IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 5, 2019, accepted July 27, 2019, date of publication July 30, 2019, date of current version August 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932120

Vehicle Speed Measurement Based on Binocular Stereovision System

LEI YANG[1], (Member, IEEE), MENGLONG LI[1], XIAOWEI SONG[1], (Member, IEEE), ZIXIANG XIONG[2], (Fellow, IEEE), CHUNPING HOU[3], AND BOYANG QU[1]

[1]School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou 450007, China
[2]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA
[3]School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Corresponding author: Lei Yang (annieyanglei@163.com)

This work was supported in part by the Science and Technology Innovation Team of Colleges and Universities in Henan Province (18IRTSTHN013), in part by the National Natural Science Foundation of China under Grant 61440031 and Grant 60902063, in part by the Foundation and Advanced Technology Research Project of Henan Province under Grant 152300410132, and in part by the Key Research Project of Colleges and Universities in Henan Province under Grant 19A510005.

ABSTRACT In this paper, we propose a novel stereovision-based system for vehicle speed measurement. Our system is set in a fixed location to capture two view stereo videos for passing vehicles by using a calibrated binocular stereovision system. An important feature of this system is the use of an optimized single shot multibox detector network that can efficiently detect license plates in the captured two view stereo videos. Then, the system performs only on the detected license plate area to extract stereo matching point pairs rapidly. The world coordinates of 3D points corresponding to each stereo matching point pair are calculated. Abnormal 3D points are removed in accordance with distance distribution. The remaining 3D point closest to the license plate center is finally selected as the exact vehicle location of the current frame. Vehicle speed is measured by dividing the distance between two selected 3D points by frame intervals. Vehicle trajectory is depicted by connecting multiple selected 3D points. Our proposed system performs non-intrusive and stealth measurement in intelligent traffic surveillance and overcomes the challenges of speed measurement for multiple vehicles on multiple lanes in different motions simultaneously. Experimental results demonstrate its accuracy. Our system has a speed measurement error range of $[-1.6, +1.1]$ km/h and a maximum error rate of 3.80%, which are within the $[-3, +2]$ km/h error limit of several countries' regulatory authorities and the 6% error rate limit of China national standard GB/T 21255-2007.

INDEX TERMS Vehicle speed measurement, vehicle feature detection, SSD network, stereovision, stereo matching.

I. INTRODUCTION

With the rapid development of intelligent transportation system (ITS), ITS has provided intelligent traffic surveillance, transportation management, dynamic information service, and vehicle control [1]–[4]. Vehicle speed measurement plays an important role in intelligent traffic surveillance. At present, the most commonly used speed measurement methods are divided into two main categories: intrusive and non-intrusive [5]–[7]. Intrusive sensors, such as inductive loop detectors (ILDs), are widely used. The time interval during which a vehicle passes through a certain length between ILDs is recorded to calculate the ave-rage speed of the vehicle [8]. However, the installation and maintenance of ILDs are complex and may damage road surfaces. Non-intrusive sensors, such as Radar and Lidar, are also widely used. For Radar, an electromagnetic wave is transmitted, and the frequency shift in the received signal is calculated [9]. For Lidar, a laser is transmitted, and the time of flight of the reflected signal is calculated [10]. Both sensors have high measurement accuracy but have detectable energy emission. When vehicles detect the energy emission, it will intentionally slow down before entering the speed measurement range. This action is undesirable for stealth measurement. Moreover, these sensors require high cost and frequent maintenance. Despite the popularity of ILDs, Radar, and Lidar in traffic surveillance, three major problems still exist and need to be solved.

The associate editor coordinating the review of this manuscript and approving it for publication was Kang Li.

1) Synchronization of vehicle speed measurement and vehicle information (e.g., license plate number, and vehicle color, brand, and mode) identification.
2) Simultaneous speed measurement of multiple vehicles.
3) Speed measurement of vehicles in curved motion.

The technological advancement in vision system has reduced the cost of cameras and increased the resolution of the captured images. Video-based vehicle speed measurement has received increasing attention [11]–[19] because it is non-intrusive, suitable for stealth measurement, and low cost. It can also provide vehicle speed measurement and information identification in the same video simultaneously. The methods in [11], [12], [14] and [17] are monocular video-based vehicle speed measurement ones. Those in [11], [12] and [14] are feature-based methods that identify vehicles from the video frames by their features. The method in [17] is a motion-based one that locates vehicles with optical flow. These monocular methods calculate the time interval of the vehicle passing through a distance and thus measure the average speed of that vehicle. The 2018 AI City Challenge Workshop in CVPR 2018 has a challenge track in 3D speed estimation [13]. The winning method [14] can solve the problem of multi-vehicle speed measurement simultaneously with the lowest RMSE. However, it uses perspective projection and requires vehicles to be in straight line motion, with a constant speed, and be parallel to the lane under test. The methods in [18] and [19] are stereo video-based vehicle speed measurement ones. The method in [18] is a feature-based one that identifies the vehicle from the stereo video frames, calculates the depth, and measures the speed. This method can solve the problem of speed measurement of vehicles in curved motion within structured environments. However, the stereo visual tracking is achieved by particle filtering, in which the manual setup of vehicle feature is required. One time of manual setup is needed for one feature, while different vehicles have varying features. Therefore, if multiple vehicle speeds need to be measured, then multiple times of manual setup are required. Thus, this method lacks intelligence. The method in [19] is a motion-based one that subtracts a static background from the video frame to detect the vehicle, calculates the depth, and measures the speed. However, this method has difficulty distinguishing between vehicles and other moving objects. Therefore, it also lacks intelligence.

In this paper, we propose a novel stereovision-based system for vehicle speed measurement. Fig. 1 shows the overall view of our proposed system. The proposed system integrates vehicle feature tracking with stereo motion estimation for accurate and intelligent vehicle speed measurement. The input left and right view videos are captured by a fixed bino-cular stereovision camera. An optimized single shot multibox detector (SSD) network, which is referred to as license plate detection (LPD)-SSD (LPD-SSD), is used to extract the typical feature (license plate) of a vehicle synchronously from the stereo videos. Vehicle tracking in the monocular video and stereo matching in the binocular stereo videos are only carried out within the detected license plate area. The actual spatial position of the vehicle is calculated in accordance with the calibrated parameters of the stereovision system and the disparity of vehicle in the stereo videos. The vehicle speed is finally calculated between two consecutive video frames. Our proposed system allows the same video to be used for speed measurement and license plate identification. It also allows speed measurement of multiple vehicles on multiple lanes and of vehicles in straight line or curved motion intelligently. However, this system is only installed in a fixed location instead of in a vehicle at present, which may be a future target of our project.

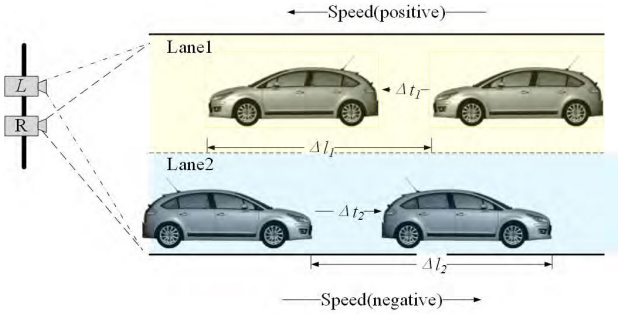


FIGURE 1. Overall view of the proposed vehicle speed measurement system based on binocular stereovision system.

The rest of the paper is divided into four parts. In Section II, we discuss the related works. In Section III, we present our proposed system, which includes system configuration, vehicle feature detection, vehicle tracking and stereo matching, and speed measurement. In Section IV, we report experimental procedures and results. In Section V, we make a conclusion.

II. RELATED WORKS

Accurate vehicle feature detection in video frames is a prerequisite for video-based vehicle speed measurement. License plate has a regular appearance, uniform contour, and relatively rich texture details and can thus be easily detected. License plate is also unique and suitable for the synchronization of speed measurement and information identification in the future. Many existing traffic surveillance systems record license plates of vehicles in traffic violations, and the infrastructure is already available [11]. Although the license plate is not attached to the ground plane, the 3D spatial position of the license plate center can be directly calculated using our binocular stereovision system. The relative 3D displacements can be obtained. Accordingly, accurate vehicle speed can be derived. Therefore, we choose license plate as the object to be detected in our system.

Traditional video object detection methods include background modeling [20], [21], frame difference [22], [23], and optical flow [24], [25]. The two former methods are based on video processing. They are suitable for detecting moving object with fixed background. However, license plate is always a part of a moving vehicle and cannot be detected separately. Meanwhile, the latter method can track objects with fixed or moving background. It can also track license plate as

a separate part but cannot detect license plate without manual help. Morphological methods are widely used in image object detection by relying on color, edge, shape, and texture attributes extracted with the Canny detector [26], Sobel operator [27], template-matching [28], conditional random field [29], or wavelet [30]. In [29], an edge operator is used to extract the vertical edge of license plate. In [30], edge density information is utilized to detect license plate. In [31], a rectangular sliding window is adopted to detect image regions of high gradient density, such as license plate. All these morphological methods are cumbersome, time consuming, and unsuitable for LPD in complex background.

Modern object detection methods have developed with the progress of convolutional neural network (CNN) [32]. Since AlexNet [33] won ILSVRC 2012 by a notable margin in 2012, object detection methods with CNN have received considerable attention. CNN-based object detection methods can detect various objects accurately and intelligently with the trained networks and models. The R-CNN series (e.g., R-CNN, Fast-RCNN, and Faster-RCNN) algorithms are typical examples of two-stage algorithms [34], [35]. They have high accuracy but need long computation time. You only look once (YOLO) [36] and SSD [37] are typical examples of one-stage algorithms. They are considerably faster than the two-stage algorithms. For example, the detection speed of SSD in VOC2007 dataset can reach 59 frames per second (FPS), whereas that of Faster R-CNN can only reach 7 FPS [37]. All these CNN-based object detection methods are suitable for LPD. In [38], a CNN network is trained and fine-tuned to detect license plate. In [39], a CNN-based MD-YOLO framework for multidirectional LPD is proposed. In [40], a region-based convolutional neural network is trained to detect license plate. However, we build LPD in our system on the basis of the framework of one-stage SSD algorithm to meet the system requirements on efficiency and accuracy fully.

III. PROPOSED SYSTEM

In the proposed vehicle speed measurement system, we use two POINT GRAY FL2G-C industrial cameras to build the stereovision system, as shown in Fig. 2. The camera has a Sony ICX445 CCD, with a 1/3 inch size, a 1288×964 maximum resolution, 1.25 M effective pixels, a 32 MB onboard buffer, and a 512 kB data flash memory. The stereovision system also includes a laptop, with a Core i7 CPU, 8 G RAM, a 2 G Nvidia Geforce 830 M discrete graphics card, and a 1 T solid-state hard disk. The laptop controls the cameras and communicates with the cameras via a USB3.0 interface. We use Zhengyou Zhang's camera calibration method [41] to calibrate our stereovision system.

The entire procedure of the proposed system is shown in Fig. 3. System input includes left and right view videos captured by the calibrated binocular stereovision system. System output includes vehicle speed and trajectory. This system consists of three main parts: vehicle feature detection, vehicle tracking and stereo matching, speed and trajectory

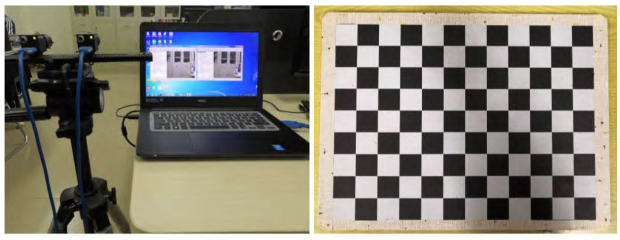


FIGURE 2. Binocular stereovision system. (a) Capture system. (b) Calibration board.

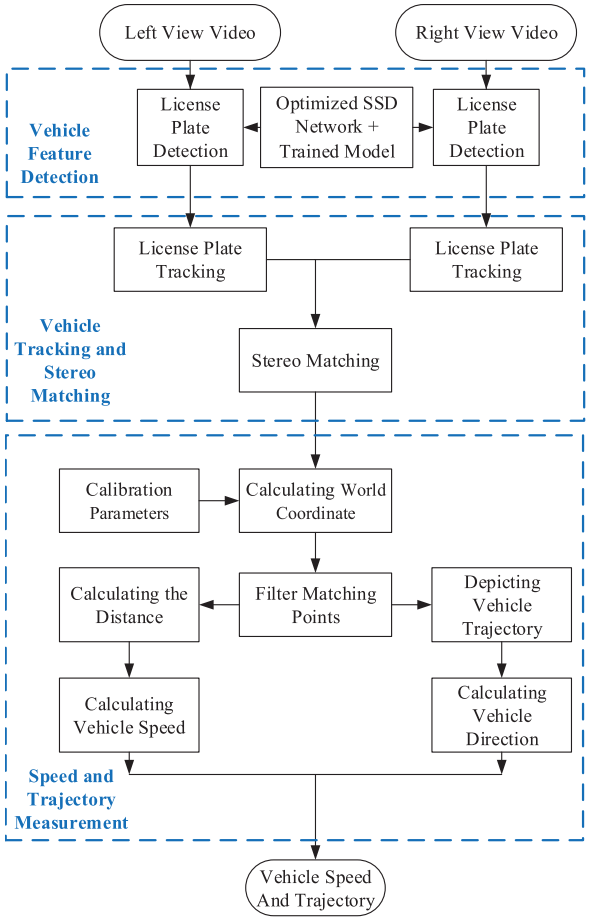


FIGURE 3. The whole procedure of the proposed vehicle speed measurement system.

measurement. In the vehicle feature detection part, an LPD model is trained on the basis of an optimized SSD network. The trained model is then used to detect all license plates in the left and right view videos. In the vehicle tracking and stereo matching part, the license plate areas in consecutive frames of the monocular video are matched to ensure independent tracking for each license plate. Then, the license plate areas in the frame pairs of the stereo videos are matched to extract several stereo matching point pairs. In the speed and trajectory measurement part, the world coordinates of the real

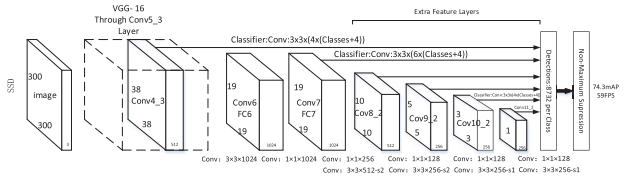


FIGURE 4. SSD network structure.

TABLE 1. Parameters of SSD Layers (pixels).

Convolution Layer		Feature Layer
Layer	Receptive Field Size	Output Size
Conv4_3	92×92	38×38
Conv7	260×260	19×19
Conv8_2	292×292	10×10
Conv9_2	356×356	5×5
Conv10_2	485×485	3×3
Conv11_2	612×612	1×1

3D point corresponding to each stereo matching point pair are calculated using the calibration parameters of the camera. The distance between each real 3D point and the origin is computed. In accordance with the distance distribution, the abnormal points are removed first. Then, the point closest to the license plate center is selected as the exact 3D location of the current frame pairs. With the world coordinates of the selected points, the distance traveled by a vehicle in a certain period of time (i.e., one frame interval or several frame intervals) is computed, and the trajectory is depicted. The vehicle speed can be calculated by dividing the distance by the time, and the vehicle movement direction can be derived.

A. VEHICLE FEATURE DETECTION

We first detect the vehicle feature in stereo video frames. As described in the related works, we choose license plate as the vehicle feature to be detected and select SSD as the network structure to detect license plate. We also improve the existing SSD network structure in accordance with the statistical distribution of license plate size in some traffic surveillance datasets. Fig. 4 shows the existing SSD network structure. Table 1 shows the receptive field size of each convolution layer and the output size of each feature layer.

We analyze 9850 traffic surveillance images in BIT-Vehicle dataset [42], with image resolution 1920×1080, to design the most suitable network structure for vehicle LPD in traffic video. The license plate sizes are statistically classified in accordance with the SSD receptive field size, as shown in Fig. 5. The statistical distribution bar graph in Fig. 5 indicates that the license plate sizes in traffic surveillance videos are all in the resolution range below 260×260 pixels. Specifically, the number of license plate sizes below 92×92 pixels accounts for 58.8%, the number of license plate sizes between 92×92 pixels and 260×260 pixels accounts for 41.2%, and the number of license plate sizes above 260×260 pixels accounts for 0%. Among them, the number of license plate sizes above 150×150 pixels only accounts for 0.17%, and the

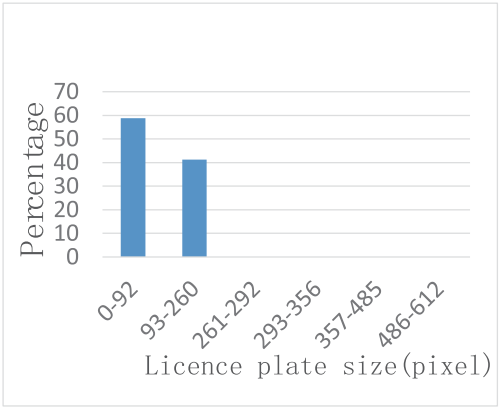


FIGURE 5. Statistical distribution of license plate sizes.

FIGURE 6. Image with the maximum license plate size in BIT-Vehicle dataset.

largest license plate size is 172×57 pixels, which is captured at the closest position of the surveillance camera, as shown in Fig. 6. Even if the resolution reaches 4K(4096×2160), that is, the length and the width are expanded 2.13 times, the size of license plate will still be 367×122 pixels and will not exceed the receptive field size of conv10_2, that is, 485×485 pixels.

Therefore, we optimize the SSD network structure in accordance with the statistical distribution of license plate sizes by removing the conv11_2 layer and changing channel number in the last layer to 2. For simplicity, we call this optimized network as LPD-SSD. The LPD-SSD network structure is shown in Fig. 7. The LPD-SSD network can increase the model training and detection speed and reduce the model data amount and can thus enhance the network efficiency.

We randomly select 7000 images from three vehicle datasets to build our license plate training and testing sets. We include 1098 images selected from the CCPD dataset (250000 images with resolution 720×1160) constructed by the University of Science and Technology of China [43], 4500 images selected from the BIT-Vehicle dataset (9850 images with resolution 1920×1080) established by University of Science and Technology of

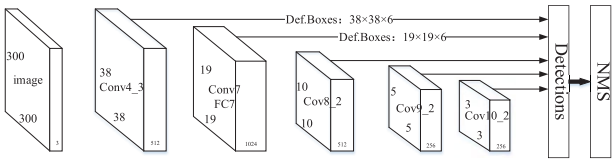


FIGURE 7. LPD-SSD network structure.

FIGURE 8. Image examples of our dataset.

Beijing [42], and all 1402 images from an open license plate dataset (resolution 1600×1200) provided by the OpenITS research project [44] sponsored by Guangdong Key Laboratory of Intelligent Transportation System. In addition, 800 images (resolution 4032×2268) with high resolution and image quality are taken by Nikon d3200 SLR camera in parking lots and streets at a distance of 1-15 m. Meanwhile, 1000 images (resolution 1920×1080) are extracted from the campus security monitoring system of Zhongyuan University of Technology, in which most vehicles are far from the cameras and the license plate sizes are very small. In summary, 8800 images are obtained. With random image clipping and scaling, the dataset size is expanded by a factor of three, that is, 26400 images. Fig. 8 shows some image examples of our license plate dataset.

We train the LPD model with half of the dataset, that is, 13200 images, as the training set in the SSD and LPD-SSD network. The other half of the dataset are the testing set in the training process. The testing set contains license plates shot from various perspectives. After 23000 steps of training, we obtain a trained SSD-based model and a trained LPD-SSD-based model. The mean average precision (mAP) of the SSD-based model reaches 0.971. The mAP of the LPD-SSD-based model reaches 0.968, which is 0.3% lower than that of SSD. As shown in Table 2, the loss function of the SSD-based model converges to 1.45. The loss function of the LPD-SSD-based model converges to 1.25, which is 0.2 lower than that of SSD. Therefore, in the training process, the detection accuracy of the LPD-SSD-based model is nearly the same as that of the SSD-based model.

We select other 600 images, which are not used in the training process, from BIT-Vehicle dataset as the validation

TABLE 2. Convergence value of loss function.

SSD		LPD-SSD	
Step	Convergence Value	Step	Convergence Value
22605	1.760966	22606	1.000497
22640	1.744639	22643	1.577889
22677	1.9097	22680	1.17034
22714	1.073728	22714	1.343024
22750	0.837441	22751	0.757618
22787	1.365314	22794	1.616417
22824	1.418205	22834	1.554723
22860	1.26742	22875	1.405214
22895	1.717385	22916	1.014247
22933	1.420003	22956	1.079638
MeanValue	1.45	MeanValue	1.25

TABLE 3. Detection accuracy and speed.

Network	Accuracy	Speed
SSD	98%	5FPS
LPD-SSD	97.8%	5.4FPS

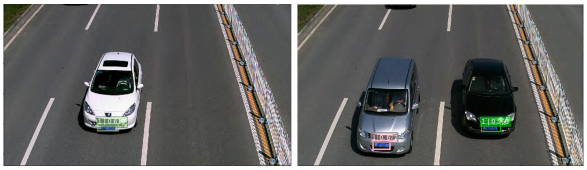


FIGURE 9. License plate detection effect.

set to validate the detection accuracy of the two models further. As shown in Table 3, the SSD-based model achieves a detection accuracy rate of 98.0%. The LPD-SSD-based model achieves a detection accuracy rate of 97.8%, which is 0.2% lower than that of the SSD in the validation process. Both have good detection ability for license plate in surveillance videos. The detection effect of LPD-SSD is shown in Fig. 9.

However, the detection speed of SSD and LPD-SSD differs. On a local server, equipped with two 6-Core Intel E5-2620 v3 @ 2.40 GHz CPUs, 32 G RAM, an 8 G Nvidia Geforce GTX 1080 independent graphics card, and a 1 T solid-state hard disk, the detection speeds of SSD and LPD-SSD are 5 and 5.4 FPS, as shown in Table 3. The speed of LPD-SSD is increased by 8%. The detection efficiency of the LPD-SSD-based model is improved, whereas the detection accuracy remains unchanged. Thus, we choose the LPD-SSD network with the trained LPD model to detect license plates in our vehicle speed measurement system. Notably, the processing speed will increase as the CPU and GPU processing power increases. The position and distance of the camera to the vehicle will also matter, and the SSD network structure can be adjusted in accordance with the environmental situation.

Blurring may occasionally occur in the captured stereo video frames, during which license plate cannot be detected. However, vehicle speed can be measured as long as two pairs of video frames contain clear license plates in the speed measurement range. Therefore, we only keep the frame pairs

FIGURE 10. License plate detected at a distance of 15 m.

FIGURE 11. License plate detected at a distance of 1 m.

with clear license plates and skip those with blurred license plates. At least 10 frame pairs can be captured for the vehicle speed of 120 km/h with a measurement range of 10 m and a frame rate of 30 FPS. These frame pairs are sufficient to remove the effect of occasional blurring. At least five frame pairs can be captured even for the vehicle speed of 240 km/h, which is a prohibited speed in traffic. These frame pairs are also sufficient to remove the effect of occasional blurring.

Fig. 10 shows the detected license plate in a pair of stereo video frames at a distance of 15 m from the vehicle to the camera. The detected license plate is small but detectable. If the vehicle is more than 15 m away from the camera, then the license plate will be too small to be detected or lack of details to perform the following stereo matching. Fig. 11 shows the detected license plate in a pair of stereo video frames at a distance of 1 m from the vehicle to the camera. If the vehicle is less than 1 m away from the camera, then the license plate will be easily blurred because of the fast displacement at a close range. By experimental verification, the system can achieve very high LPD accuracy in the speed measurement range of 1-15 m. The resolution of license plates in this speed measurement range is sufficiently high for detection and matching in our system. Therefore, we conduct vehicle speed and trajectory measurement in the speed measurement range of 1-15 m in the following experiments.

B. VEHICLE TRACKING AND STEREO MATCHING

Multiple vehicles may appear simultaneously on the road. Thus, multiple license plates may be detected in one frame in the preceding vehicle feature detection process. Each license plate should be precisely tracked in the consecutive frames of the monocular video to measure the speed of each vehicle separately. In addition, license plates in the frame pairs of the stereo videos should be matched to calculate the spatial position of the vehicle at a certain time. Therefore,

FIGURE 12. Tracking result of the same license plate in consecutive frames of the monocular video.

FIGURE 13. Stereo matching result of the same license plate in frame pairs of stereo videos.

the license plates in the stereo videos should be accurately matched.

In essence, vehicle tracking and stereo matching are the same because they both belong to matching problem between two images. For vehicle tracking, the scale of the license plate varies with the distance between the vehicle and the camera. Therefore, the matching algorithm should be scale invariant. For stereo matching, the translation and rotation of the license plate vary with the range and angle of the stereovision camera. Thus, the matching algorithm should be translation and rotation invariant. The matching algorithms that meet the aforementioned requirements include scale-invariant feature transform (SIFT) [45], speed up robust features (SURF) [46], locality preserving matching [47], [48], and vector field consensus matching [49]. In our system, we use SURF to perform license plate tracking in consecutive frames of the monocular video. We also use SURF to perform feature points matching on the license plate area in the frame pairs of the stereo videos. As shown in Fig. 12, the license plate in the left image is used as the matching template, and the two detected license plates in the right image are matched separately. The tracking result is accurately located on the right license plate. As shown in Fig. 13, the license plates in frame pairs of the stereo videos are matched. Matching point pairs are extracted, and a homography matrix is used to remove most mismatching point pairs for the first time [50], [51]. The remaining matching point pairs are used for speed and trajectory measurement in the following part. We limit the SURF operation only on the detected license plate area to save processing time.

C. SPEED AND TRAJECTORY MEASUREMENT

Although we perform SURF operation only on the detected license plate area, we still extract many matching point pairs. In our experiments, the number of matching point pairs is approximately 60 at the farthest distance (i.e., 15 m) and is nearly 140 at the nearest distance (i.e., 1 m). We further reduce the number of the extracted matching point pairs by narrowing the license plate area to a small circular area to increase the measurement efficiency. We choose a circular

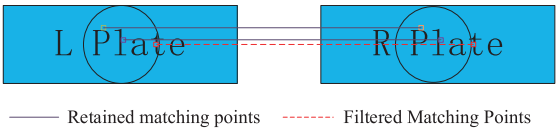


FIGURE 14. Filter of matching point pairs.

area because it is easy to calculate. We take the center of the detected license plate area as the center because it is easy and quick to locate. We take the height of the detected license plate area as the diameter because the circular area with this diameter covers nearly a quarter of the total license plate area. The matching point pairs with left point in left circle and right point in right circle are retained, and the remaining matching point pairs are abandoned. As shown in Fig. 14, we can retain two pairs of matching points connected by solid lines and remove another pair of matching points connected by dashed lines. With this circle, we can effectively reduce the computational complexity by at least a quarter. We can reduce the number of matching point pairs to approximately 10 at the farthest distance (i.e., 15 m) and nearly 30 at the nearest distance (i.e., 1 m). Such number is suitable for the following error elimination process. If the diameter becomes small, then the number of matching point pairs at the farthest distance of 15 m will be too small to perform the statistical analysis in the following error elimination process and thus affect accuracy.

MISMATCH-POINT-PAIRS-FILTER(A)

for $i = 1$ **to** n
  $matched\_pairs(i) = left\_point(i) + right\_point(i)$
  if $left\_point(i)$ *not in left circle area*
    $matched\_pair(i)$ *is abandoned*
  else If $right\_point(i)$ *not in right circle area*
    $matched\_pair(i)$ *is abandoned*
  $PL = Remaining\ matched\_pairs\ in\ P$
  $\lambda = Deviation\ value\ of\ distance\ between$
    $camera\ and\ point(i)\ in\ PL$
  if $\lambda > 20$
    $matched\_pair(i)$ *is abandoned*
$Q = Remaining\ matched\_pairs\ in\ PL$

After matching point pairs are filtered, the world coordinates of the real 3D points corresponding to each retained stereo matching point pairs are calculated using the calibration parameters of the stereovision camera. In our system, a convergent binocular stereovision model [52] is used, that is, the optical axes of the left and right cameras intersect at a point in front of the cameras, as shown in Fig. 15. With the left camera as the main camera, the external parameters of the system, such as the relative displacement $T = (l,m,n)^T$, and the relative rotation $T = (\alpha,\beta,\gamma)^T$, and the internal parameters of the two cameras, such as focal length, principal point cc, and distortion kc, can be obtained by Zhengyou Zhang's camera calibration method. As shown in Fig. 16, the world coordinates $p_i(x_i,y_i,z_i)$ of a 3D point corresponding to a matching point pair can be calculated with the internal

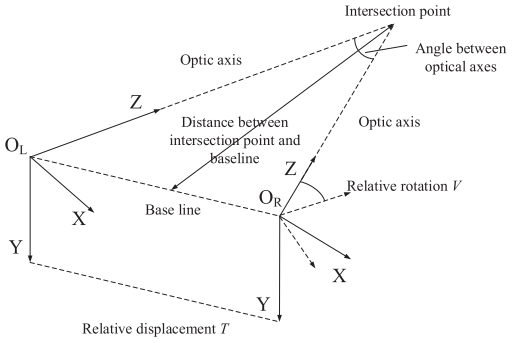


FIGURE 15. Geometry of a convergent binocular stereovision system.

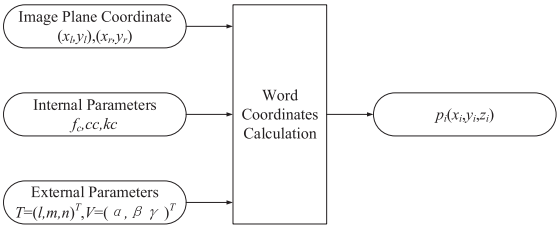


FIGURE 16. Schematic diagram of world coordinates calculation.

parameters, the external parameters, and the 2D coordinates of the matching point pair in the left and right image planes, that is, (x_l,y_l) and (x_r,y_r) [53].

The CCD plane center of the left camera is assumed to be the origin of the coordinate system. Thus, the distance d_i from the target 3D point $p_i(x_i,y_i,z_i)$ to the camera can be expressed as

$$d_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{1}$$

The relative distance d_{ij} between two 3D points $p_i(x_i,y_i,z_i)$ and $p_j(x_j,y_j,z_j)$ can be calculated as

$$d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \tag{2}$$

To eliminate the stereo measurement error in the stereo matching further, we calculate the mean value μ and the standard deviation σ of the distance d_i for all retained 3D points and compute the z_score Z_i of each d_i as follows:

$$\mu = \frac{\sum_{i=1}^{N} d_i}{N} \tag{3}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \mu)^2} \tag{4}$$

$$Z_i = \frac{(d_i - \mu)}{\sigma} \tag{5}$$

Table 4 shows some exemplar distance measurement results of 3D points. From Z_i in Table 4, the z_score of point 5 reaches -2.756. On this basis, the distance of point 5 has a large deviation and should be discarded to increase the measurement accuracy. We eliminate the points with the absolute value of Z_i greater than 1. We select the world

TABLE 4. Exemplar distance measurement results of 3D points.

point	$d_i(mm)$	Z_i
1	8936	0.326
2	8927	0.241
3	8996	0.9
4	8948	0.44.
5	8612	-2.756
6	8898	-0.035
7	8916	0.136
8	8933	0.298
9	8912	0.098
10	8939	0.355
	μ= 8901.7	σ=105.1

coordinates of the point closest to the license plate center in the remaining points as the exact spatial location of the target vehicle in the current stereo frame pair.

As analyzed above, the speed measurement range of our system is 1-15 m. We then assume that the stereo cameras shoot at 10 FPS and each pair of stereo frames are not blurred. For correct measurement, the maximum distance traveled by a vehicle in the time interval $\Delta t = 1/framerate = 1/10 = 0.1$ s between two consecutive frames should not exceed $\Delta l = 15\text{-}1 = 14$ m. Accordingly, our system can theoretically measure a maximum vehicle speed of $V_{max} = 504$ km/h

$$V_{\max} = \frac{\Delta l}{\Delta t} = \frac{14\ m}{0.1\ s} = 140\ \text{m/s} = 504\ \text{km/h} \tag{6}$$

That is sufficiently high for all traffic vehicles. Generally speaking, N pairs of stereo frames would be captured depending on the vehicle speed:

$$N = \left[\frac{\Delta l}{V\Delta t}\right] + 1 = \left[\frac{140}{V}\right] + 1 \tag{7}$$

where N represents the number of video frames that can be captured for a vehicle passing through the speed measurement range and V represents the vehicle speed (m/s). For example, if a vehicle travels at a speed of 100 km/h, that is, 27.8 m/s, then the system can capture six pairs of stereo frames. Then, we can calculate the vehicle speed V in accordance with two spatial positions $p_k(x_k,y_k,z_k)$ and $p_{k+1}(x_{k+1},y_{k+1},z_{k+1})$ of the vehicle and the time interval Δt between the corresponding consecutive frames:

$$V = \frac{\sqrt{(x_{k+1}-x_k)^2 + (y_{k+1}-y_k)^2 + (z_{k+1}-z_k)^2}}{\Delta t} \tag{8}$$

The 3D vehicle trajectory can be depicted using the spatial locations of $p_{k-1}(x_{k-1},\ y_{k-1},\ z_{k-1})$, $p_k(x_k,y_k,z_k)$ and $p_{k+1}(x_{k+1},y_{k+1},z_{k+1})$ of the vehicle, as shown in Fig. 17.

For simplicity, we project the 3D vehicle trajectory onto the ground plane (XOY plane) to observe the vehicle movement direction. As shown in Fig. 18, vectors $\vec{a}$ and $\vec{b}$ represent the

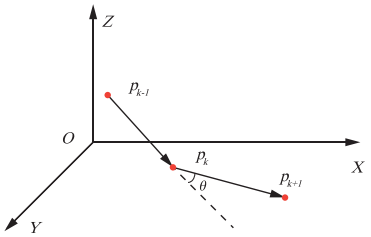


FIGURE 17. 3D vehicle trajectory.

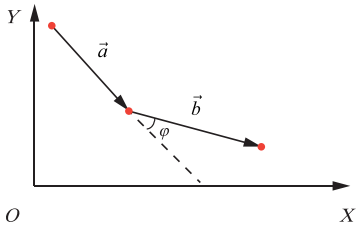


FIGURE 18. 2D projection of a 3D vehicle trajectory.

FIGURE 19. P-Gear P-510 professional satellite speed meter.

moving vectors in XOY plane of vehicle traveling from point p_{k-1} to point p_k and from point p_k to point p_{k+1}, respectively:

$$\begin{aligned}\vec{a} &= (x_a, y_a) = (x_k - x_{k-1}, y_k - y_{k-1}) \\ \vec{b} &= (x_b, y_b) = (x_{k+1} - x_k, y_{k+1} - y_k)\end{aligned} \tag{9}$$

According to the vector cross product formula, we can calculate the steering angle φ of the vehicle (10)–(12), as shown at the bottom of the next page:

If $\varphi = 0$, then the vehicle travels in straight line motion. If $\varphi > 0$, then the vehicle turns left. If $\varphi < 0$, then the vehicle turns right.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

In practical vehicle measurement test, we adjust the focal length of the two POINT GRAY FL2G-C industrial cameras. In this manner, these cameras can shoot the license plate clearly in the measurement range of 1-15 m. The cameras shoot at a relatively slow speed of 10 FPS to avoid blurring. That is, we can perform vehicle speed measurement 10 times per second. For comparison, we take the speed measured by a

FIGURE 20. Single vehicle speed measurement experiments. (a) Straight line motion with a constant speed. (b) Curved motion with a constant speed. (c) Straight line motion with a changing speed.

P-Gear P-510 professional satellite speed meter in the vehicle as the ground truth. This speed meter utilizes the data of the GPS+GLONASS dual satellite positioning system to measure speed and exchanges results with mobile phone 10 times per second by Bluetooth 4.0. The measurement error is 2%. The real-time speed and trajectory of the vehicle are recorded and displayed using a mobile app, as shown in Fig. 19.

We set up five vehicle speed measurement scenarios, which can be divided into two groups: GROUP I - single-vehicle speed measurement experiments and GROUP II - multi-vehicle speed measurement experiments. In GROUP I, we measure the speed of a single vehicle in three motions: straight line motion with a constant speed (Fig. 20(a)), curved motion with a constant speed (Fig. 20(b)), and straight line motion with a changing speed (Fig. 20(c)).

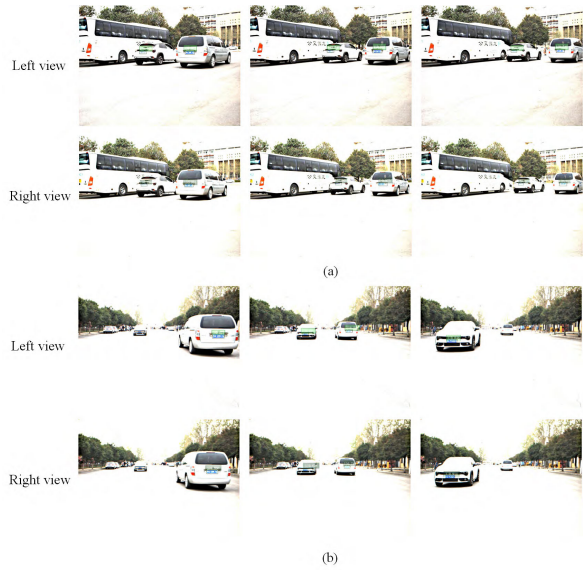


FIGURE 21. Multi-vehicle speed measurement experiments. (a) Multiple vehicles in the same direction. (b) Multiple vehicles in opposite directions.

In GROUP II, we measure the speeds of two vehicles traveling at a constant speed in the same direction (Fig. 21(a)) and opposite directions (Fig. 21(b)), separately. The maximum speed in our experiments is no more than 50 km/h due to the speed limit on campus. The exemplary images of the five vehicle speed measurement scenarios are shown in Fig. 20 and Fig. 21.

B. SINGLE VEHICLE SPEED MEASUREMENT EXPERIMENTS

We measure the speed of a single vehicle in three different motions. We set up the speed measurement system on the right side of the road by keeping an angle of 15° from the road. However, we can select other suitable system locations if we perform an accurate calibration.

Fig. 22 shows the speed measurement graphic results of a single vehicle in three scenarios. Fig. 22(a) shows the result of a straight line motion with a constant speed of 30 km/h. Fig. 22(b) shows the result of a curved motion with a constant speed of 30 km/h. Fig. 22(c) shows the result of a straight line motion with a changing speed from 40 km/h to 50 km/h. The blue line with cross represents the ground truth speed measured by satellite. The red line with circle represents the speed measured by our proposed system and fits the blue line with cross very well.

$$\vec{a} \times \vec{b} = \begin{vmatrix} x_a & y_a \\ x_b & y_b \end{vmatrix} = x_a \cdot y_b - x_b \cdot y_a = |\vec{a}| \cdot |\vec{b}| \cdot \sin\varphi \tag{10}$$

$$\sin\varphi = \frac{x_a \cdot y_b - x_b \cdot y_a}{|\vec{a}| \cdot |\vec{b}|} = \frac{(x_k - x_{k-1}) \cdot (y_{k+1} - y_k) - (x_{k+1} - x_k) \cdot (y_k - y_{k-1})}{\sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2} \cdot \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2}} = C \tag{11}$$

$$\varphi = \arcsin C \tag{12}$$

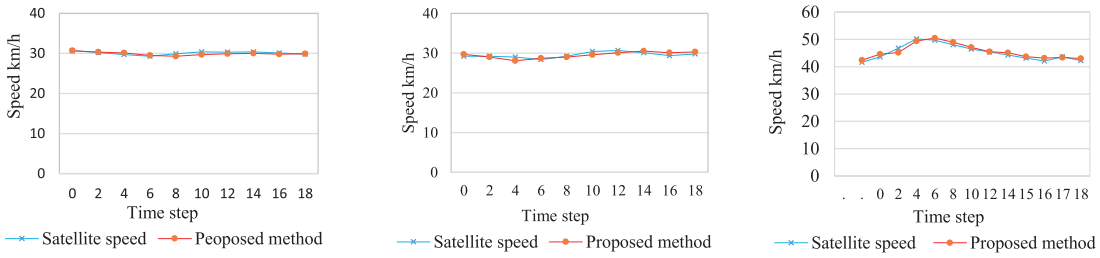


FIGURE 22. Speed measurement graphic results of GROUP I. (a) Straight line motion with a constant speed. (b) Curved motion with a constant speed. (c) Straight line motion with a changing speed.

TABLE 5. Speed values of GROUP I(a).

Time Step	Satellite Speed(km/h)	Proposed (km/h)	Error (km/h)	Error rate (%)
0	30.7	30.7	0	0
2	30.2	30.3	0.1	0.33
4	29.7	30.1	0.4	1.35
6	29.3	29.5	0.2	0.68
8	29.9	29.3	-0.6	-2.01
10	30.4	29.7	-0.7	-2.3
12	30.3	29.9	-0.4	-1.32
14	30.4	30	-0.4	-1.32
16	30.1	29.8	-0.3	-1.00
18	29.8	29.9	0.1	0.34

TABLE 6. Speed values of GROUP I(b).

Time Step	Satellite Speed(km/h)	Proposed (km/h)	Error (km/h)	Error rate (%)
0	29.2	29.7	0.5	1.71
2	29.2	29	-0.2	-0.68
4	29	28.1	-0.9	-3.10
6	28.4	28.7	0.3	1.06
8	29.2	29	-0.2	-0.68
10	30.4	29.6	-0.8	-2.63
12	30.7	30.1	-0.6	-1.95
14	30.1	30.5	0.4	1.33
16	29.4	30.1	0.7	2.38
18	29.8	30.3	0.5	1.68

TABLE 7. Speed values of GROUP I(c).

Time Step	Satellite Speed(km/h)	Proposed (km/h)	Error (km/h)	Error rate (%)
0	43.6	44.6	1	2.29
2	46.8	45.2	-1.6	-3.42
4	50.2	49.4	-0.8	-1.59
6	49.7	50.5	0.8	1.61
8	48.1	49	0.9	1.87
10	46.5	47.1	0.6	1.29
12	45.4	45.5	0.1	0.22
14	44.3	45.1	0.8	1.81
16	42.1	43.2	1.1	2.51
18	42.3	43	0.7	1.65

TABLE 8. Steering angles of GROUP I(b).

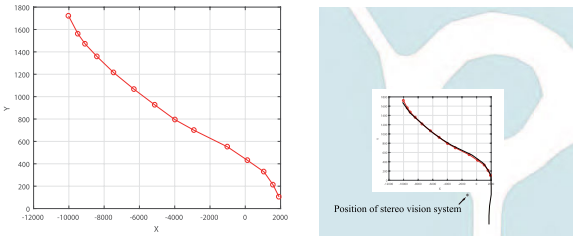
Number	1	2	3	4	5
Angle	4.301°	2.937°	0.881°	1.285°	0.567°
Number	6	7	8	9	10
Angle	0.201°	1.574°	0.525°	−1.581°	−0.24°
Number	11	12			
Angle	−6.405°	−5.099°			

The specific speed values measured in GROUP I(a), (b), and (c) are shown in Table 5, Table 6, and Table 7, respectively. We also calculate the error and the error rate. The maximum error and error rate of GROUP I(a), which appear in the time step 10, are −0.7 km/h and −2.30%, respectively. The maximum error and error rate of GROUP I(b), which appear in the time step 4, are −0.9 km/h and −3.10%. The maximum error and error rate of GROUP I(c), which appear in the time step 2, are −1.6 km/h and −3.42%. In GROUP I, the maximum error is −1.6 km/h, and the maximum error rate is −3.42%; correspondingly, the speed measured by satellite is 46.8 km/h, whereas the speed measured by our proposed system is 45.2 km/h.

The steering angles measured in Group I(b) are shown in Table 8. The trajectory is shown in Fig. 23(a). The trajectory comparison is shown in Fig. 23(b). The results show that the vehicle trajectory measured by our system is highly consistent with the vehicle trajectory recorded by the satellite speed meter.

C. MULTI-VEHICLE SPEED MEASUREMENT EXPERIMENTS

We measure the speeds of multiple vehicles on multiple lanes and in different directions simultaneously. In the multi-vehicle speed measurement experiments, we set up the speed measurement system on the right side of the road by keeping an angle of 15° from the road for vehicles traveling in the same direction. We set up the speed measurement system in the middle of the road by keeping it parallel to the road for vehicles traveling in opposite directions. We can select other suitable system locations if precise calibration is performed.



(a) (b)

FIGURE 23. Trajectory comparison of single vehicle in a curved motion with a constant speed. (a) Trajectory of our proposed system. (b) Trajectory comparison with P-Gear P-510.

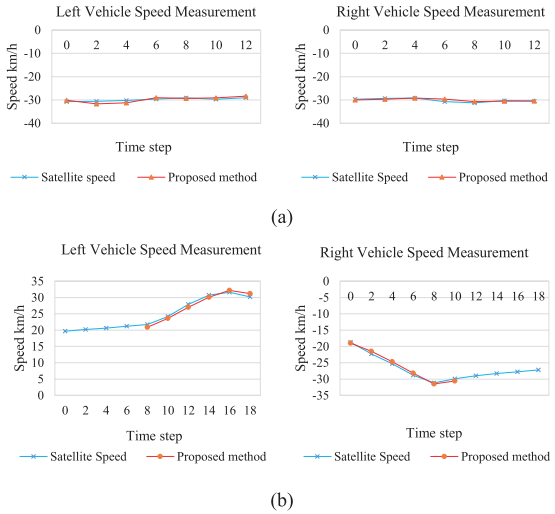


FIGURE 24. Speed measurement graphic results of GROUP II. (a) Two vehicles in the same direction. (b) Two vehicles in opposite directions.

In accordance with the distance between the vehicle and the camera, we set the speed to positive value when the vehicle is coming near and to negative value when the vehicle is moving away. Fig. 24 shows the speed measurement graphic results of multiple vehicles in two scenarios. Fig. 24(a) shows the result of two vehicles moving away in the same direction and both in straight line motions with constant speeds of 30 km/h. Fig. 24(b) shows the result of two vehicles traveling in opposite directions. The left SUV comes near in a straight line motion with a changing speed from 20 km/h to 30 km/h, whereas the right MPV moves away in a straight line motion with a changing speed from 20 km/h to 30 km/h. The blue line with cross represents the ground truth speed measured by satellite. The red line with circle represents the speed measured by our proposed system and fits the blue line with cross very well. However, Fig. 24(b) shows that the speed cannot be measured when the vehicles are far and not within the measurement range of 1-15 m. For the left SUV, the speed can only be measured from time step 8 to 18. For the right MPV, the speed can only be measured from time step 0 to 10. However, this condition does not affect system performance because the speed of each vehicle is measured separately.

TABLE 9. Speed values of GROUP II(a).

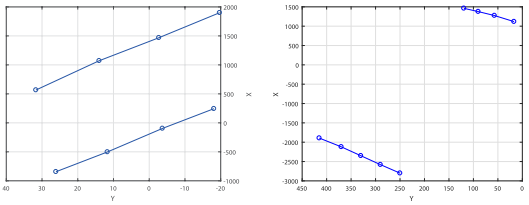
Time Step	Left SUV/ Right MPV			
	Satellite Speed (km/h)	Proposed (km/h)	Error (km/h)	Error rate(%)
0	-30.7/-29.7	-30.1/-30	-0.6/0.3	-1.95 / 1.01
2	-30.6/-29.3	-31.7/-29.7	1.1/0.4	3.80 / 1.37
4	-30.2/-29.2	-31.2/-29.2	1.0/0	3.31 / 0
6	-29.6/-30.7	-29/-29.6	-0.6/-1.1	-2.02 / -3.58
8	-29.1/-31.2	-29.3/-30.7	0.2/-0.5	0.69 / -1.60
10	-29.7/-30.3	-29.1/-30.5	-0.6/0.2	-2.02 / 0.66
12	-29.0/-30.5	-28.4/-30.4	-0.6/-0.1	-2.07 / -0.33

TABLE 10. Speed values of GROUP II(b).

Time Step	Left SUV/ Right MPV			
	Satellite Speed (km/h)	Proposed (km/h)	Error (km/h)	Error rate(%)
0	19.7 / -18.7	NA / -19	NA / 0.3	NA / 1.60
2	20.2 / -22.3	NA / -21.5	NA / -0.8	NA /-3.59
4	20.6 / -25.3	NA / -24.7	NA / -0.6	NA /-2.37
6	21.2 / -28.8	NA / -28.2	NA / -0.6	NA /-2.08
8	21.7 / -31.2	20.9 / -31.5	-0.8 / 0.3	-3.69 / 0.96
10	24.2 / -29.9	23.6 / -30.6	-0.6 / 0.7	-2.48 / 2.34
12	27.9 / -29	27 / NA	-0.9 / NA	-3.22 / NA
14	30.7 / -28.3	30.1 / NA	-0.6 /NA	-1.95 / NA
16	31.6 / -27.8	32.2 / NA	0.6 / NA	1.90 / NA
18	30.2 / -27.2	31.2 / NA	1.0 / NA	3.31 / NA

The specific speed values measured in GROUP II(a) and (b) are shown in Table 9 and Table 10, respectively. We also calculate the error and the error rate. The maximum error and error rate for the left SUV of GROUP II (a), which appear in the time step 2, are 1.1 km/h and 3.80%. The maximum error and error rate for the right MPV of GROUP II (a), which appear in the time step 6, are -1.1 km/h and -3.58%. The maximum error and error rate for the left SUV of GROUP II (b), which appear in the time step 18 and 8, are -1.0 km/h and -3.69%. The maximum error and error rate for the right MPV of GROUP II (b), which appear in the time step 2, are -0.8 km/h and -3.59%. The maximum error for the left SUV of GROUP II is 1.1 km/h, and the maximum error rate is 3.80%; correspondingly, the speed measured by satellite is -30.6 km/h, whereas the speed measured by our proposed system is -31.7 km/h. The maximum error for the right MPV of GROUP II is -0.8 km/h, and the maximum error rate is -3.59%; correspondingly, the speed measured by satellite is -22.3 km/h, whereas the speed measured by our proposed system is -21.5 km/h. Fig. 25(a) shows the trajectories of two vehicles in the same direction. Fig. 25(b) shows the trajectories of two vehicles in opposite directions. Each vehicle keeps driving with a trajectory parallel to each other.

Overall, the measured speed has a maximum negative error of -1.6 km/h and a maximum positive error of $+1.1$ km/h,



(a) (b)

FIGURE 25. Trajectories of multi-vehicle experiments. (a) Two vehicles in the same direction. (b) Two vehicles in opposite directions.

TABLE 11. Error comparison for vehicle speed measurement.

	Mean Error (km/h)	RMSE(km/h)	Max Error (km/h)
Luvizon et al. [11]	-0.5	1.36	[-4.68,+6.00]
Tang et al. [14]	NA	6.59	NA
VSS-SURF [19]	NA	1.29	[-2.0, +2.0]
PROPOSED SYSTEM	-0.02	0.65	[-1.6,+1.1]

which are within the [−3 km/h, +2 km/h] error interval decided by the regulatory authorities of many countries [11]. Furthermore, the measured speed has a maximum error rate of 3.80%, which is within the 6% error rate determined by the China national standard GB/T 21255-2007 of motor vehicle speed meter [54].

We compare the speed measurement error with three other video-based vehicle speed measurement methods, namely, Luvizon et al.'s method [11], Tang et al.'s method [14], and VSS-SURF [19], as shown in Table 11. The error mean of our proposed system is −0.02 km/h, which is less than the error mean of [11] with −0.5 km/h. The RMSE of our proposed system is 0.65 km/h, which is also less than the RMSEs of [11], [14], and [19] with 1.36, 6.59, and 1.29 km/h. The maximum error of our proposed system is [−1.6, +1.1] km/h, which is less than that of [11] with [−4.68, +6.00] km/h and that of [19] with [−2.0, +2.0] km/h. In summary, our system not only can measure the speed of multi-vehicle on multi-lane in different motions simultaneously but also can achieve better measurement accuracy than other methods.

V. CONCLUSION

In this study, we solved the problem of intelligently measuring vehicle speeds on the basis of videos captured by the binocular stereovision system. We presented a system composed of vehicle feature detection, vehicle tracking and stereo matching, speed and trajectory measurement. We integrated the vehicle feature detection within the SSD framework to propose a novel LPD-SSD network particularly for accurate and intelligent LPD. Several matching point pairs were obtained by vehicle tracking in consecutive frames of the monocular video and stereo matching in frame pairs of stereo videos. Real 3D points calculated from matching point pairs were filtered in accordance with the distance distribution, and the best ones representing the vehicle positions were chosen. The distance was computed, and the trajectory was depicted. The speed was calculated, and the direction was derived. The speed measurement performance of our proposed system was verified by the experiments of single- and multi-vehicle speed measurement in five different scenarios. The results show that our method is efficient and reliable in the practical application of vehicle speed measurement. In our experiments, the measured speeds have an error range of [−1.6, +1.1] km/h and a maximum error rate of 3.80%, which are within the [−3, +2] km/h error limit of several countries' regulatory authorities and the 6% error rate limit of China national standard GB/T 21255-2007. In particular, our method is non-intrusive, suitable for stealth measurement, and low cost. It can measure the speeds of multiple vehicles on multiple lanes in the same direction or in different directions simultaneously. It can also measure the speed and trajectory of the vehicle in straight line or curved motion. It deals with a speed measurement range of 10 m where distance estimation errors are in the order of centimeters. Moreover, the method has no strict restrictions on system location if appropriately calibrated. In the future work, we intend to combine vehicle speed measurement with vehicle information identification of license plate in the same video at the same time. We also aim to train a compound vehicle feature detection model to identify a considerable amount of feature information adaptively if the license plate information is unavailable or inadequate.

REFERENCES

[1] A. Auer *et al.*, "History of intelligent transportation systems," U.S. Dept. Transp., Intell. Transp. Syst. Joint Program Office, Tech. Rep. FHWA-JPO-16-329, 2016.

[2] Z. Halim, R. Kalsoom, S. Bashir, and G. Abbas, "Artificial intelligence techniques for driving safety and vehicle crash prediction," *Artif. Intell. Rev.*, vol. 46, no. 3, pp. 351–387, 2016.

[3] K. M. Alam, M. Saini, and A. E. Saddik, "Toward social Internet of vehicles: Concept, architecture, and applications," *IEEE Access*, vol. 3, pp. 343–357, Mar. 2015.

[4] O. Kaiwartya, A. H. Abdullah, Y. Cao, A. Altameem, M. Prasad, C.-T. Lin, and X. Liu, "Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects," *IEEE Access*, vol. 4, pp. 5356–5373, 2016.

[5] E. Minge, J. Kotzenmacher, and S. Peterson, "Evaluation of non-intrusive technologies for traffic detection," Minnesota Dept. Transp., Res. Services Sect., Saint Paul, MN, USA, Tech. Rep. MN/RC 2010-36, 2010.

[6] J. Wu, H. Xu, Y. Zheng, and Z. Tian, "A novel method of vehicle-pedestrian near-crash identification with roadside LiDAR data," *Accident Anal. Prevention*, vol. 121, pp. 238–249, Dec. 2018.

[7] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[8] J. Han, J. W. Polak, J. Barria, and R. Krishnan, "On the estimation of space-mean-speed from inductive loop detector data," *Transp. Planning Technol.*, vol. 33, no. 1, pp. 91–104, 2010.

[9] S.-L. Jeng, W.-H. Chieng, and H.-P. Lu, "Estimating speed using a side-looking single-radar vehicle detector," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 607–614, Apr. 2014.

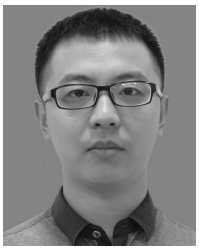
[10] X. Mao, D. Inoue, S. Kato, and M. Kagami, "Amplitude-modulated laser radar for range and speed measurement in car applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 408–413, Mar. 2012.

[11] D. C. Luvizon, B. T. Nassu, and R. Minetto, "A video-based system for vehicle speed measurement in urban roadways," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1393–1404, Jun. 2016.

[12] D. F. Llorca, C. Salinas, M. Jiménez, I. Parra, A. G. Morcillo, R. Izquierdo, J. Lorenzo, and M. A. Sotelo, "Two-camera based accurate vehicle speed measurement using average speed at a fixed point," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 2533–2538.

[13] M. Naphade, M.-C. Chang, A. Sharma, D. C. Anastasiu, V. Jagarlamudi, P. Chakraborty, T. Huang, S. Wang, M.-Y. Liu, R. Chellappa, J.-N. Hwang, and S. Lyu, "The 2018 NVIDIA AI City challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 53–60.
[14] Z. Tang, G. Wang, H. Xiao, A. Zheng, and J.-N. Hwang, "Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 108–115.
[15] Z. Tang, G. Wang, T. Liu, Y.-G. Lee, A. Jahn, X. Liu, X. He, and J.-N. Hwang, "Multiple-kernel based vehicle tracking using 3D deformable model and camera self-calibration," 2017, *arXiv:1708.06831*. [Online]. Available: https://arxiv.org/abs/1708.06831
[16] Z. Tang, M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D. Anastasiu, and J.-N. Hwang, "CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification," 2019, *arXiv:1903.09254*. [Online]. Available: https://arxiv.org/abs/1903.09254
[17] J. Lan, J. Li, G. Hu, B. Ran, and L. Wang, "Vehicle speed measurement based on gray constraint optical flow algorithm," *Opt.-Int. J. Light Electron Opt.*, vol. 125, no. 1, pp. 289–295, Jan. 2014.
[18] J. Zhu, L. Yuan, Y. F. Zheng, and R. L. Ewing, "Stereo visual tracking within structured environments for measuring vehicle speed," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 10, pp. 1471–1484, Oct. 2012.
[19] A. E. Bouziady, R. O. H. Thami, M. Ghogho, O. Bourja, and S. E. Fkihi, "Vehicle speed estimation using extracted surf features from stereo images," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, 2018, pp. 1–6.
[20] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 827–832, May 2005.
[21] A. Tourani, A. Shahbahrami, A. Akoushideh, S. Khazaee, and C. Y. Suen, "Motion-based vehicle speed measurement for intelligent transportation systems," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 4, pp. 42–54, 2019.
[22] H. Min, S. Huazhong, L. Qian, X. Yongquan, and C. Gang, "Moving object detection method based on NMI features motion detection frame difference," *Adv. Sci. Lett.*, vol. 6, no. 1, pp. 477–480, 2012.
[23] C. Zhan, X. Duan, S. Xu, Z. Song, and M. Luo, "An improved moving object detection algorithm based on frame difference and edge detection," in *Proc. 4th Int. Conf. Image Graph. (ICIG)*, Aug. 2007, pp. 519–523.
[24] D. F. P. Bouthemy and C. Kervrann, "Optical flow modeling and computation: A survey," *Comput. Vis. Image Understand.*, vol. 134, pp. 1–21, May 2015.
[25] A. Agarwal, S. Gupta, and D. K. Singh, "Review of optical flow technique for moving object detection," in *Proc. 2nd Int. Conf. Contemp. Comput. Informat. (IC3I)*, 2016, pp. 409–413.
[26] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2963–2970.
[27] M. Garg and S. Goel, "Real-time license plate recognition and speed estimation from video sequences," *ITSI Trans. Elect. Electron. Eng.*, vol. 1, no. 5, pp. 1–4, 2013.
[28] A. H. Ashtari, M. J. Nordin, and M. Fathy, "An iranian license plate recognition system based on color features," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1690–1705, Aug. 2014.
[29] B. Li, B. Tian, Y. Li, and D. Wen, "Component-based license plate detection using conditional random field model," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1690–1699, Dec. 2013.
[30] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 147–156, Jan. 2000.
[31] D. Zheng, Y. Zhao, and J. Wang, "An efficient method of license plate location," *Pattern Recognit. Lett.*, vol. 26, no. 15, pp. 2431–2438, 2005.
[32] I. Ševo and A. Avramović, "Convolutional neural network based automatic object detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 740–744, May 2016.
[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.
[36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.
[37] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
[38] S. Z. Masood, G. Shu, A. Dehghan, and E. G. Ortiz, "License plate detection and recognition using deeply learned convolutional neural networks," 2017, *arXiv:1703.07330*. [Online]. Available: https://arxiv.org/abs/1703.07330
[39] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 507–517, Feb. 2018.
[40] M. A. Rafique, W. Pedrycz, and M. Jeon, "Vehicle license plate detection using region-based convolutional neural networks," *Soft Comput.*, vol. 22, no. 19, pp. 6429–6440, 2018.
[41] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. ICCV*, vol. 99, Sep. 1999, pp. 666–673.
[42] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015.
[43] Z. Xu, W. Yang, A. Meng, N. Lu, H. Huang, C. Ying, and L. Huang, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 255–271.
[44] Guangdong Key Laboratory of Intelligent Transportation System. *Open-Its Research Project*. Accessed: Aug. 6, 2019. [Online]. Available: http://www.openits.cn/
[45] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Comput. Vis. Image Understand.*, vol. 113, no. 3, pp. 345–352, Mar. 2009.
[46] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[47] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.
[48] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.
[49] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.
[50] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, Aug. 2019.
[51] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.
[52] X. Song, L. Yang, Y. Wu, and Z. Liu, "A new depth measuring method for stereo camera based on converted relative extrinsic parameters," in *Proc. Int. Symp. Photoelectron. Detection Imag. Sensors Appl.*, vol. 8908, 2013, p. 890826.
[53] Y. Wu, "Research of traffic surveillance algorithm based on two-view stereo vision," Ph.D. dissertation, Zhongyuan Univ. Technol., Zhengzhou, China, 2014.
[54] *Motor Vehicle Speed Detector*, China National Standard GB/T 21255-2007, 2007.

LEI YANG (M'09) received the Ph.D. degree in communication and information system from Tianjin University, Tianjin, China, in 2007. She was a Visiting Researcher with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, from July 2017 to July 2018. She is currently an Associate Professor with the School of Electronics and Information, Zhongyuan University of Technology, Zhengzhou, China. Her research interests include 3D image processing, 3D display, and computer vision.



MENGLONG LI received the B.S. degree in electronic and information engineering from the Information Engineering University of PLA, Zhengzhou, China, in 2011. He is currently pursuing the M.S. degree with the School of Electronics and Information, Zhongyuan University of Technology, Zhengzhou, China. His research interests include 3D image processing, stereo vision, and computer vision.

XIAOWEI SONG (M'09) received the Ph.D. degree in signal and information processing from Tianjin University, Tianjin, China, in 2007. He was a Visiting Researcher with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, from 2012 to 2013. He is currently a Professor with the School of Electronics and Information, Zhongyuan University of Technology, Zhengzhou, China. His research interests include 3D image and video processing, 3D display, and computer vision.

ZIXIANG XIONG received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, in 1996.

Since 1999, he has been with the Department of Electrical and Computer Engineering, Texas A&M University, where he is currently a Professor and the Associate Department Head. His main research interests lie in image/video processing, computer vision, virtual/augmented reality, big data, and communications. He received the NSF Career Award, in 1999, the ARO Young Investigator Award, in 2000, and the ONR Young Investigator Award, in 2001. He is a co-recipient of the *2006 IEEE Signal Processing Magazine* Best Paper Award, the top 10% paper awards at the 2011 and 2015 IEEE Multimedia Signal Processing Workshops, an IBM Best Student Paper Award at the 2016 IEEE International Conference on Pattern Recognition, and the Best Demo Paper Award at the 2018 IEEE International Conference on Multimedia and Expo. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (1999–2005), the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002–2005), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002–2006), the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (Part-B) (2005–2009), and the IEEE TRANSACTIONS ON COMMUNICATIONS (2008–2013). He is currently an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA.

CHUNPING HOU received the M.Eng. and Ph.D. degrees in electronic engineering from Tianjin University, Tianjin, China, in 1986 and 1998, respectively.

Since 1986, she has been the faculty of the School of Electronics and Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and the 3D Imaging Institute. Her current research interests include 3D image processing, 3D display, wireless communication, and the design and applications of communication systems.

BOYANG QU received the B.E. and Ph.D. degrees from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a Professor with the School of Electronics and Information, Zhongyuan University of Technology, China.