to limit the number of input pixels processed. Variations on the Hough transform, such as the Randomized Hough Transform[10,11], or the approach described by Xie and Ji[12], can reduce computation time, but are error-prone in noisy or partially occluded conditions.

The Hough transform was also used by Lai and Tsai[13] to detect the wheels of passing cars, using the orientation and centre of the wheel to calculate relative heading and position of the two vehicles.

More general feature descriptors include Haar Features[14]. These are commonly used for face detection, using Adaboost to train classifiers[15,16] and a cascade architecture. The cascade allows background regions to be quickly ignored by the simplest classifiers, so that more computation time can be spent by the higher-level classifiers on promising 'object-like' regions of the image. An implementation of the Haar Cascade classifier is available as part of the OpenCV computer vision library[17].

This approach was used effectively by Chavez-Aragon et al.[18] to detect parts of nearby vehicles. Real-time processing was achieved by using geometric arguments to limit the region of image searched to a 'Feasible Search Zone', drastically reducing computation time.

More complex methods for image feature detection and classification exist, such as part-based models[19], often used for pedestrian detection. Although part-based models can be very accurate, they are generally computationally demanding.

In vehicle-based pedestrian detection work by Bertozzi et al.[20], an innovative camera calibration method was devised, in order to create an efficient mapping of the ground plane from image to world coordinates. The method avoided the need for full camera calibration and image distortion correction. Initial images of a calibration grid on the ground captured by the system allowed the generation of a direct pixel-to-ground coordinate mapping. This enabled efficient real-time processing.

Figure 2. Camera and ultrasonic setup

System outline and test plan

The aim of the work in this paper was to investigate whether a vision-based system could be used to measure cyclists' motion relative to an HGV with one or two cameras instead of the 10 or 12 ultrasonic sensors needed by the CVDC system. The primary advantage of a camera-based system compared to an ultrasonic system is discrimination between moving objects and stationary objects (which can be immediately ignored). This could reduce false alarms due to detection of street furniture or walls. Although complex vision-systems for object detection exist, it was proposed that simple shape-based detection might be sufficiently accurate.

System outline

The effectiveness of the proposed vision system is dependent on the type of imaging system used (for example the choice of lens) and the configuration in which it is installed on the vehicle (location, orientation and number of cameras). In establishing a suitable combination of these factors, the following criteria were considered:

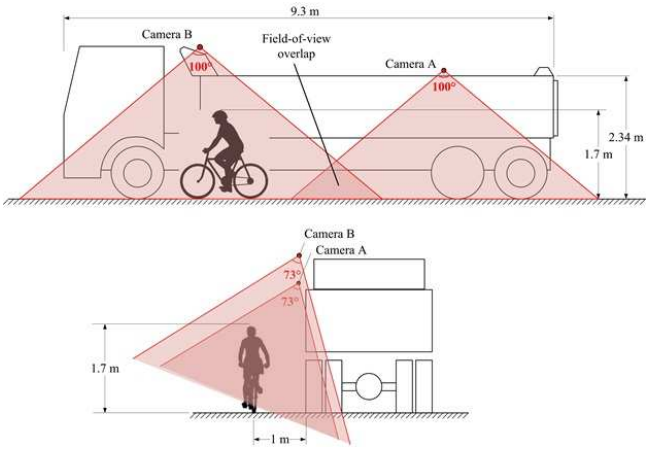


Figure 3. Camera configuration and field-of-view, shown approximately to scale with a cyclist at 1 m from the HGV

(i) The ground area covering the full length of the HGV should be visible, using the minimum number of cameras possible.

(ii) The camera field-of-view should cover the region of interest but should minimise the inclusion of background scenery and other moving objects.

(iii) Potential classification features such as wheels should be visible.

(iv) Occlusion problems should be minimised in instances where more than one cyclist is present.

Sample images were obtained from various points on the test vehicle to determine the best location for the camera. The chosen configuration is illustrated in Figure 3. The vehicle is shown to scale and a representative silhouette of a cyclist is included for reference. The system was mounted on the same rigid construction vehicle used by Jia[6] (see Figure 2).

The high mounting point of the cameras was a key decision in the design of the system. An important benefit of the top-down view is that lateral position errors arising from image processing are minimised, because a single pixel uncertainty in the measurement in image coordinates corresponds to a much smaller lateral distance if the cameras are looking almost straight down compared to if the cameras were mounted at wheel height and looking 'across' the ground plane. This vantage point also addresses the fourth point above, by minimising potential occlusion. The visibility of potential classification features (as in the third requirement) may be slightly reduced, compared to lower mounting points, but this was shown not to be prohibitive.

To address the first and second criteria, two ultra-compact Point Grey Flea3®. USB 3.0 cameras fitted with Fujinon 2.8 to 8 mm wide-angle lenses with a maximum field-of-view of 100° were selected[21,22]. The cameras were located longitudinally so as to achieve maximum coverage along the full length of the vehicle. There is a region of overlap between the two views which is important for the transition of tracking information between the two cameras. The front camera (camera B) was mounted slightly higher than the rear camera (camera A) due to the available height at that point on the tipper bucket. A small outwards tilt to the cameras extended the lateral viewing distance.

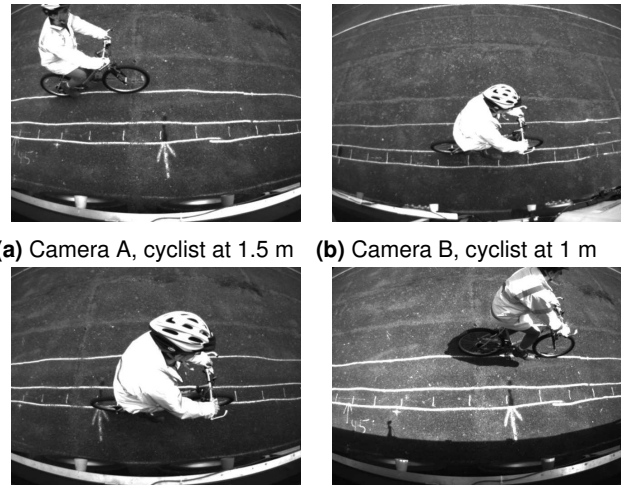
Test program

Tests were carried out on an open area of tarmac at Bourn Airfield, near Cambridge, UK. Parallel passing manoeuvres between the test vehicle and a cyclist were carried out at various passing distances.

A straight line was marked on the road as a guide for the driver to follow, such that the line roughly approximated the left side of the HGV. Parallel to this, lines were marked at distances of 0.75 m, 1 m and 1.5 m as guidelines for the cyclist. Transverse tick marks at 0.5 m spacing were included in order to estimate cyclist and vehicle speed during post-processing. The lines are visible in Figure 4.

Three runs of each test were conducted to allow for variations. A total of 18 sets of data were recorded, six each at 0.75 m, 1 m and 1.5 m spacing (consisting of three repetitions at HGV speeds of 5 km/h and 8 km/h).

A schematic of the instrumentation layout is shown in Figure 5. The cameras were connected via USB 3.0 to a dedicated computer located inside the driver's cab. Synchronous greyscale images were captured from both cameras at 20 frames per second (fps) with a resolution of



(a) Camera A, cyclist at 1.5 m **(b)** Camera B, cyclist at 1 m

(c) Camera A, cyclist at 0.75 m **(d)** Camera A, cyclist at 1.5 m, sunny

Figure 4. Sample images at various lateral separations and lighting intensities

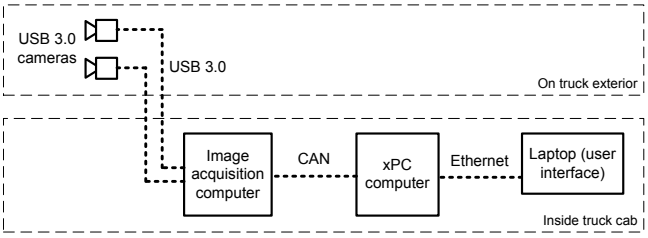


Figure 5. Instrumentation schematic for image acquisition

640×480. A slave computer running MATLABs xPC Target toolbox was used for data-logging, and a laptop computer was used as the primary user interface. CANbus was used for communication between the camera computer and xPC slave unit.

Image processing

Strategy

Although the selected camera system and its particular configuration has a number of benefits, it also presents some challenges. Firstly, the elevated camera positioning results in the cyclist's shape varying with lateral distance. For example, circular wheels are viewed as thin ellipses from above at close range, and the proportion of head and torso in the cyclist's silhouette grows with proximity to the vehicle. Similarly, the cyclists shape is variable from the left to the right of the camera field-of-view, due to camera position and lens distortion. These effects are highlighted in Figure 4a to 4c.

Secondly, as with any vision-based system, variations in lighting can be problematic. Figure 4d shows the effects of strong light conditions on an image. The shadow covering the rear wheel makes it more difficult to distinguish from the background.

The overall image processing strategy of the system can be summarised into four parts:

(i) Wheel detection, to identify the presence of a cyclist in the image.

(ii) Contact point location, to locate the positions of the contact points between wheels and road.

(iii) Ground mapping, to convert image coordinates to world coordinates.

(iv) Cyclist tracking using a bicycle model and Kalman Filter, to mitigate spurious or occluded detections and predict trajectories.

These are discussed separately in the following sections.

Wheel detection

Wheel detection was used to detect the presence of a cyclist in the images. Both ellipse-based and classifier-based methods for wheel detection were explored. Wheels were chosen as recognisable features since they are common to all bicycles with only minor variations. Ellipse-fitting methods considered included Edge Following[23,24], Genetic Algorithm-based approaches[25], and the Hough transform[12]. The approach to the Hough Transform described by Xie and Ji[12] was implemented in Python, and a frame rate of 10 fps was achieved. However, the algorithm is not robust to occlusion of one side of the wheel, which was a common occurrence. As an alternative, the Edge Following approach was implemented but achieved only 1 fps maximum processing speed.

The classifier-based method of Viola and Jones[14] was implemented in Python, using the OpenCV computer vision library[17]. The output of the classifier is a bounding box surrounding the detected wheel feature. This method was found to be most suitable, and yielded an acceptable frame rate of more than 20 fps.

Training data for detection and classification work is available for cyclists, including the SUN[26] and KITTI[27] databases. However, the cyclist images in these datasets are largely from a ground level reference and are not suitable for the highly oblique view which results from our raised camera setup. As a consequence, the data from two of the three runs of each test were used as training data for the third run. This is of course not suitable for a generalised system which should be robust to varied cyclists and backgrounds. However, it was deemed suitable at this proof-of-concept stage. More generalised training data will be obtained and used to retrain the classifiers in future work.

Due to the relatively small number of training images available (approximately 900 images per camera for each test run), separate classifiers were trained for each lateral distance from the HGV. Positive image regions were marked manually and images without wheels visible were used as negative training data.

Figure 6 shows examples of positive training images. Between 150 and 300 positive images were used to train each classifier, depending on how many frames the wheels were visible for. The positive samples were scaled to 24×24 pixels. Only 75 negative images were used due to lack of variation between the images.

This method was fast enough for real-time implementation. However, the location accuracy was not sufficiently precise since the detection bounding box could move relative to the feature it enclosed. Due to the relatively small amount of training data available, the classifiers were not very robust. In order to guarantee detection of the wheels, the detection threshold was kept low, which led to a high rate of false positive detections. Figure 7 shows example outputs of the detection step.

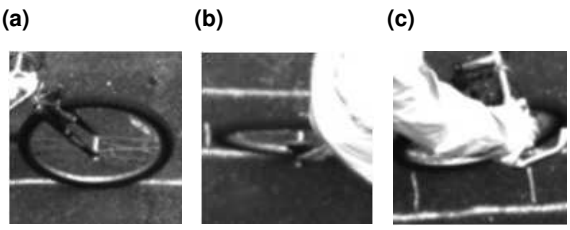


Figure 6. Examples of positive training images

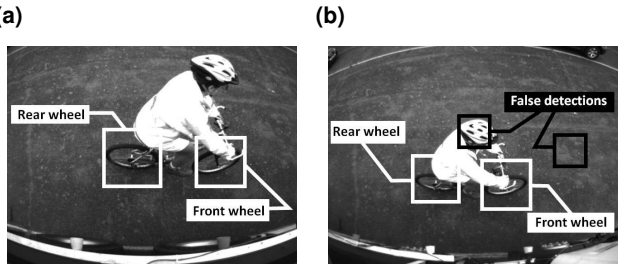


Figure 7. Examples of (a) correctly and (b) wrongly detected features

It should be noted that this combination of testing and training images is insufficient for robust implementation. First, the number of negative training images should be significantly larger than the number of positive training samples, and secondly, both testing and training were carried out on the same style of bicycle. This is due to the lack of availability of training data for the oblique camera angles used in this work. Some ad-hoc testing has been carried out on classifiers trained with multiple different cyclists and bicycles and found to work well.[28]

The current work is intended as a proof-of-concept of the detection-location-tracking-prediction system as a whole, thus it was considered that a partially-trained classifier would give sufficiently accurate results. Future work should include training the classifier on a much larger database of images collected from the elevated camera angle.

Contact point location

Once wheels have been detected, the contact point with the road must be determined. Several methods were considered for finding the ground-wheel contact point within the

detected bounding boxes. The methods considered and their performance are summarised as follows.

(i) The Hough transform[12] was used to fit ellipses inside the feature bounding box.

(ii) Fitzgibbon's algorithm[29] was used to fit ellipses, combined with RANSAC[30] to remove outliers.

(iii) A simplified version of the Starburst algorithm[31] was used to limit the number of pixels in the box.

(iv) The ground point was assumed to be a fixed distance down the centreline of the bounding box, where the distance varied according to the passing distance between HGV and cyclist in order to prevent loss of accuracy at close range.

(v) The ground contact point was taken as the lowest point on an edge in the cropped image. A grey-scale threshold of 50 was first applied to remove bright patches, such as road markings (Figure 8b), then the images were normalised to maximise the contrast between the tyres and the road (Figure 8c), and finally Canny edge detected[32] with a high threshold to ensure that noise from the road surface was removed (Figure 8d). The cropped image was then searched in columns to find the lowest edge pixel (Figure 8e). The threshold and normalisation steps largely remove susceptibility to lighting conditions, though more work is needed to ensure complete robustness.

For methods (i) to (iii), the cropped images were first pre-processed with a Gaussian blur and Canny edge detection[32]. The Hough transform (i) was computationally expensive and unreliable due to noise and occlusion in the images. Combining Fitzgibbon's algorithm with RANSAC (ii) was a more reliable method of ellipse fitting, but still took too long to run. The simplified Starburst algorithm (iii) was inaccurate due to the noise and occlusion in the images. Using the full Starburst algorithm might be more accurate, but would again be computationally expensive. Assuming a fixed position within the bounding box (iv) was accurate when the wheel was near to the centre of the field-of-view but introduced errors of up to 3 cm at the edges of the image due to the camera distortion. This approach has the benefit that the contact point can be estimated even when it is occluded by the cyclist. Edge detection and minimum point selection (v) was fast and accurate but was less accurate if the contact point was occluded. Therefore this was the method chosen except for close range tests where the fixed location method was used instead. It is possible for the contact point to be occluded in the tests at longer range. However method (v) simply returns the location of the edge point which is closest to the truck, which is likely to be whatever is obscuring the wheel. This method is therefore still fairly accurate, and so can be used, especially at longer test distances where accuracy is less critical because the cyclist is further from danger.

Ground mapping

A method was required for converting the detected cyclist's position into a world coordinate system. One approach to this would be to rectify the distorted images, and then use known camera parameters to perform a full 3D calibration. However, this adds a computationally expensive processing stage. Ground mapping was proposed as a simpler alternative, which does not require an undistorted image.

The aim of the ground mapping process was to convert the coordinates of the points of contact between the bicycle wheels and the ground from the image coordinate system to a global coordinate system (relative to the HGV). The point on the ground directly below the front left corner of the vehicle was chosen for the origin of the HGV coordinate system. Defining bicycles in the ground plane by their contact points simplified the calibration of the camera to a planar mapping. There is an intermediate step needed to transform the image coordinates into the HGV coordinate system because the origin of the HGV coordinate system was not visible in either

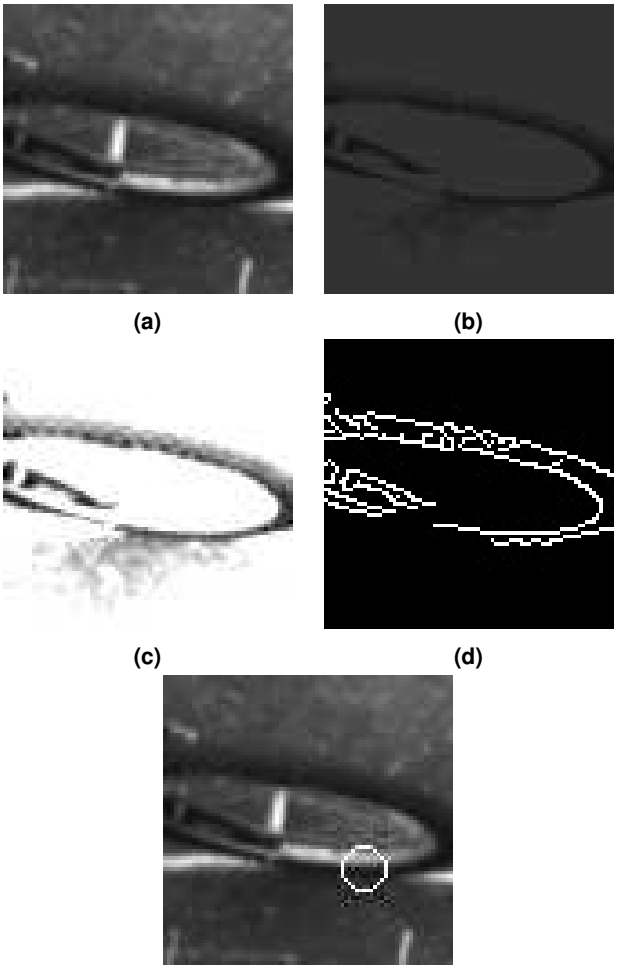


Figure 8. Stages in the extraction of the ground contact point. (a) Cropping (b) Thresholding (c) Normalisation (d) Edge Detection (e) Selection of the lowest pixel

camera's field-of-view. The contact point between the left-most HGV wheel visible in each camera's field-of-view and the ground was chosen as the origin for the intermediate HGV-based coordinate system. Both intermediate HGV-based coordinate systems were then translated into the world coordinate system.

The cameras were calibrated by positioning the vehicle next to a calibration grid (Figure 9).

The image coordinates (u, v) of the grid intersection points were extracted manually; the world coordinates of each grid intersection point (x, y) were already known. The camera lenses introduced barrel distortion which is approximately quadratic[33]. Therefore, a quadratic function was used to approximate the transformed shape of the grid in both dimensions. Shape-function-based interpolation was used

Figure 9. Calibration grid processed to cover the entire image

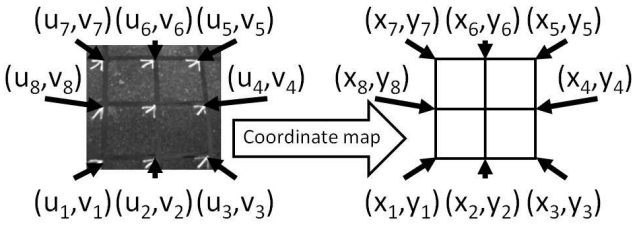


Figure 10. General illustration of coordinate mapping

to generate a map from image to world coordinates, as described by Silva et al.[34]. The mapping is described as follows:

$$u(x,y) = c_1 + c_2x + c_3y + c_4x^2 + c_5xy \ldots + c_6y^2 + c_7x^2y + c_8xy^2 \quad (1)$$

$$v(x,y) = d_1 + d_2x + d_3y + d_4x^2 + d_5xy \ldots + d_6y^2 + d_7x^2y + d_8xy^2 \quad (2)$$

where c_i and d_i are constant coefficients. For an example intersection point (x_1, y_1) the mapping to (u_1, v_1) would be as follows:

$$u_1(x_1,y_1) = c_1 + c_2x_1 + c_3y_1 + c_4x_1^2 \ldots + c_5x_1y_1 + c_6y_1^2 + c_7x_1^2y_1 + c_8x_1y_1^2 \quad (3)$$

$$v_1(x_1,y_1) = d_1 + d_2x_1 + d_3y_1 + d_4x_1^2 \ldots + d_5x_1y_1 + d_6y_1^2 + d_7x_1^2y_1 + d_8x_1y_1^2 \quad (4)$$

A one-meter square as shown in Figure 10 has intersection coordinates (x_1, y_1) to (x_8, y_8), where (x_1, y_1) = (0, 0), (x_2, y_2) = (0.5, 0), (x_3, y_3) = (1, 0), (x_4, y_4) = (1, 0.5), (x_5, y_5) = (1, 1), (x_6, y_6) = (0.5, 1), (x_7, y_7) = (0, 1) and (x_8, y_8) = (0, 0.5). These values of (x_i, y_i) can be substituted into (2) to yield values of u_i and v_i for i = 1 to 8. For example:

$$u_2(0.5, 0) = c_1 + 0.5c_2 + 0.25c_4 \tag{5}$$

$$v_2(0.5, 0) = d_1 + 0.5d_2 + 0.25d_4 \tag{6}$$

These expressions for all eight intersection points can be written in matrix form:

$$\mathbf{u} = A\mathbf{c} \tag{7}$$

where

$$\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 \end{bmatrix}^\mathsf{T} \tag{8}$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0.5 & 0 & 0.25 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0.5 & 1 & 0.5 & 0.25 & 0.5 & 0.25 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0.5 & 1 & 0.5 & 0.5 & 1 & 0.25 & 0.5 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0.5 & 0 & 0 & 0.25 & 0 & 0 \end{bmatrix} \tag{9}$$

$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 \end{bmatrix}^\mathsf{T} \tag{10}$$

Similarly

$$\mathbf{v} = A\mathbf{d} \tag{11}$$

Given a set of known image coordinates $(\mathbf{u}, \mathbf{v})$, $\mathbf{c}$ can be found by

$$\mathbf{c} = A^{-1}\mathbf{u} \tag{12}$$

$$\mathbf{d} = A^{-1}\mathbf{v} \tag{13}$$

This process of finding $\mathbf{c}$ and $\mathbf{d}$ from images of the ground grid constitutes the only calibration needed for the cameras. Given a world coordinate point (x, y), the coordinates of the transformed point in the image (u, v) can be found by substituting $x = x_1$, $y = y_1$, c_1 to c_8 = $\mathbf{c}$ and d_1 to d_8 = $\mathbf{d}$ into Equations 3 and 4.

However, given image coordinates (u, v), the corresponding point (x, y) is not directly obtainable. To calculate (x, y) an initial guess of the world coordinates is made and mapped to image coordinates. This is compared to the target point and the guess adjusted according to the error. This process is repeated until the transformed coordinates converge to the target point. A separate ground plane calibration map was required for each camera in this case, owing to their different heights and small variations in pitch and internal parameters.

To speed up the real-time element of the program the conversion was calculated in advance for every pixel, and stored in a lookup table. The calibration process should only be required once per vehicle unless the cameras are moved.

From the 40 grid intersection points shown in Figure 9, only eight are required for each 2D quadratic approximation. To maximise the accuracy, the calibration image was split into ten patches, in two rows of five, with overlap between the two rows to ensure continuity in the more critical lateral direction. Each patch was treated separately. This introduced small discontinuities between the patches which could be reduced by using a finer calibration grid, but this was not considered necessary. All measurements of position and velocity were transformed into *relative* measurements between cyclist and vehicle.

The resolution of the images limited the precision of the manual extraction of grid coordinates to approximately ± 3 pixels which can be shown to correspond to an error in world coordinates of up to 4 cm. This calibration assumes the HGV was perfectly aligned with the grid while the

calibration images were taken, which was not necessarily the case. This could introduce an additional offset to the final position outputs. In total, errors due to the coordinate conversion of up to 7 cm are likely, although the uncertainty will vary across the field-of-view with higher uncertainty corresponding to regions of highest distortion on the images.

Once the image coordinates of the wheel-ground contact points were extracted from the images, they were passed through the coordinate map, and then translated so as to be relative to the global origin under the front left corner of the HGV.

Cyclist tracking

In order to track the cyclist's motion using a Kalman Filter, the cyclist was modelled by converting the coordinates of the contact points of the front and rear wheels to a yaw angle, wheelbase and position of the centre-of-mass of the cyclist. Due to the relatively high rate of false positive detections of the wheels, the positions of the front and rear wheels were used to validate each other: a bicycle detection would not be confirmed unless both wheels were detected in the correct relative positions. This check was performed in world coordinates and so the acceptable relative position was governed by an approximate bicycle wheelbase of 1.2 m, and a maximum expected yaw angle of $\pm 5^\circ$ relative to the x-axis.

Any detection with a plausible wheelbase was compared to detections from the previous frame, and a maximum velocity limit of 25 cm per frame in the direction of travel and 8 cm per frame laterally was imposed at 20 fps. These values were determined from an assessment of feasible cyclist motions. The bicycle was then tracked and its future position predicted, so that in future frames only one wheel needed to be detected, and checked against the expected position.

A simple Kalman Filter[35] was added to reduce measurement noise. Constant accelerations both parallel and perpendicular to the direction of motion were assumed. This also had the effect of providing motion estimates even in ranges where detections were missed. The positions of the front and rear wheels were averaged to output a list of positions of the approximate centre of the cyclist (mid-wheelbase) in each frame.

The prediction equations of the Kalman filter were:

$$\hat{\mathbf{X}} = \mathbf{X}_{k-1} + \dot{\mathbf{X}}_{k-1}\Delta t \tag{14}$$

$$\mathbf{p} = \mathbf{P} + \mathbf{Q} \tag{15}$$

where $\mathbf{X}$ is the state vector (lateral and longitudinal displacement and velocity of the center of the cyclist's wheelbase), $\hat{\mathbf{X}}$ is the prior estimate of the state vector, $\mathbf{P}$ is the error in the estimate, $\mathbf{p}$ the prior estimate of the error, and $\mathbf{Q}$ is the process covariance.

The update equations were:

$$\mathbf{K} = \mathbf{p}(\mathbf{p} + \mathbf{R})^{-1} \tag{16}$$

$$\mathbf{X}_k = \hat{\mathbf{X}} + \mathbf{K}(\mathbf{z} - \hat{\mathbf{X}}) \tag{17}$$

$$\mathbf{P} = (1 - \mathbf{K})\mathbf{p} \tag{18}$$

where $\mathbf{z}$ is the observed states, $\mathbf{R}$ is the model covariance, and $\mathbf{K}$ is the Kalman gain.

The prediction equations produce estimates of the system states and their uncertainties. The update equations take these estimates and the observations from the image processing and calculate a weighted sum, giving a higher weighting to the more certain predictions. The model covariance was set to 1 for all states, and the process covariance set to 1.2 for the position states and 2 for the velocity states. These values were approximated from inspection of the covariance of the unfiltered states and then adjusted to give suitable results.

A Python implementation of the whole algorithm ran at an average of 7.7 fps on a 3.6 GHz laptop. Analysis by Jia[6] showed that for effective intervention in the HGV motion, the system should predict 1.5 seconds ahead. At typical closing

speeds, this requires a minimum of 7.5 fps. The algorithm frame rate was therefore deemed suitable.

Error analysis

As no 'ground truth' position of the cyclist was available, the position of the ground contact point was manually extracted from each of the images in order to remove the effect of imperfect following of the nominal position lines by both cyclist and HGV. However, there are errors associated with this process, both in the mapping itself and due to imperfect alignment between the vehicle and the calibration grid during the capture of calibration images.

This manual measurement was designated M, the camera system's measurement C and the 'true' position of the cyclist T. The maximum uncertainty between the manual measurement and the true position, $\varepsilon_{MT,max}$ was approximated as the sum of two components—a 3 cm uncertainty in the drawing of the calibration grid, and a three pixel uncertainty in the accuracy of manually selecting points from images. This three pixel uncertainty was converted to world coordinates at different lateral distances from the vehicle, representing increasing distances at higher separation as anticipated from the angle of the cameras, equalling 2 cm at 0.75 m, 3 cm at 1 m and 4 cm at 1.5 m. Combined with the 3 cm uncertainty from drawing the grid, this gave $\varepsilon_{MT,max}$ = 5 cm, 6 cm and 7 cm at 0.75 m, 1 m and 1.5 m respectively.

The standard deviation of the uncertainty, σ_{MT} was estimated from the maximum error between the manually-extracted position and true position. Assuming the errors followed a Gaussian distribution, 99% of the data lie within three standard deviations of the mean, leading to $\sigma_{MT} \approx \varepsilon_{MT,max}/3$ giving σ_{MT} = 1.67 cm at 0.75 m, 2.00 cm at 1 m, and 2.33 cm at 1.5 m. The mean uncertainty was assumed to be zero. Although a slight bias possibly occurred due to the nature of the data extraction task, this was likely to be small and impossible to quantify.

The error between the camera measurement and the manual measurement (ε_{CM}) had a different mean and standard deviation for each test run. The total error between the camera measurement and true position ε_{CT} was calculated as the sum of ε_{CM} and ε_{MT}. The mean (μ) and standard deviation (σ) were found by assuming that the errors ε_{CM} and ε_{MT} were uncorrelated, according to:

$$\varepsilon_{CT} = \varepsilon_{CM} + \varepsilon_{MT} \tag{19}$$

$$\mu_{CT} = \mu_{CM} + \mu_{MT} \tag{20}$$

$$\sigma_{CT}^2 = \sigma_{CM}^2 + \sigma_{MT}^2 \tag{21}$$

The standard deviations were also normalised as a percentage of the nominal passing distance.

The camera system can continue to make estimates of the cyclist's position using the Kalman Filter if a previously-detected wheel becomes occluded, so there is no loss of data at the edges of the fields-of-view of the cameras. This contrasts with the manually-extracted position data points, which cannot be extrapolated in the case of an occluded cyclist and therefore do not cover the same longitudinal range as the camera-measured position. There are data points missing in the region around X = 5 m and at the highest and lowest values of X. This corresponds to points close to the edge of either camera's field-of-view, where one wheel is occluded, so manual position extraction is impossible. The camera system can detect a wheel even if the contact point is fully occluded, if the view of the cyclist is sufficiently similar to training images. Additionally, the camera system can predict from previous positions, or from detection of a single wheel. This loss of data points is more significant at d = 0.75 m where the wheels are occluded further from the edges of the camera field-of-view.

As an initial validation of the wheel detection algorithm, the relative longitudinal velocity between cyclist and HGV was compared with manual measurements taken from the

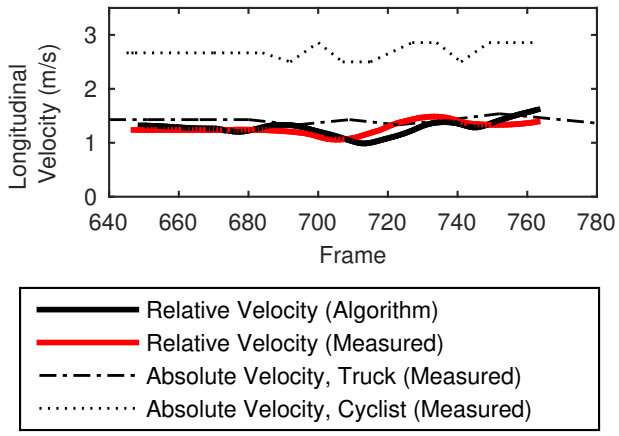


Figure 11. Comparison of calculated and measured longitudinal velocities

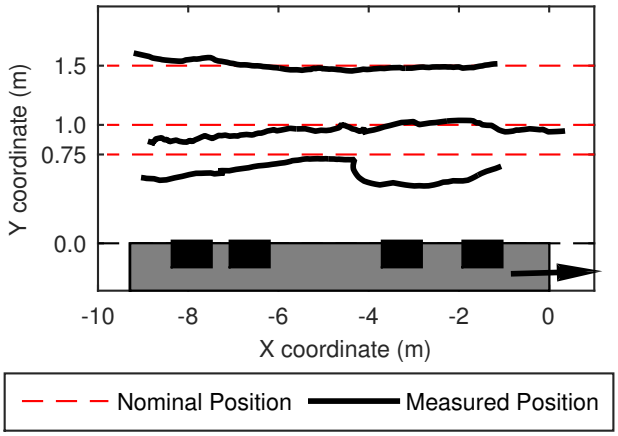


Figure 12. Output of camera-based detection system over three different lateral distances from the side of the HGV

lateral tick marks on the ground. The manual extraction was approximate due to discretisation errors: there may not be an image at the exact moment a wheel passes a tick mark. The manual speed measurement was smoothed by taking a moving average.

Results and discussion

Figure 11 shows the relative and absolute longitudinal velocities of the HGV and cyclist for the cyclist nominally at 1 m from the side of the HGV (d = 1 m). The results indicate reasonable agreement between the algorithm and the measured speeds.

Figure 12 shows the trace of the camera-estimated position for three of the test runs at different lateral distances. The nominal position is the location of the marked lines on the road at d = 0.75 m, 1 m and 1.5 m from the HGV, shown in dashed lines on the figure. These results show reasonable performance in the estimation of the cyclist's position relative to the nominal position, although comparison to the nominal position is of limited value as the cyclist and the HGV may not have followed their respective lines precisely. However, the estimated position is within a 10 cm window of the nominal line in most cases. Errors are higher at smaller values of d because the wheels were more often occluded by the cyclist's body. This reduced the number of observations, thus reducing the robustness of the Kalman Filter. The occlusion of the wheels at the edges of the fields-of-view of the separate cameras also contributed to the large discontinuity in position at the join between the left and right cameras at $X \approx$ -4.5 m for d = 0.75 m, as the position estimate there was based on prediction rather than observations.

Table 1. Average detection errors across all tests.

Nominal distance (m)	1.5	1.0	0.75
σ_{CM} (cm) (measured)	3.27	3.66	4.72
σ_{MT} (cm) (estimated)	1.67	2.00	2.33
σ_{CT} (cm) (calculated)	3.67	4.17	5.26
Normalised σ_{CT}	2.4%	4.2%	7.0%

In total, 18 sets of testing images were recorded—six each at 0.75 m, 1 m and 1.5 m. These included a range of passing speeds between the HGV and the cyclist, and a range of lighting conditions, from overcast to bright sunlight, with combinations of HGV and cyclist shadows in different orientations. The lighting conditions had no noticeable effect on the accuracy of the detection. Table 1 summarises the average results for each of the three test distances.

Figure 13 shows a comparison of the camera-measured position and the manually-extracted position for one test run at each of the test distances.

At 1.5 m distance (Figure 13a), the camera detection matches closely the manually extracted positions, with a

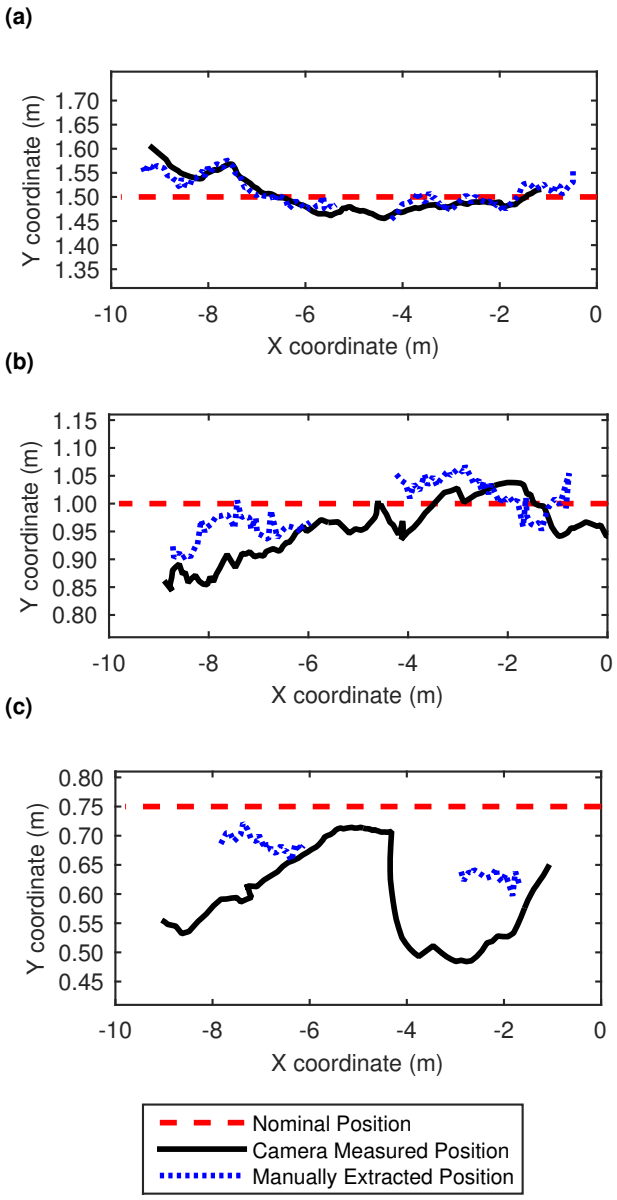


Figure 13. Comparison between camera measurements and manually-extracted data points for a single run at (a) 1.5 m separation (b) 1 m separation (c) 0.75 m separation

maximum error of 4.1 cm at $X = -9.1$ m. This is close to the width of the bicycle tyre and also to the uncertainty introduced by the calibration system at that distance which had a standard deviation of ± 1.67 cm.

Across all six test runs at a nominal spacing of 1.5 m, the camera system performed very well. The errors between the camera system and the manually extracted coordinates were very small, with a standard deviation across all six runs of only 3.27 cm—only slightly greater than the tyre width. The error between the camera measurements and manual measurements dominates the error between the manual measurements and true value. This implies that the calibration process is very accurate and reliable, even at d = 1.5 m.

At 1 m distance (Figure 13b), occlusion prevents manual measurements between the two cameras (X = -6 m to -3 m). The errors relative to the manually extracted position are slightly larger, but still close to the 3 pixel uncertainty window (corresponding to 3 cm at d = 1 m), again implying that the image processing system can find the ground contact point at least as accurately as a human. The maximum error for the displayed test run was 10.9 cm at X = -7.4 m, but the standard deviation across all tests was 3.66 cm. The higher maximum error suggests that the camera system is slightly less robust at closer range, as the wheels become more liable to occlusion, but the overall accuracy is similar. As predicted, at closer range, the uncertainty in the calibration process drops, as the camera 'looks down' on the closer points instead of 'across' them, allowing the ground point to be more accurately defined. This causes the errors due to the detection stage to become even more dominant as the range reduces.

At 0.75 m distance (Figure 13c), the camera system is much less reliable. Occlusion strongly limits the areas where manual extraction can be performed. The maximum errors are much larger (15.2 cm at -2.7 m), although the standard deviation is still under 8 cm across all the test runs. The calibration process is more accurate at the closer distance, so the errors in the detection stage dominate, leading to a standard deviation in the error between the camera measurement and the true position of up to 5 cm. This is largely due to significant occlusion of the wheels at close range by the cyclists body. Since the classifier training dataset included images of partially occluded wheels the system can still estimate the position of the wheel, but accuracy is reduced compared to the fully visible case.

The errors at the closest nominal distance were noticeably larger than at further distances. Inspection of the output of the detection stage for these runs shows significant loss of accuracy at the gap between the images taken by the two cameras. This often caused the Kalman Filter to fail for all or part of the test, without enough observations to inform the model. This loss of detections was most significant at the closest distance because the wheels are more easily occluded when the cyclist is close to the camera, leaving only two small patches, one in the centre of each camera's field-of-view where the detection system was working well. This pattern was consistent across all the runs at 0.75 m. Reducing the separation between the cameras, or increasing the number of cameras would eliminate this blind-spot in the field-of-view, and improve the accuracy significantly.

Jia[6] quoted the accuracy of the ultrasonic measurement system as a standard deviation of 3.4 cm at a nominal passing distance of 1 m. This is very similar to 3.92 cm for the camera system in a similar test. However it should be remembered that a component of this value is an uncertainty in the manually extracted position (as the true position was unknown, unlike the ultrasonic tests) and the standard deviation of the camera detection alone was 3.66 cm. The output of the camera system is the world coordinates of the point midway between the ground contact points of the bicycle's wheels, whereas the output of the ultrasonic system was the distance of the cyclists shoulder from the side of the HGV. The translation from the ground point to the shoulder would introduce discrepancies between the camera and ultrasonic systems due to roll motion of the cyclist, and the position and angle of the cyclist's torso relative to the bicycle. However, for the purposes of predicting trajectories rather than merely detecting proximity, the point in the ground plane is the more reliable predictor of future motion, which is a benefit of the camera system.

A significant disadvantage of the camera system compared to the ultrasonic system is the loss of accuracy at close range. However, this could be mitigated by smaller separation between the cameras (possibly increasing the number of cameras required) and also by lowering the camera, to reduce the angle between the camera and the ground plane, thus reducing occlusion of the wheels by the cyclist's torso.

Conclusions and further work

Conclusions

(i) A camera system was been developed to measure the motion of cyclists on the nearside of Heavy Goods Vehicles. The system consisted of two downward-facing cameras mounted high on the side of the vehicle. A calibration grid marked on the ground was used for initial calibration. Cyclist wheels were detected using boosted classifiers and validated using geometrical arguments. The point of contact between the wheel and the ground was extracted and converted into world coordinates using a coordinate mapping generated from the calibration grid.

(ii) The system was evaluated using test data from a number of parallel passing manoeuvres between a cyclist and HGV. The system was generally able to track the position of the cyclist to within 10 cm at distances of 1 m or greater from the HGV. The detection step was accurate to ± 4 cm (standard deviation) at most points. The remainder of the error was introduced by the mapping to world coordinates. At lateral distances of less than 1 m the system was found to be significantly less accurate due to occlusion and distortion of the image features. Quantification of the error was hampered by the lack of a ground truth to compare to.

(iii) The system was slightly less accurate than Jia's ultrasonic system, most significantly when the cyclist was close to the HGV. The camera-based approach also suffers in poor lighting or weather conditions, meaning

a solely camera-based approach is likely unrealistic. However, the camera system addresses many of the limitations of the ultrasonic system, including complexity and cost of installation and the ability to differentiate between multiple cyclists. A hybrid system using cameras to identify cyclists and a few ultrasonic sensors to accurately locate them would be a possible enhancement.

Further work

(i) Additional image features such as helmets or handlebars could be detected and used to validate wheel detections. This would need to be done in image coordinates as the other features are not in the ground plane and so cannot be located in world coordinates without a more complex calibration stage.

(ii) Processing time could be reduced by limiting feature searching to a zone close to the previous detection (with the size of the search zone controlled by the cyclists velocity as tracked by the system). A full image search could be included periodically to detect any new cyclists.

(iii) A robust and efficient implementation of ellipse detection would likely be a more reliable method of locating the ground contact point than the current solution. There is also a need for a way to recognise when the ground contact point is occluded so as to use the alternative method of estimating at a fixed position in the wheel bounding box.

(iv) For the tests described here, lighting conditions were favourable, although there was variation in light intensity. Image normalisation (to intensity and contrast) could help to produce robustness to lighting conditions.

(v) The system does not work at night. The use of night-vision cameras could be investigated. Since the classifiers are based on shape features, they should be adaptable enough to work on night-vision images. Care would need to be taken to shield the cameras from intense lighting such as headlights, which would wash-out the images.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

C. Eddy was supported by the UK Engineering and Physical Sciences Research Council (EPSRC). C.C. de Saxe was supported by the Cambridge Commonwealth, European and International Trust, UK, and the Council for Scientific and Industrial Research (CSIR), South Africa.

Acknowledgements

The authors would like to thank Yanbo Jia for providing test data and assistance, and Max Lowther for input into the vision processing. They would also like to thank Mark Starosolsky of Laing O'Rourke for providing the test vehicle.

The authors would like to acknowledge the members of the Cambridge Vehicle Dynamics Consortium who supported the work in this paper. At the time of writing, the Consortium consisted of the University of Cambridge with the following partners from the heavy-vehicle industry: Anthony Best Dynamics, Brigade Electronics, Denby Transport, Firestone, Goodyear, Haldex, Motor Industry Research Association, SDC Trailers, SIMPACK, Tinsley Bridge, Tridec, Volvo Trucks and Wincanton.

References

1. Road accidents and safety: statistical tables index. Department for Transport, London, 2014. URL https://www.gov.uk/government/statistical-data-sets.
2. Robinson TL and Chislett W. Commercial vehicle safety priorities - ranking of future priorities in the UK - based on detailed analysis of data from 2006-2008. *Project Report, Transportation Research Laboratory* 2010; .

3. Jia Y. *An automated cyclist collision avoidance system for Heavy Goods Vehicles*. Phd thesis, University of Cambridge, 2014.
4. Safety Shield Systems. Cycle Safety Shield - Safety Shield Systems. URL http://safetyshieldsystems.com/cycle-safety-shield/.
5. Fusion Processing. CycleEye — Fusion Processing. URL http://www.fusionproc.com/products/.
6. Jia Y and Cebon D. Field testing of a cyclist collision avoidance system for heavy goods vehicles. *IEEE Transactions on Vehicular Technology* 2016; 65(6): 4359–4367.
7. Ardeshiri T, Larsson F, Gustafsson F et al. Bicycle tracking using ellipse extraction. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*. pp. 1–8.
8. Illingworth J and Kittler J. A survey of the hough transform. *Computer Vision, Graphics, and Image Processing* 1988; 44(1): 87–116.
9. Yuen H, Princen J, Illingworth J et al. Comparative study of Hough Transform methods for circle finding. *Image and Vision Computing* 1990; 8(1): 71–77.
10. Xu L, Oja E and Kultanen P. A new curve detection method: Randomized Hough transform (RHT). *Pattern Recognition Letters* 1990; 11(5): 331–338.
11. McLaughlin RA. Randomized Hough Transform: Improved ellipse detection with comparison. *Pattern Recognition Letters* 1998; 19(3-4): 299–305.
12. Yonghong X and Qiang J. A new efficient ellipse detection method. In *Pattern Recognition, 2002. Proceedings. 16th Inter*, volume 2. IEEE, pp. 957–960.
13. Lai CC and Tsai WH. Estimation of moving vehicle locations using wheel shape information in single 2-D lateral vehicle images by 3-D computer vision techniques. *Robotics and Computer-Integrated Manufacturing* 1999; 15(2): 111–120.
14. Viola P and Jones M. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (CVPR '01)*, volume 1. IEEE Comput. Soc, pp. I–511–I–518.
15. Freund Y and Schapire RE. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* 1997; 55(1): 119–139.
16. Freund Y and Schapire RE. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence* 1999; 14(5): 771–780.
17. Bradski G. The OpenCV Library, 2000.
18. Chavez-Aragon A, Laganiere R and Payeur P. Vision-based detection and labelling of multiple vehicle parts. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 1273–1278.
19. Felzenszwalb P, Girshick R, McAllester D et al. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2010; 32(9): 1627–1645.
20. Bertozzi M, Boggi A, Medici P et al. Stereo Vision-Based Start-Inhibit for Heavy Goods Vehicles. In *2006 IEEE Intelligent Vehicles Symposium*. IEEE, pp. 350–355.
21. Point Grey Research Inc. Flea 3 specifications sheet., 2012.
22. Fujinon. YV2.8x2.8SA02 CCTV lens specifications sheet., 2014.
23. Rosin P and West G. Nonparametric Segmentation of Curves into Various Representations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1995; 17(12): 1140–1153.
24. Mai F, Hung Y, Zhong H et al. A hierarchical approach for fast and robust ellipse extraction. *Pattern Recognition* 2008; 41(8): 2512–2524.
25. Mainzer T. Genetic Algorithm for Shape Detection. Technical report, University of West Bohemia, Pilsen, 2002.
26. Xiao J, Hays J, Ehinger KA et al. SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 3485–3492.

27. Geiger A, Lenz P and Urtasun R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition*.
28. Lowther M. *Identifying Cyclists Using Vision Data*. Master's thesis, University of Cambridge, 2017.
29. Fitzgibbon A, Pilu M and Fisher R. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1999; 21(5): 476–480.
30. Fischler MA and Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 1981; 24(6): 381–395.
31. Winfield D and Parkhurst D. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, volume 3. IEEE, pp. 79–79.
32. Canny J. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1986; PAMI-8(6): 679–698.
33. Claus D and Fitzgibbon A. A Rational Function Lens Distortion Model for General Cameras. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1. IEEE, pp. 213–219.
34. Silva GH, Riche RL, Molimard J et al. Exact and efficient interpolation using finite elements shape functions. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique* 2012; 18(3-4).
35. Ali NH and Hassan GM. Kalman Filter Tracking. *International Journal of Computer Applications* 2014; 89(9).

Appendix

Nomenclature

A	Transformation matrix
c_i	i_{th} coefficient of mapping polynomial in the u direction
$\mathbf{c}$	Vector of c_i coefficients
C	Camera system measurement of cyclist position
d	Nominal lateral distance from the side of the vehicle
d_i	i_{th} coefficient of mapping polynomial in the v direction
$\mathbf{d}$	Vector of d_i coefficients
$\mathbf{K}$	Vector of state Kalman gains
M	Manually measured cyclist position
$\mathbf{p}$	Vector of prior estimates of the state errors
$\mathbf{P}$	Vector of state estimation errors
$\mathbf{Q}$	Vector of state process covariances
$\mathbf{R}$	Vector of state model covariances
T	True cyclist position
$\mathbf{u}$	Vector of u coordinates
$\mathbf{v}$	Vector of v coordinates
X	Longitudinal position relative to the vehicle
$\mathbf{X}$	Vector of bicycle states
$\hat{\mathbf{X}}$	Vector of prior estimates of the states
Y	Lateral position relative to the vehicle
$\mathbf{z}$	Vector of state observations
(u, v)	Location of a point in image coordinates
(x, y)	Location of a point in world coordinates
ε_{ij}	Uncertainty between the measurements i and j
μ_{ij}	Mean error between the measurements i and j
σ_{ij}	Standard deviation of the uncertainty between the measurements i and j