Biologically Inspired Composite Vision System for Multiple Depth-of-field Vehicle Tracking and Speed Detection

Lin Lin, Ramesh Bharath*, and Xiang Cheng

Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576

Abstract. This paper presents a new vision-based traffic monitoring system, which is inspired by the visual structure found in raptors, to provide multiple depth-of-field vision information for vehicle tracking and speed detection. The novelty of this design is the usage of multiple depth-of-field information for tracking expressway vehicles over a longer range, and thus provide accurate speed information for overspeed vehicle detection. A novel speed calculation algorithm was designed for the composite vision information acquired by the system. The calculated speed of the vehicles was found to conform with the real-world driving speed.

1 Introduction

Object tracking has been one of the most attractive topics in computer vision, and it has various practical applications so far, such as human-computer interaction [1, 2], video surveillance [3–5], vehicle navigation [6], traffic monitoring and control [7–9], and motion analysis [10]. In this paper, the focus is on traffic monitoring in expressways for automatic overspeed vehicle detection. In the past few years, various vision-based methods have been designed to solve the traffic monitoring problems using a single [11–14] or stereo camera [15, 16]; however, the performance of the above-mentioned systems is limited by a small tracking range due to the fixed depth-of-field of the cameras [17, 18]. These systems perform vehicle tracking near the installed location, and hence they can only track and calculate vehicle speed within a small distance. Therefore, these systems are better suited for traffic monitoring situations such as congestion control and intersection monitoring. In practice, a long tracking range is crucial when high-speed vehicle monitoring is needed.

In contrast, the traditional speed detection approaches using sensors such as LIDAR/RADAR have several drawbacks [19, 20]. These approaches generally work in this way: the sensors detect the presence of a possible overspeeding vehicle and trigger a camera to capture the image of the overspeeding vehicle. However, with a large amount of vehicles present in the scene, the detector might know one of the vehicles is overspeeding but not be able to single it out. Furthermore, the interference caused by big vehicles leads to unreliable results for speed

* Corresponding author.

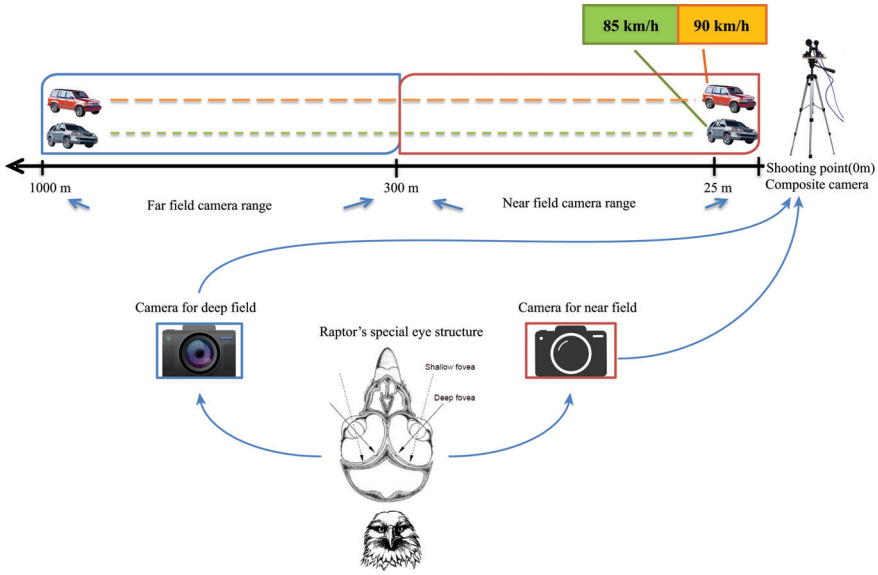


Fig. 1. Overview of the Composite Vision System.

detection. Finally, the lack of coherence between sensor data and visual vehicle identity data seriously cripples law enforcement. Therefore, there is a need for a vision-based traffic monitoring system that can address all the above problems.

Reference [21] proposed a composite image sensor inspired by raptors' vision for deep-field object tracking, which uses multiple depth-of-field information to detect the presence of vehicles over a longer range. Although the object tracking range was increased, only a limited portion of the tracking result was used as a switch to activate another camera for license plate detection. In fact, this activation can be achieved by a single camera or other sensors. In other words, the visual data was not fully utilized, and the object tracking effort over a longer range did not serve a suitable application. In this paper, we propose a speed detection application using the multiple depth-of-field vision information, which utilizes the full tracking range.

As shown in Figure 1, the proposed system uses a composite camera that can view both near field and far field synchronously, and monitor vehicle movements up to 1000 m away from the shooting point. The proposed system defines the tracking result, while vehicles are present in the composite camera's field of view. Later on, each vehicle's speed and identity information are output as a snapshot, which can be easily used for law enforcement. As a result, the drawbacks of present traffic monitoring systems are overcome by providing a long tracking range and coherence between speed information and vehicle identity. In summary, the proposed system provides a new solution to improve the capabilities of vision-based traffic monitoring systems.

The rest of the paper is organized as follows. Sec. 2 presents the composite

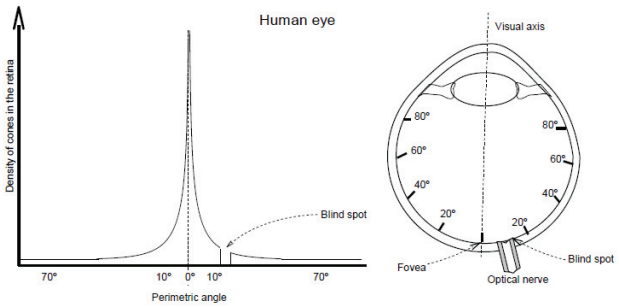


Fig. 2. Primates' retina photoreceptor distribution.

camera design with implementation details; Sec. 3 presents the vehicle tracking algorithm, followed by Sec. 4 with the proposed vehicle speed calculation algorithm, and Sec. 5 with experimental results and discussion. Finally, conclusions and future works are presented in Sec. 6.

2 Composite Camera Design

According to Tucker and Snyder [22, 23], raptors such as falcons and eagles have two sets of foveae, which enables them to simultaneously observe objects from near and far distances, as shown in Figure 1. The visual data obtained by the two sets of foveae are combined into a single nerve and sent to raptor's brain for processing. However, since the study about raptor's foveae to brain information mapping is limited, we adopt the primate retina model and log-polar transform [24] to simulate raptor's internal mapping.

By observing the non-uniform distribution of the cones around primates' fovea, as shown in Figure 2, a logarithmic relationship for information around the fovea structure [25] can be established. Using log-polar mapping, the vision information received in each camera can be transformed into log-polar space regardless of the depth-of-field. Therefore, the problem of combining information from multiple depth-of-field is neatly solved as shown in Figure 3. Furthermore, log-polar transformation (LPT) with ideal center point provides scale and rotation invariance. As a result, the scale change of vehicles caused by forward vehicle movement will be converted to horizontal shifting in the log-polar space with a fixed shape. Hence, the transformed LPT image could possibly provide relatively unchanged vehicle shape during the tracking process.

To have seamless stitching, the composite camera requires two cameras with different view angles. In addition, the two cameras should be placed as close as possible to reduce errors. As shown in Figure 4, these nested cameras should have a special relation to achieve seamless stitching of different depth-of-field information. For ideal log-polar mapping and stitching, the camera relation factor K_b (eqn. 1) is set to be 10. In other words, the two camera lenses should have roughly about 10 times difference in their view angle.

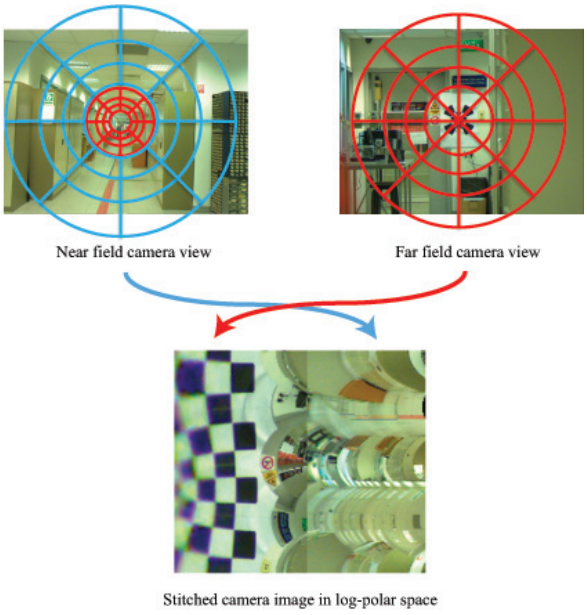


Fig. 3. Composite image stitching example. Best viewed in color.

$$K_b = \frac{\rho_b}{\rho_{max}} = \frac{tan(\theta_{far}/2)}{tan(\theta_{near}/2)} \tag{1}$$

$$K_b \approx \frac{\theta_{far}}{\theta_{near}}, if \theta_{far} \to 0, \theta_{near} \to 0 \tag{2}$$

2.1 Composite Camera Implementation

On the basis of the theoretical design introduced earlier, a market research showed that there is no readily available camera for synchronous viewing with different depth-of-fields and view angles. Therefore, the composite camera was implemented by choosing individual cameras and integrating them in a flexible hardware mount. As indicated in Figure 5, USB 3.0 technology was selected after comparing various industrial standards. Further market survey about USB 3.0 cameras led us to choose Basler cameras, shown in Figure 6(a).

This composite camera design is different from the work of [21], because USB 3.0 cameras provide high frame rate with minimal frame dropping and also support synchronous video acquisition with easy plug-and-play ability with laptops. This provides future development possibilities to easily install the vehicle monitoring system on mobile platforms such as police cars. On the contrary, ref. [21] used ad hoc surveillance cameras that have several disadvantages. Surveillance camera standards such as Gigabit Ethernet/Fire Wire/Camera Link are not so

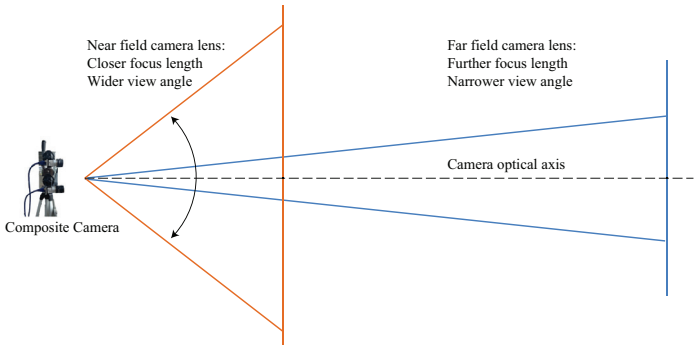


Fig. 4. Dual lens composite camera relationship.

	FireWire	Gigabit Ethernet	USB 2.0	USB 3.0	Camera Link
Bandwidth	80MB/s	100MB/s	40MB/s	440MB/s	680MB/s
Cable length	10m	100m	5m	3m	7m
Consumer acceptance	Declining	Excellent	Excellent	Excellent	None
Multiple cameras	Excellent	Good	Fair	Excellent	Fair
Power delivery	Excellent	Excellent (POE)	Fair	Good	None
Vision Standard	IIDC DCAM	GigE Vision	No	USB3 Vision	Camera Link

Fig. 5. Industrial Camera Standards [26].

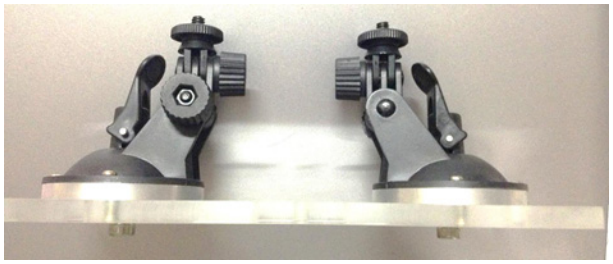
suitable for outdoor vehicle monitoring, because they require a special connection interface, or separate power supply, or separate communication devices for data acquisition. These additional requirements limit the mobility and flexibility of traffic monitoring system using surveillance cameras. Besides, our composite camera design includes a flexible hardware mount, which was designed to adjust the cameras' positional relationship arbitrarily (Figure 6(c)).

Two lenses were carefully selected for the implementation of the composite camera to simulate raptor's vision. One of them is a wide angle camera which provides a 76.7 degree view angle, and the other has a narrow view angle about 7.9 degrees. Therefore, the two of them have approximately 10 times difference in viewing angle. The successfully built composite camera is shown in Figure 6(b). The next objective was to synchronously capture videos from both these cameras.

Our initial trial in using MATLAB to synchronously capture videos failed, because it could not differentiate the two identical cameras with different lenses. Hence, some alternative software solutions were tested. Finally, a C# program with AForge.Net library was implemented to handle synchronized video acquisition from both cameras. After implementing the software, two plausible locations in Singapore - Tampines Expressway and Ayer Rajah Expressway - were selected

(a) Basler cameras.

(b) Composite camera system.



(c) Hardware platform (without cameras).

Fig. 6. Composite Camera built using USB 3.0 industrial cameras.

to capture high-speed vehicle movements. Figure 7 shows the field test conducted on an overhead bridge 4.5 m above the expressway lanes.

2.2 Multiple Depth-of-field Data Processing

The stitching of the multiple depth-of-field images acquired from the composite camera is shown in Figure 8. For the two Cartesian images acquired synchronously from the composite camera, log-polar transformation is applied individually. To achieve the scale and rotation invariant property in log-polar space, the road vanishing point has to be carefully determined, which serves as the central point for the log-polar transformation. After the transformation, there appears an over-sampled region around the inner part of the near view and an under-sampled region around the outer part of the far vision, shown in Figure 8 as the blue areas. To stitch the two views, some duplicated regions in both views are discarded. The final stitched result has the scale and rotation invariant property of log-polar transformation [27], which helps to maintain the size of each vehicle from the shooting point. Moreover, well-separated road lanes assist tracking in particular lane(s) of interest. Notice that the stitched multiple

Fig. 7. Composite Camera Field Test.

depth-of-field view has a significant reduction in image content, which helps to fasten the object tracking process and therefore increase the computational efficiency. As shown in Figure 8, the Cartesian image size of both views is 600 x 900, whereas the stitched log-polar space composite view result is 136 x 509 only.

3 Vehicle Tracking

The seamless stitching from different depth-of-field cameras extended the object tracking range from about 300 m in the traditional vision-based vehicle tracking practice to up to 1000 m. To counter the complicated outdoor conditions, such as illumination change (shadow of cloud/tree/vehicles, brightness change, and reflections) and mechanical vibration, Gaussian mixture model (GMM) was adopted to extract and separate the moving vehicles from the background. Figure 9 shows the vehicle extraction result using GMM algorithm.

With successful object extraction by GMM, Kalman filter [28] was adopted as the object tracking method for multiple vehicles in the scene. The main advantage of using Kalman filter is that it is able to model the vehicle's acceleration in the video due to the prospective projection. Furthermore, it provides tolerance to a certain degree of occlusion by predicting the vehicle position based on previous vehicle states. Finally, Kalman filter provides a distance parameter to tolerate distortion and noise of object movements. Briefly, Kalman filter allows the system to track multiple vehicles while maintaining some prediction and tolerance to the complex vehicle motions, such as lane switching and variable acceleration.

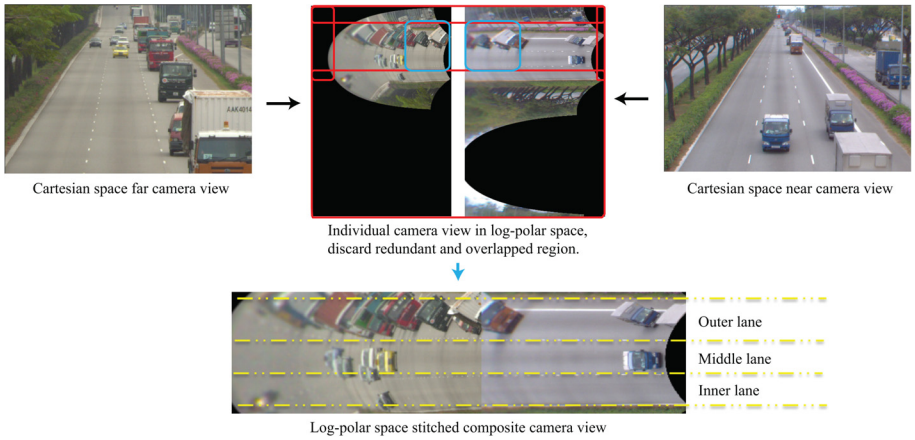


Fig. 8. A typical stitching process of the multiple depth-of-field images.

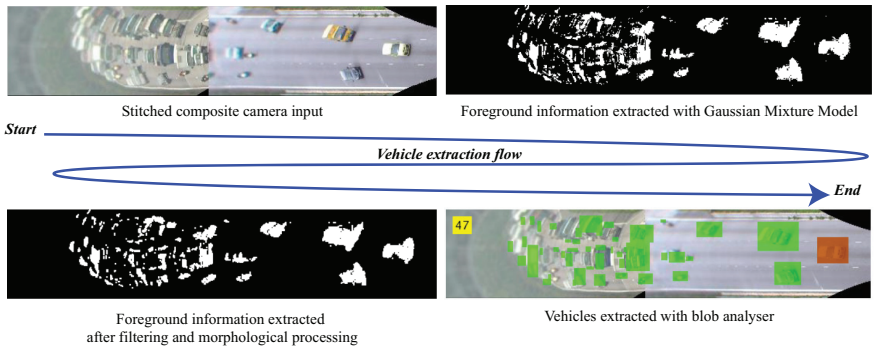


Fig. 9. Vehicle extraction with Gaussian mixture model.

4 Proposed Vehicle Speed Calculation

The most commonly adopted solution for vehicle speed detection is using LIDAR or RADAR signal along with surveillance cameras. One significant drawback is the lack of ability to determine the correct overspeed vehicle, because of insufficient coherence between speed information and the vehicle image saved. For instance, upon activation by the RADAR signal, the surveillance camera could capture an image with more than one vehicle in the scene including the overspeed vehicle. Another defect is that the RADAR/LIDAR accuracy is highly affected by interference from large vehicles. The proposed composite vision system aims to provide a vision-based solution that is able to gather speed information of a vehicle with its corresponding image, and thus overcome the drawbacks in the current speed monitoring systems.

4.1 Vehicle Speed Calculation

Because the vehicles move faster from deep to near field due to perspective projection, direct speed calculation from pixel coordinates is impractical. Therefore, an algorithm is proposed to transform the composite image location to the real-world location, and determine the vehicle speed in kilometer per hour. The vehicle speed calculation involves four main steps that are as follows:

1. Transform stitched log-polar space coordinates to single individual log-polar space coordinates.

2. Transform individual log-polar space coordinates to camera Cartesian space coordinates.

3. Transform camera Cartesian space coordinates to real world coordinates.

4. Use tracking time information and real world location to calculate vehicle speed

Step 1: An arbitrary position (u, v) in the stitched log-polar coordinates can be transformed to the single individual log-polar space coordinate (U,V) by the following relation, where $u_{\mathrm{InnerRingCrop}}$ is the number of rings cropped out, $v_{\mathrm{LowerWedgeCrop}}$ is the number of wedges cropped out, and $u_{\mathrm{stitchline}}$ is the position of the stitch line.

1. When u,v falls in the far-field view range, that is, on the left of the stitching line

$$U = u + u_{InnerRingCrop} \tag{3}$$

$$V = v + v_{LowerWedgeCrop} \tag{4}$$

2. When u,v falls in the near-field view range, that is, on the right of the stitching line

$$U = u + u_{stitchline} + u_{InnerRingCrop} \tag{5}$$

$$V = v + v_{LowerWedgeCrop} \tag{6}$$

Step 2: The method to transform individual log-polar space coordinate (U,V) to their corresponding camera Catesian space coordinate (x,y) is defined as follows:

$$Distance = r_{min} \times e^{\frac{U \times log(\frac{r_{max}}{r_{min}})}{n_r - 1}} \tag{7}$$

$$Angle = V \times \frac{2\pi}{n_w} \tag{8}$$

$$x = Distance \times \cos(Angle) + x_c \tag{9}$$

$$y = Distance \times \sin(Angle) + y_c \tag{10}$$

Where n_r represents the number of rings, n_w is number of wedges, x_c and y_c are chosen road vanishing point position, r_{max} and r_{min} are maximum and minimum radii used in stitching process. The dots in Figure 10 demonstrate successful transformation between the two coordinates.

Step 3: The transformation from camera Cartesian space coordinates to real-world coordinates (x,z) follows the method proposed by Wu [11]. Taking into

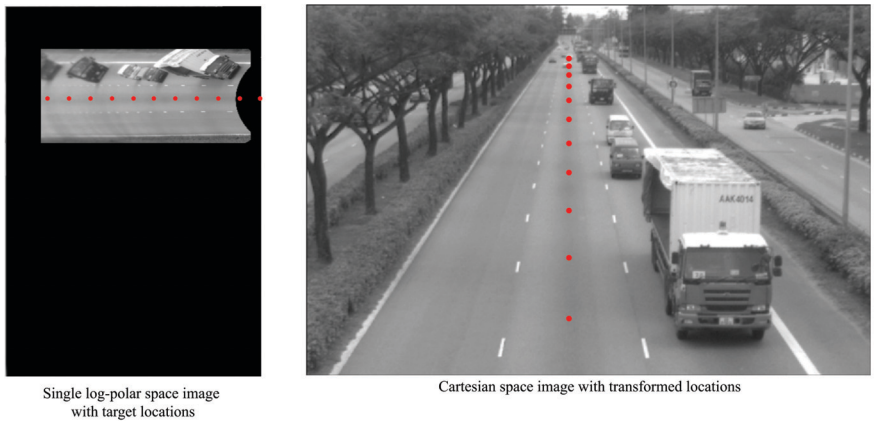


Fig. 10. Speed calculation step 2 demonstration.

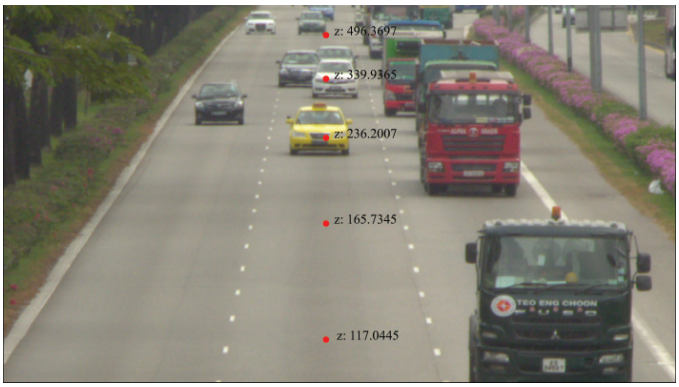


Fig. 11. Speed calculation step 3 demonstration.

consideration the composite cameras height above the road and tilt angle θ from the road's forward direction, the real-world locations x (transverse direction on the road surface) and z (longitudinal or forward direction on the road surface) can be obtained. To validate the transformation result, the lane markers separation distance (12 m) defined by international traffic standard was used. As shown in Figure 11, the estimated distance roughly matches with the real-world location.

Step 4: Finally, with the correct location information, the distance traveled by a vehicle in the real world can be calculated based on the Euclidean distance between the two points of interest. The speed calculation is effectively achieved by using the timing information provided during vehicle tracking process. About 20 calculation windows were used to average out the calculated speed of a vehicle, which is tracked from approximately 1000 m to 25 m away from the shooting point.

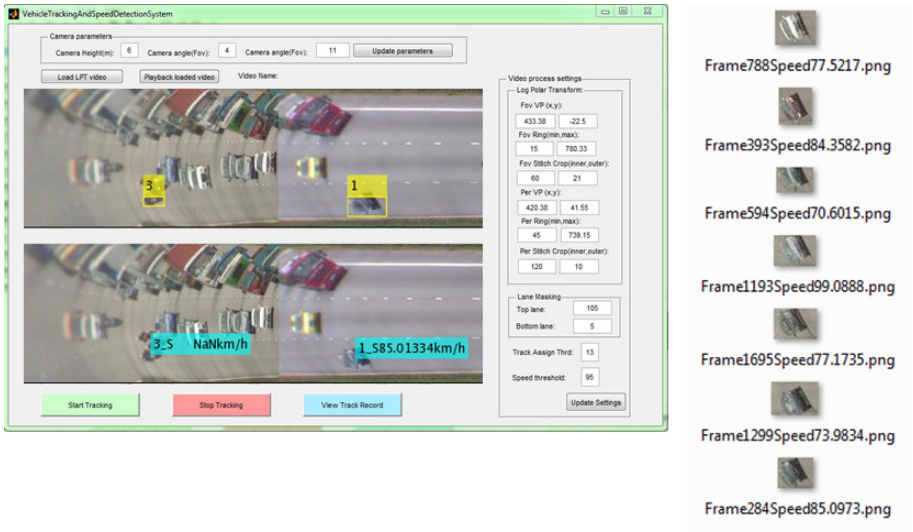


Fig. 12. Results of the composite vision system.

5 Experimental Results and Discussion

5.1 Vehicle Speed Calculation Based on Multiple Depth-of-Field Vision

To test the performance of the composite vision system, we arranged a car to drive with a known speed around 75 km/h and then recorded the video using our composite camera setup. With the recorded video, the composite vision system tracked the speed information with 20 calculation windows and calculated the average vehicle speed to be 78 km/h. This is a reasonable speed detection result, because the car's original speed was slightly varied around 5 km/h due to driving conditions on the expressway.

After establishing the baseline result, we tested the system on a variety of traffic conditions. Figure 12 shows the results from a video captured at Ayer Rajah Expressway in Singapore; the upper half of the MATLAB GUI displays the tracking result while the lower half shows the calculated speed. After each vehicle exits the scene, a snapshot of it is stored along with the time stamp and average speed. Notice that NaN appears as one of the vehicles' speed in Fig.12. This is because the speed calculation algorithm waits for tracking to stabilize, which typically takes a few seconds.

Because the vehicle motion is tracked from far field to near field, the composite camera provides adequate information for the system to perform long distance tracking and speed calculation. This benefits the accuracy and reliability of the speed calculation. With the multiple depth-of-field viewing, the system is able to track the vehicle from approximately 1000 m away to 25 m nearby, which is a big improvement compared to conventional practices up to 300 m

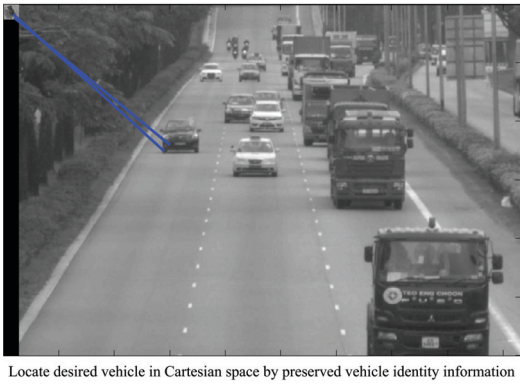


Fig. 13. Example of preserving over-speeding vehicle identity

using vision-based methods. Various parameters need to be considered for the speed calculation procedure, such as road lane vanishing point setting, log-polar transform parameters, and individual camera settings. However, the speed calculation is most affected by the precision of object tracking. From experiments in other parts of the island, we found that the most challenging aspect was low lighting conditions. In these cases, the speed detection result was not reliable due to tracking difficulties.

5.2 Benefits of Vehicle Identity Preservation

As discussed earlier, the main drawback of the present traffic monitoring systems is the lack of vehicle identity information. The composite vision system solves this problem by providing a snapshot of the exact overspeeding vehicle. Figure 12 shows the saving of snapshots during the tracking process. To augment evidence of the overspeeding vehicle, license plate detection with the aid of a high-resolution camera triggered by the composite vision system can be implemented. As a proof-of-concept, we conducted some experiments with the snapshot information to locate the vehicle in the Cartesian image. As shown in Figure 13, the overspeeding vehicle can be precisely located using SIFT keypoint matching, which can possibly help the high-resolution camera take a close-up picture.

6 Conclusion

We proposed a composite vision system with multiple depth-of-field viewing ability that largely extended the tracking range of traditional traffic monitoring systems. By defining the overspeeding vehicle using the tracking result, strong coherence between identity and speed information was established. Having deep field object tracking ability, the composite vision system can handle high-speed

vehicle tracking and can compensate the drawbacks of present speed monitoring systems. Moreover, the system has the potential to do real-time tracking in complex road conditions and multiple lanes. It also has the potential for both airborne unmanned vehicle navigation due to its versatility in being autonomous.

The future work of the proposed system includes the addition of a third camera for license plate detection. Besides, more cameras could be added to extend the tracking distance. Last but not least, the performance and reliability will be further investigated in the future work as well.

Acknowledgement. The authors would like to thank Mr. Jack Chin from Sodavision for helping us gather detailed information about camera devices and their accessories. We would also like to thank the software engineers from Basler Asia service for giving advice and guidance on using the Basler camera SDK and AForge.Net library to develop the programs.

References

1. Rehg, J.M., Kanade, T.: Visual tracking of high dof articulated structures: an application to human hand tracking. In: European Conference on Computer Vision. Springer (1994) 35–46
2. Jacob, R.J., Karn, K.S.: Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. Mind **2** (2003) 4
3. Foresti, G.: Object recognition and tracking for remote video surveillance. IEEE Transactions on Circuits and Systems for Video Technology **9** (1999) 1045–1062
4. Cohen, I., Medioni, G.: Detecting and tracking moving objects for video surveillance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (1999) 325
5. Javed, O., Shah, M.: Tracking and object classification for automated surveillance. In: European Conference on Computer Vision. Springer (2002) 343–357
6. Saripalli, S., Montgomery, J., Sukhatme, G.: Visually guided landing of an unmanned aerial vehicle. IEEE Transactions on Robotics and Automation **19** (2003) 371–380
7. Coifman, B., Beymer, D., McLauchlan, P., Malik, J.: A real-time computer vision system for vehicle tracking and traffic surveillance. Transportation Research Part C: Emerging Technologies **6** (1998) 271–288
8. Kamijo, S., Matsushita, Y., Ikeuchi, K., Sakauchi, M.: Traffic monitoring and accident detection at intersections. IEEE Transactions on Intelligent Transportation Systems **1** (2000) 108–118
9. Tai, J.C., Tseng, S.T., Lin, C.P., Song, K.T.: Real-time image tracking for automatic traffic monitoring and enforcement applications. Image and Vision Computing **22** (2004) 485–501
10. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. Pattern recognition **36** (2003) 585–601
11. Wu, J., Liu, Z., Li, J., Caidong, G., Si, M., Tan, F.: An algorithm for automatic vehicle speed detection using video camera. In: 4th International Conference on Computer Science Education, ICCSE '09. (2009) 193–196
12. Clady, X., Collange, F., Jurie, F., Martinet, P.: Cars detection and tracking with a vision sensor. In: Proceedings of the Intelligent Vehicles Symposium, IEEE (2003) 593–598

13. Roessing, C., Reker, A., Gabb, M., Dietmayer, K., Lensch, H.: Intuitive visualization of vehicle distance, velocity and risk potential in rear-view camera applications. In: Intelligent Vehicles Symposium (IV), IEEE (2013) 579–585
14. Wang, C.C., Thorpe, C., Suppe, A.: Ladar-based detection and tracking of moving objects from a ground vehicle at high speeds. In: Proceedings of the Intelligent Vehicles Symposium, IEEE (2003) 416–421
15. Kormann, B., Neve, A., Klinker, G., Stechele, W., et al.: Stereo vision based vehicle detection. In: VISAPP (2). (2010) 431–438
16. Toulminet, G., Bertozzi, M., Mousset, S., Bensrhair, A., Broggi, A.: Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis. IEEE Transactions on Image Processing **15** (2006) 2364–2375
17. Richard A. Messner, P.B.M.: Mobile digital video system for law enforcement. In: VTC Spring. (2002) 468–472
18. Schoepflin, T.N., Dailey, D.J.: Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. IEEE Transactions on Intelligent Transportation Systems **4** (2003) 90–98
19. Simmoneau, B.: I-team: Controversy over speed cameras. News (2013)
20. Solicitors, M.D.: Guide to speed detection devices (2014)
21. Melnyk, P.: Biologically inspired composite image sensor for deep field target tracking. PhD thesis, University of New Hampshire (2008)
22. Tucker, V.A.: The deep fovea, sideways vision and spiral flight paths in raptors. Journal of Experimental Biology **203** (2000) 3745–3754
23. Snyder, A.W., Miller, W.H.: Telephoto lens system of falconiform eyes. Nature **275** (1978) 127 – 129
24. Messner, R.A., Szu, H.H.: An image processing architecture for real time generation of scale and rotation invariant patterns. Computer vision, graphics, and image processing **31** (1985) 50–66
25. Schwartz, E.L.: Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. Biological cybernetics **25** (1977) 181–194
26. Richmond, B.: A Practical Guide to USB 3.0 for Vision applications. Point Grey Research, Inc, 12051 Riverside Way. (2013)
27. Bailey, J.G., Messner, R.A.: Log-polar mapping as a preprocessing stage for an image tracking system. In: Robotics Conferences, International Society for Optics and Photonics (1989) 15–22
28. Brown, R.G., Hwang, P.Y., et al.: Introduction to random signals and applied Kalman filtering. Volume 3. John Wiley & Sons New York (1992)