

Research Article

A Traffic Parameters Extraction Method Using Time-Spatial Image Based on Multicameras

Jun Wang and Deliang Yang

School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Jun Wang; wangjun88100720@emails.bjut.edu.cn

Received 8 September 2012; Revised 4 December 2012; Accepted 5 December 2012

Academic Editor: Liguozhang

Copyright © 2013 J. Wang and D. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on the traffic monitoring system consisting of image processing units (IPUs), network communication, and data fusion server, this paper proposes an approach for the extraction of traffic parameters in the highway based on time-spatial image and data fusion. This system used a simple method to install cameras, which can capture the video images from the left to the right perspectives synchronously, for the sake of reducing the influence from vehicles height or width by fusing the image plane information of multicameras. Firstly, based on the synchronous and adaptive camera calibration method, images from different cameras are projected onto a same world coordinate, which is a top view picture of the same detecting region. Secondly, in order to improve the environment adaptability, an advanced method of time-spatial image is proposed to extract traffic information. At the same time, through network communication, traffic information picked up by each camera is transformed to the SDF timely. Finally, combined with probability fusion map, SDF fuses information grabbed by all cameras to improve the accuracy of traffic parameters extraction. The experimental results show that the proposed method can be implemented quickly and accurately either in crowded state or in fluent state.

1. Introduction

With the development of image processing technology, traffic monitoring system based on multicameras has received more and more attention. Many research achievements have been transformed into commercial products, such as the vehicle infrastructure integration (VII) project in the USA, SARETEA project in Europe, and advanced cruise assist highway system (AHS) project in Japan. Among them, papers [1, 2] describe a traffic monitoring system which is composed of three parts by ATUs, Network Communications, and SDF, based on the architecture of TRAVIS. Based on background updating technology, each ATU, composed of a series of video sensors, extracts the traffic information such as vehicle type and location from the video frames. Using the grid-based fusion or foreground probability map fusion technology, SDF fuses the traffic information from the ATUs connected with SDF via the internet. The experimental results show that this system is flexible, versatile, and practical. Based on the vehicle infrastructure integration (VII) system and the theory of shock wave, Koutsia et al. [3] designed a traffic monitoring

system by placing video sensors on different road sections. The model calculates the average vehicle velocity of every road section, tracks the moving target using Spatial-Temporal Markov Random Field model (S-T MRF model), predicts traffic state of the current road section and downstream sections, and reduces the incidence of traffic incidents such as traffic congestion.

Currently, on the area of microcosmic traffic parameters extraction, such as the length of the vehicle, most researches focused on the technology of 3D model reconstruction using multiple cameras, which can get a high accuracy and prolific information about the shape of the vehicle, but it is more time consuming and complex simultaneously. Generally, macroscopic traffic parameters are extracted by background updating and optical flow through tracking the moving vehicles, which is more precise at the cost of simplicity of algorithm attributed to updating background and retrieving optical flow points in time and prone to become invalid under the traffic jam. Fujimura et al. and Lamosa et al. [4, 5] proposed an image fusion method based on multilevel probability fusion maps, which analyzes each pixel's foreground probability to

obtain a new foreground image firstly, then analyzes the plane projective map of diverse height between the object and its plane to detect the existence of vehicles and compute their height and width. This method can segment the vehicles accurately, but shape information is detected imprecisely. A method of traffic parameters extraction based on time-spatial image using a single camera was proposed by Zhu et al. [6]. It will extract macroscopic traffic parameters, for instance, traffic flow and average speed, and microcosmic parameters containing height and width of the vehicle, through the analysis of PVI (the panoramic view image and EPI (the epipolar plane image)). The way is resilient to circumstances and less demanding for background update, while it has more sophisticated mathematical operation. Reference [7] adopts multivirtual line (MVDL), of which the method sets up several vehicle-detecting lines along the direction of the lane to obtain a series of PVI images. The idea is excellent in traffic flow statistics, but the method is more reliable on the angle of cameras, which is difficult to avoid.

Based on the traffic monitoring system consisting of image processing units (IPUs), network communication, and data fusion server (SDF), this paper proposes an approach for traffic parameters extraction in the highway, such as vehicle number, vehicle velocity and vehicle type, based on time-spatial image [6] and image fusion. This system used a simple method to install cameras, which can capture the video images from the left to the right perspectives synchronously, for the sake of reducing the influence from vehicles height or width by fusing the image plane information of multicameras. Firstly, based on the synchronous and adaptive camera calibration method [8, 9], project images from different cameras onto a same world coordinate, which is a top view picture of the same detecting region. Secondly, in order to improve the environment adaptability, an advanced method of time-spatial image is proposed to process the video images and extract traffic parameters. At the same time, through network communication, traffic information picked up by each camera is transformed to the data fusion server (SDF) in time. Finally, combined with probability fusion maps (PFM) [1, 2, 4, 5], information grabbed by all of related cameras is fused to improve the accuracy of traffic parameter extraction by the SDF. The experimental results show that the proposed method can be implemented quickly and accurately either in crowded state or in fluent state. Therefore, in the future, combined with pattern recognition and data mining technology, this method can be used to realize the traffic state identification and prediction and provide basis for section optimization.

This paper is organized as follows. Section 2 will demonstrate the system architecture, including the functions and implementation methods of IPUs, network communication, and SDF. The procedure of image processing and parameters extracting concerning IPU will be elaborated in Sections 3. Sections 4 and 5 will briefly describe the realization process or basic theories concerning the functions of SDF and network communication, respectively. Section 6 gives the experimental results, and a brief conclusion is also presented in Section 6.

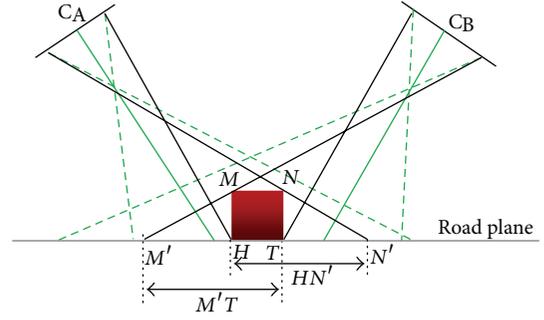


FIGURE 1: Perspective projection of cameras with different angle. Green lines denote the visual range of cameras.

2. Problem Description and System Framework

2.1. Problem Description. In order to guarantee the accuracy of the information extraction, in the existing methods based on computer vision, the video cameras, which are mounted in the position near the target section, access the road traffic information by using nearly perspective method. Thus the shape information of vehicles will be greatly affected in the 2D plane. Just as what is shown in Figure 1.

Assume that the THMN plane is the section along the vehicle width direction and HT is the vehicle width. Affected by the height of the vehicle, the vehicle width in the image flat is HN' in the perspective C_A , while it is $M'T$ in the perspective C_B . $M'T$.

If we fuse the image plane information of the two cameras into D ($D = A \cap B$), the influences on the vehicle width shown in the 2D plane from cameras of different perspective could be cancelled each other out, and thus the vehicle width range can be compressed to HT ($HN' \cap M'T = HT$). By this way the vehicle position can be detected accurately. As shown in Figure 2(e) and in Figure 2(d), the green, red, and yellow regions denote the vehicle contours in the camera 1, 2, in Figures 2(a), and 2(b) and the fusion image, respectively.

Therefore, in order to guarantee the accuracy of the information extraction, this paper will, on the one hand, access the clear and abundant traffic image information by using the nearly perspective method with multicameras video sensors. On the other hand, in order to reduce the influence from vehicles such as height and width, put forward a method to realize the traffic monitoring system consisting of IPU, network communication, and SDF, to extract the traffic parameters such as road vehicle number, vehicle velocity, and vehicle type.

2.2. System Introduction. Based on the traffic monitoring system consisting of IPUs, network communication, and SDF, this paper proposes an approach for traffic parameters extraction in the highway, such as vehicle number, vehicle velocity, and vehicle type, as shown in Figure 3.

(A) *Image Processing Unit (IPU)*. There is a corresponding relationship between cameras and IPUs. And each IPU is

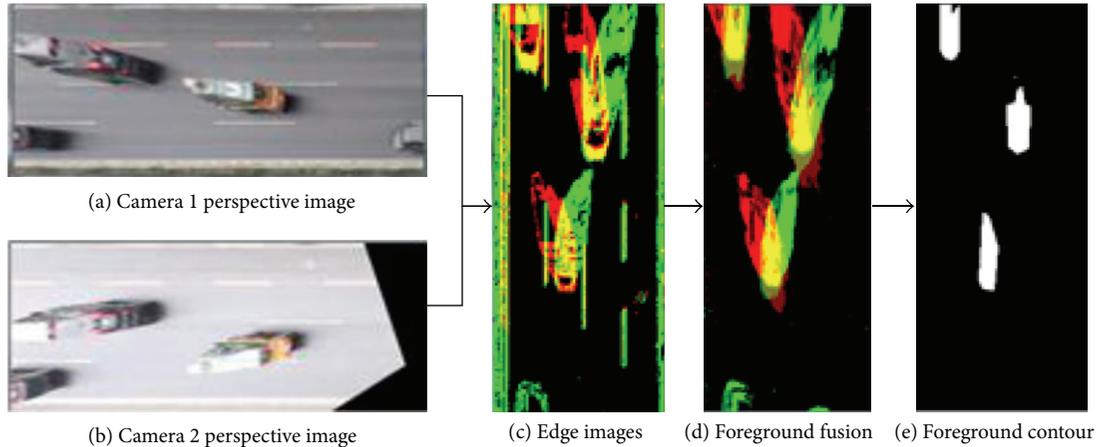


FIGURE 2: Foreground fusion.

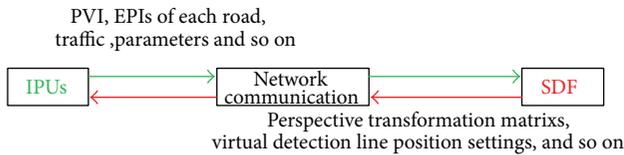


FIGURE 3: Traffic monitoring system.

used to capture related video streaming, to process vision information, and to extract real traffic formation. In this paper, an advanced method based on the method mentioned in [6, 10] will be used to generate PVI and EPI, which are useful for traffic information extraction, such as traffic flow, vehicle velocity, type, and so on.

There is no requirement on the camera number, height, and shooting mechanism of each camera in this paper, but the following installation conditions should be satisfied.

- (1) The target road section should be highway.
- (2) The cameras should be installed on the left and the right perspectives simultaneously on the target road section, in order to capture the video images from the left and the right perspectives synchronously.
- (3) The cameras should be installed on the same road section, upstream road section, or downstream road section, to ensure that the directions of all the moving vehicles in the target road section are the same.

(B) *Network Communication*. Each IPU, which is an independent unit, exchanges information with SDF by the internet. In this paper, based on method of [1, 2], the internet protocol (IP), including the transmission control protocol (TCP) and the user datagram protocol (UDP), will be used to transmit traffic information from IPU to SDF, and the configuration information of IPU can be controlled by the SDF using the IP protocol similarly. In this way, time-spatial image and traffic parameters can be transmitted to SDF synchronously from IPU. Accordingly, the image perspective switching matrix of

each IPU and the position of virtual detecting lines can be set flexibly by SDF.

(C) *Sensor Data Fusion Server (SDF)*. The following three functions can be achieved by SDF.

- (1) IPU camera calibration: this paper will use the method of [8, 9, 11] to calibrate the camera group synchronously and adaptively and calculate the perspective transformation matrix of each camera to ensure that the images from different cameras can be projected onto a top view picture with the same size and direction, as shown in the perspective images of the camera 1 and camera 2 in Figures 2(a) and 2(b).
- (2) Virtual detection and tracking lines position settings: based on the perspective images and the methods of [7, 10, 12], the position of the detection line vertical to the direction of traffic movement to detect whether the vehicle is passing and the lines parallel to the direction of traffic movement to track the existing vehicles, which are shown in Figure 4 by red lines, can be set to generate the time-spatial image.
- (3) IPU information fusion: based on time-spatial image of each IPU and the method of PFM [1, 2, 4, 5], the fused image of the PVI and EPIs can be obtained, respectively, by image fusion on the pixel level. And then a series of traffic parameters can be extracted using the traffic parameter extraction strategy based on the time-spatial fusion image.

3. Image Processing Unit

In order to improve environment adaptability of the system, the method of time-spatial image [6, 7, 10, 13] will be used to process the video image. First of all, the images from all cameras should be projected onto the same horizontal area by the method of perspective transformation. Secondly, the PVI and the EVI of each lane will be obtained using the method of time-spatial image. Finally, the traffic parameters such as vehicle type, vehicle velocity, and traffic flow of the vehicles



FIGURE 4: The set of virtual detecting lines.

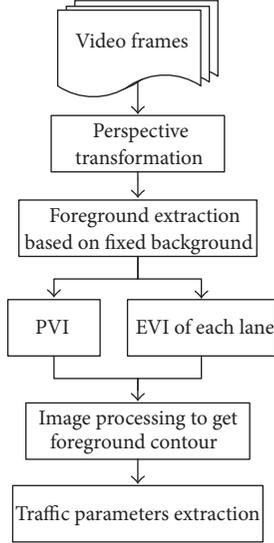


FIGURE 5: The traffic parameters extraction method of IPU.

through the virtual detection line can be obtained. The traffic parameters extraction method of IPU is shown in Figure 5.

3.1. Generation and Processing of Time-Spatial Image. As shown in Figure 4, based on the perspective images and the methods of [7, 10, 12], the position of the detection line vertical to the direction of traffic movement to detect whether the vehicle is in the road and the line parallel to the direction of traffic movement to track the existing vehicles can be set, and the corresponding PVI and EPIs of each lane can also be generated. Then extract the binary image of vehicle outline from the PVI and EPIs based on the image processing technology such as morphology.

(A) Generation of Time-Spatial Image. The generation process of PVI and EPI is shown in Figure 6. FPI_{t-i} ($1 \leq i < N$) is defined as the perspective image of video frame $t-i$; the red line vertical to the direction of traffic movement, defined as $dvdl$, is the virtual detecting line for vehicle detecting, and the other red line parallel to the direction of traffic movement, defined as $tvdl$, is the virtual detecting line for vehicle tracking.

The relation between PVI and EPI is shown in Figure 7. Both numbers of pixels in f axis of PVI and EPI are N , which is the number of frames used to create time-spatial image one time. And l_{dvdl} and l_{tvdl} are the number of pixels of $dvdl$ and $tvdl$ in Figure 6, respectively.

The value of pixel (f, y_{dvdl}) in PVI is defined as formula (1). The value of pixel (f, y_{tvdl}) in EPI is defined as formula (2).

$$\begin{aligned} PVI(f, y_{dvdl}) \\ = FPI_{t+i}(dvdl_x, \min_y_{dvdl} + y_{dvdl}), \end{aligned} \quad (1)$$

$$\begin{aligned} EPI(f, y_{tvdl}) \\ = FPI_{t+i}(dvdl_x - l_{tvdl} + x_{tvdl}, tvdl_y), \\ f = (t+i) \% N + 1. \end{aligned} \quad (2)$$

In formula (1) and (2), FPI_{t+i} denotes the FPI $t+i$, $dvdl_x$ and $tvdl_y$ are the x coordinate of $dvdl$ and the y coordinate of $tvdl$ in Figure 6, respectively, and \min_y_{dvdl} is the minimum of y coordinate of $dvdl$ in Figure 6.

(B) Image Processing of Time-Spatial Image. Based on the image processing technology, such as morphology [7, 12], extracting the binary image of vehicle outline from the PVI and EPIs, one example of processing result is shown in Figure 8.

3.2. Traffic Parameters Extraction. This paper proposed a simple and effective method to extract traffic parameters, based on the hypothesis that vehicles move with constant speed when passing the virtual detection line ($dvdl$ and $tvdl$). Based on time-spatial images of PVI and EPIs analyses, basic traffic information, including vehicle existence, approximate length, and velocity, is extracted.

Firstly, through the analysis of PVI and EPIs synchronously, vehicle existence detection of each lane is achieved. Then, the vehicle approximate length is extracted. Next, the vehicle velocity can be calculated readily. Finally, based on the three kinds of traffic information mentioned previously, variety of traffic parameters such as the number of passed vehicles, the vehicle velocity, the vehicle type, the distance between successive vehicles, and the road occupancy rate can be obtained easily:

$$\text{counts} = \sum_{t=1}^{N-1} D(t) \wedge D(t+1), \quad (4)$$

$$D(t) = \begin{cases} \text{true}, & \text{if } P(t) \cap E(t) = \text{true}, \\ \text{false}, & \text{others,} \end{cases} \quad (5)$$

$$P(t) = \begin{cases} \text{true}, & \text{if } PVI(t, y_{dvdl}) = 255, \\ & 0 \leq y_{dvdl} \leq l_{dvdl}, \text{ exists,} \\ \text{false}, & \text{others,} \end{cases} \quad (6)$$

$$E(t) = \begin{cases} \text{true}, & \text{if } EPI(t, y_{dtvl}) = 255, \\ & 0 \leq y_{dtvl} \leq l_{tvdl}, \text{ exists,} \\ \text{false}, & \text{others.} \end{cases} \quad (7)$$

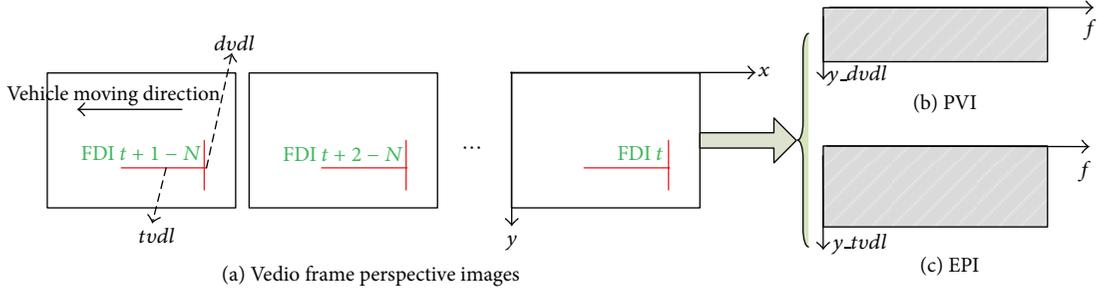


FIGURE 6: Generation process of PVI and EPI.

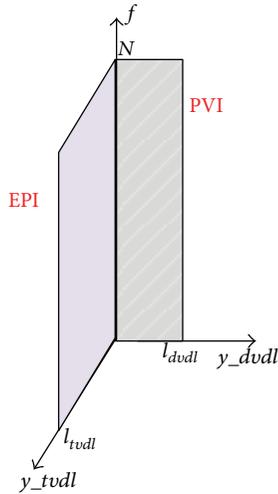


FIGURE 7: The relation between PVI and EPI.

In Figure 9 and formula (4), counts is defined as the existing number of vehicles in the time interval of PVI and EPI related. The area between two adjacent dotted red lines denotes that a vehicle is detected, while the area between two adjacent dotted green lines denotes that no vehicle is detected. Therefore, only foreground is detected simultaneously in PVI and EVI, vehicle can be inspected in this period, as shown in formula (4)–(7), among which t is defined as the horizontal ordinate of time-spatial images, and it represents the time label of video frame as discussed in Section 3.1, and $D(t)$, $P(t)$, and $E(t)$ denotes the ultimate detecting result of vehicle existence, the detecting result of PVI and EPI at the moment of video frame t , respectively.

4. Sensor Data Fusion Server

The SDF server mainly realizes multicameras synchronous calibration, virtual detection and tracking lines position settings, time-spatial image collecting and fusion using a constant polling cycle, and traffic parameter extraction based on the fusion image using the method described in Section 3.2. Among them, the method of virtual detection and tracking lines position settings has been discussed in Section 3.1. This section describes the basic theories of multicameras synchronous calibration and image fusion.

4.1. Multicameras Synchronous Calibration. A calibration technique, which is based on a 3×3 homographic transformation and uses both point and line correspondences, was proposed in paper [11] to map images into a reference world coordinate system corresponding to an approximate bird's eye view of the real scene, so that any point can be converted from image coordinates to ground coordinates and vice versa.

Assume that the surface of the road is flat; this paper refers to paper [11] and ignores the height information in the world coordinate system, to map multicameras into the same reference world coordinate system corresponding to an approximate bird's eye view of the same real scene. Define $C^w = (x^w, y^w)$ and $C^I = (x^I, y^I)$ as the point of the world coordinate system and the image coordinate, respectively. So there exists a fixed perspective transformation matrix M and variable λ meeting formula (8), and between them λ is a variable related to corresponding coordinate point:

$$\lambda \begin{bmatrix} x^w \\ y^w \\ 1 \end{bmatrix} = M \cdot \begin{bmatrix} x^I \\ y^I \\ 1 \end{bmatrix}, \quad (8)$$

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & 1 \end{bmatrix}.$$

According to Figure 10, define the direction and the vertical direction of lane lines as the x^w axis and y^w axis in the world coordinate system, respectively. Denoting that (x_j^w, y_j^w) and (x_j^I, y_j^I) are one of point correspondence in the world and the image coordinate systems, respectively, and one of corresponding calibration line is defined as $y^w + c_i^w = 0$ and $a_i^I \cdot x^I + b_i^I \cdot y^I + c_i^I = 0$ in the world and the image coordinate systems, respectively. Referencing paper [11], variables are defined as follows:

$$P_j = \begin{bmatrix} x_j^I & y_j^I & 1 & 0 & 0 & 0 & -x_j^w \cdot x_j^I & -x_j^w \cdot y_j^I \\ 0 & 0 & 0 & x_j^I & y_j^I & 1 & -y_j^w \cdot x_j^I & -y_j^w \cdot y_j^I \end{bmatrix},$$

$$T_j^P = [x_j^w, y_j^w],$$

$$L_i = \begin{bmatrix} 0 & 0 & 0 & 1 & -\frac{a_i^I}{b_i^I} & 0 & c_i^w & -\frac{c_i^w \cdot a_i^I}{b_i^I} \\ 0 & 0 & 0 & 0 & -\frac{c_i^I}{b_i^I} & 1 & 0 & -\frac{c_i^w \cdot c_i^I}{b_i^I} \end{bmatrix}, \quad (9)$$

$$T_i^L = [0, c_i^w].$$

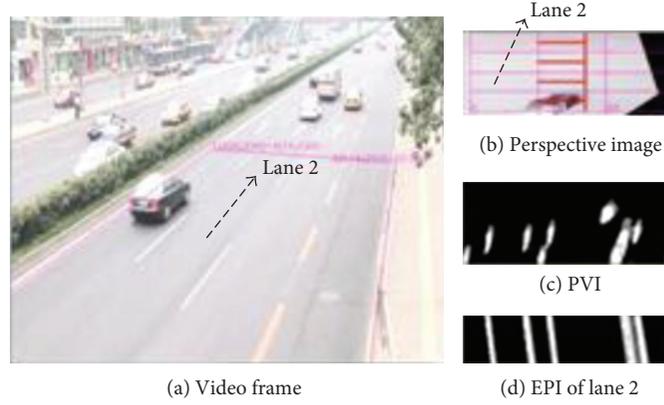


FIGURE 8: Example of processing result.

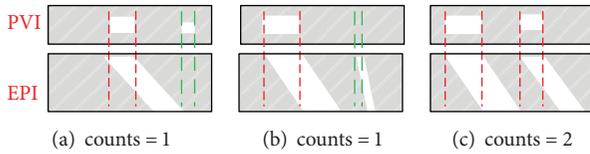


FIGURE 9: Vehicle existence detection: PVI and EPI analyses to detect passed vehicles.

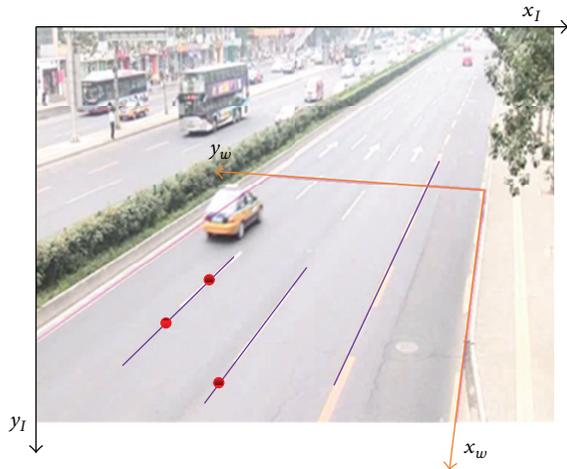


FIGURE 10: Coordinate systems and calibration points and lines. The red circles and purple lines respectively represent point and line correspondences.

Consequently, related equations defined below can be deduced, and with at least two line correspondences, parameters such as m_{22} , m_{31} , and m_{32} can be calculated, and then using at least three point correspondences, rest parameters of \overline{M} can be figured out:

$$\begin{aligned}
 A \cdot \overline{M} &= B, \\
 A &= [P_1^T, P_2^T, P_3^T, L_1^T, L_2^T, \dots]^T, \\
 B &= [T_1^P, T_2^P, T_3^P, T_1^L, T_2^L, \dots]^T, \\
 \overline{M} &= [m_{11}, m_{12}, m_{13}, m_{21}, m_{22}, m_{23}, m_{31}, m_{32}]^T.
 \end{aligned} \tag{10}$$

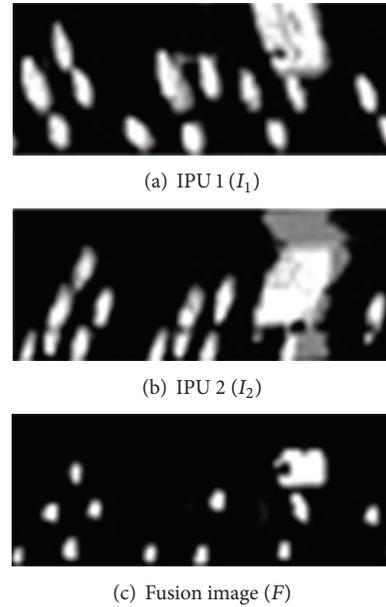


FIGURE 11: Image fusion.

\overline{M} can be solved by the method of the least square, as shown in formula (11):

$$\overline{M} = (A^T \cdot A)^{-1} A^T \cdot B. \tag{11}$$

In the SDF platform, utilizing perspective matrix, all of images related to specific cameras are projected to a same bird's eye view in the same world coordinate system. And then check the contact ratio with respect to all bird's eye views. If the contact ratio is too low, the SDF could adjust the point coordinate or linear equation correspondences. The procedure is repeated until the contact ratio of all bird's eye views is high enough in order to realize multicameras synchronous calibration.

4.2. Probability Fusion Map (PFM). As discussed in Section 2.1, because of the fact that height information is ignored in camera calibration and internal and external



FIGURE 12: Network communication procedure.

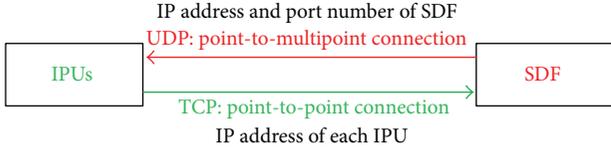


FIGURE 13: Network connection process.

camera factors including the camera orientation, the aperture angle, and the off-plane distance of the world point, there exists certain distortion to show moving targets in the bird's eye view through the perspective transformation, where the closest points to the surface will be the least distorted.

For any pixel in the 2D plane, it is only falsely projected as the foreground of vehicle by several video cameras. So that weighting the images of each IPU can weaken the influence, as can be seen in Figure 2. In this paper, the method of probability fusion map (PFM), which is firstly mentioned in paper [5], is adopted and optimized in terms of camera weight estimation and image fusion formula.

4.2.1. Camera Weight-Matrix Estimation. According to the characteristics of error transmissibility, paper [11] deduced that visual error generated by the conversion process from 2D image plane to 3D physical plane is related not only to camera pixels error but also to both inside and outside camera parameters. Paper [5] has demonstrated that the perspective transformation matrix was related to both inside and outside camera parameters, namely that based on the perspective transformation matrix and the camera pixels error, the deviation in the corresponding bird's eye view, which is called visual error before, can be estimated.

From formula (8), it is known that the world coordinate point (x^w, y^w) and the camera image coordinate point (x^I, y^I) have relations described in formulas (12) and (13). Therefore, the relationship between the visual error of bird's eye view about the world coordinate system and the system error of camera can be defined in formulas (14) and (15).

$$x^w = \frac{m_{11} \cdot x^I + m_{12} \cdot y^I + m_{13}}{\lambda}, \quad (12)$$

$$y^w = \frac{m_{21} \cdot x^I + m_{22} \cdot y^I + m_{23}}{\lambda}, \quad (13)$$

$$\sigma_{x^w,i} = \frac{m_{11}^2 \cdot \sigma_{x^I} + m_{12}^2 \cdot \sigma_{y^I}}{\lambda_i^2}, \quad (14)$$

$$\sigma_{y^w,i} = \frac{m_{21}^2 \cdot \sigma_{x^I} + m_{22}^2 \cdot \sigma_{y^I}}{\lambda_i^2}. \quad (15)$$

Assume that the deviation of all cameras is the same, that is, $\sigma_{x^I} = \sigma_{y^I}$, to define a bigger camera weight when the visual deviation is smaller while to define a smaller camera weight when the visual deviation is bigger, the relationship between visual error and the weight of the related camera is evaluated as follows:

$$\begin{aligned} w_{(x^w, y^w),i} &= \frac{1/(\sigma_{x^w,i} \cdot \sigma_{y^w,i})}{\sum_{i=1}^n 1/(\sigma_{x^w,i} \cdot \sigma_{y^w,i})} \\ &= \frac{\lambda_i^4 / [(m_{11}^2 + m_{12}^2) \cdot (m_{21}^2 + m_{22}^2)]}{\sum_{i=1}^n \lambda_i^4 / [(m_{11}^2 + m_{12}^2) \cdot (m_{21}^2 + m_{22}^2)]}. \end{aligned} \quad (16)$$

In formula (16), $w_{(x^w, y^w),i}$ is defined as the weight about point (x^w, y^w) of camera i related to IPU i .

4.2.2. Image Fusion. Traffic parameter extraction methods based on data fusion are mainly divided into 2 kinds, data fusion on feature layer and on pixel layer [2, 3]. Through grid-based fusion on feature layer [3], traffic parameters from each IPU are fused to estimate traffic parameters precisely with less network bandwidth demand. Compared with those kinds of method, methods of data fusion on pixel layer have increased computational and network bandwidth requirements, but they can very robustly resolve occlusions among multiple views, just as what is discussed in Section 2.1.

Denoting the value of pixel to be $I_i(x^w, y^w)$ about point (x^w, y^w) in IPU i . Assume that there are n IPUs in the system. F represents the result image of image fusion, and $F(x^w, y^w)$ denotes the value of pixel about point (x^w, y^w) in F . The result image of image fusion is generated as follows:

$$F(x^w, y^w) = \begin{cases} 255, & \text{if } \sum_{i=1}^n w_{(x^w, y^w),i} \cdot I_i(x^w, y^w) \geq T, \\ 0, & \text{others.} \end{cases} \quad (17)$$

In formula (17), $w_{(x^w, y^w),i}$ has the same meaning described in Section 4.2.1. T denotes a threshold used for vehicle existence detection. Figure 11 gives the result of image fusion using the method of PFM when $T = 500$.

5. Network Communication

The traffic monitoring system proposed in this paper is based on client-server network model where each IPU acts as a client, while the SDF plays the role of listening server. With the help of wired or wireless local area network (LAN) facilitates, each IPU sends video information, such as time-spatial images, to the server via point-to-point connection using TCP protocol, and each IPU is remotely controlled by

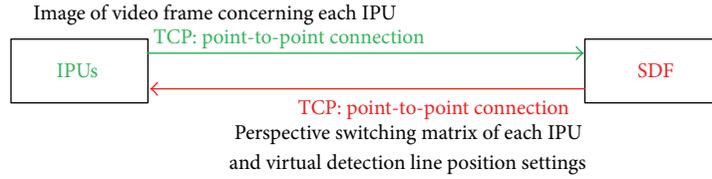


FIGURE 14: Transmission process concerning video image configuration information.

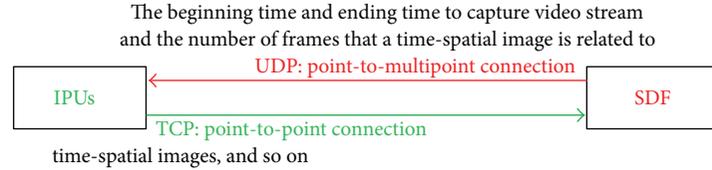
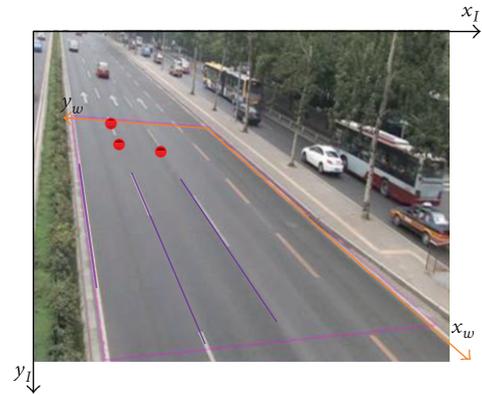


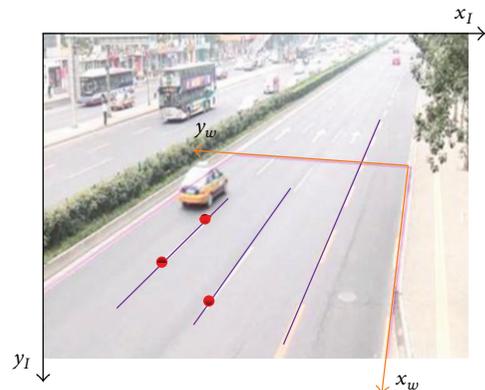
FIGURE 15: Process of synchronous image transmission.



FIGURE 16: Set of cameras.



(a) Camera left



(b) Camera right

FIGURE 17: Coordinate systems and calibration points and lines. The red circles and purple lines respectively represent point and line correspondences. The thick black line indicates the image coordinate system C_I , and the thick orange line indicates the world coordinate system C_w .

the SDF server through an independent signaling channel via point-to-point or point-to-multipoint connection, which enables the SDF server user to start or stop capturing frames, transmit perspective switching matrix and the settings of virtual detection line, and so on.

This paper is based on software platforms to realize real-time information fusion of multicameras; therefore, the system shall satisfy 3 key points. Firstly, it would be better that all cameras connected to a unique IPU separately have the same pattern. Hence, in this paper all cameras in the system have the same mode, resolution, and frame rate. Secondly, IPUs and SDF should be in the state of clock synchronization. Thirdly, within a certain time cycle, SDF must receive synchronous time-spatial images from each IPU. To fulfill the second and third points, on one hand, the network time protocol (NTP) is used in order to synchronise the system clocks of each related computer in the traffic monitoring system with reference to SDF server's system clock. On the other hand, considering network transmission delay caused by factors such as the capacity of internet channel and software processing speed, this paper references [2, 3, 14] to adjust parameters of the beginning time and ending time to capture video stream and the number of frames that a Time-Spatial image related to, according to the accuracy and time efficiency of receiving information, in order to ensure that

each IPU cannot only synchronously capture video streaming but also generate and send time-spatial image within a certain time window to SDF. And then SDF issues the parameters to each IPU to ensure synchronization.

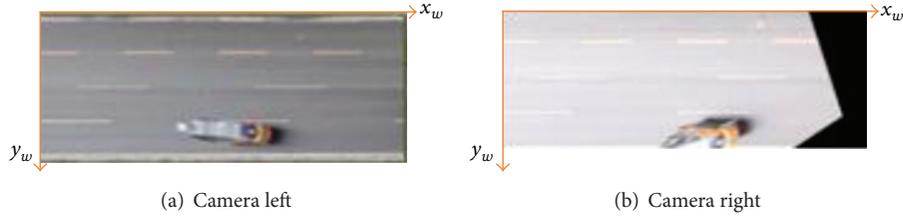


FIGURE 18: Calibration result: approximate bird's eye views of the testing region.

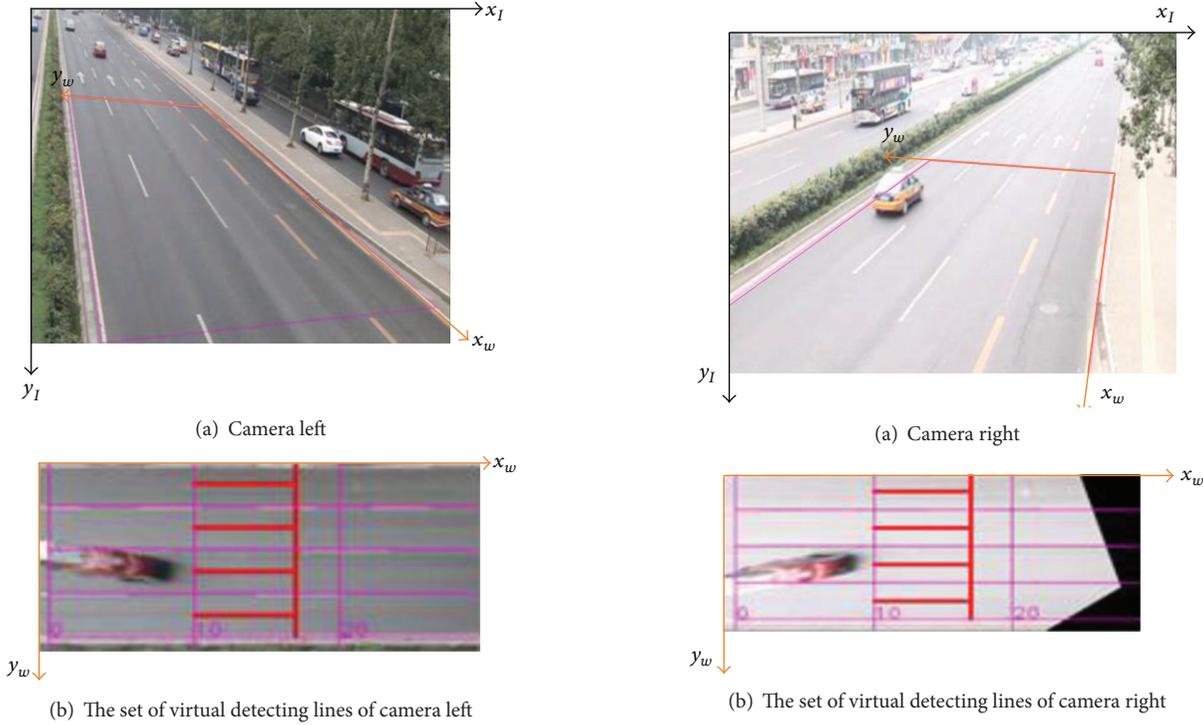


FIGURE 19: Set of virtual detecting lines in I_{left} .

FIGURE 20: Set of virtual detecting lines in I_{right} .

In general, network communication is realized through socket, and network communication procedure is divided into three steps, as shown in Figure 12.

(A) *Network Connection Process.* After installation, all IPUs in the traffic monitoring system are set in standby mode. At this phase, each IPU is listening in a UDP port for information in an independent signaling channel from the SDF. The UDP port has been set up before connection, and the signaling channel is specifically defined in order to separate data transmission of the IPUs from signals sent by SDF.

When SDF server sends the information packet which consists of SDF's IP address and its listening port number using UDP to the IPUs, network connection process starts. The next step is for each IPU to send respective IP address and request TCP connection to the listening SDF server. After all handshake procedures have been established and the meeting time has been reached, the network connection



FIGURE 21: PVI of camera left (I_{left}) generated by video frames from frame 24,000 to frame 24,300.

process succeeds, and the system enters the next stage. The process of network connection is briefly shown in Figure 13.

(B) *Transmission Process concerning Video Image Configuration Information.* At this stage, each IPU captures several video frames and then transmits them to the SDF based on TCP. Next, on the basis of video images about each IPU, SDF accomplishes synchronous camera calibration, to map each IPU frames into a same reference world coordinate system corresponding to an approximate bird's eye view of the same



FIGURE 22: EPIs of camera left (I_{left}) generated by video frames from frame 24,000 to frame 24,300.

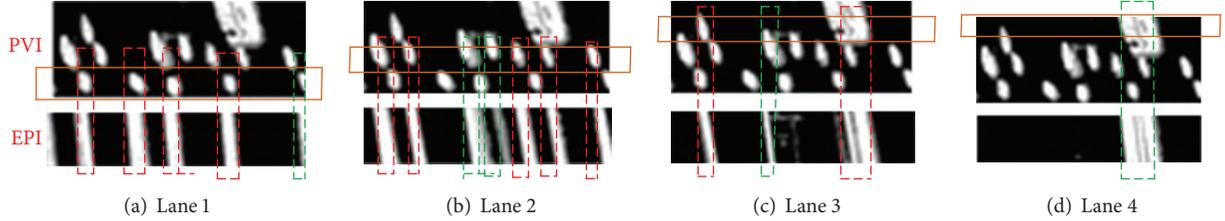


FIGURE 23: Vehicle existence detection of camera left generated by video frames from frame 24,000 to frame 24,300. In the figure, the orange rectangular box represents the position of homologous lane. And the red rectangular box and green rectangular box, respectively, represent the correct-detected and false-segmented vehicles.



FIGURE 24: PVI of camera right (I_{right}) generated by video frames from frame 24,000 to frame 24,300.

scene. Finally, the results of camera calibration, including the perspective switching matrix of each IPU and the settings of virtual detection lines, are sent to the homologous IPU, respectively.

After this stage, the system enters normal operation, in the phase of which, SDF and IPU software maintain timers to control frame capture, generate and transform synchronous time-spatial image with a const cycle. The transmission process is briefly shown in Figure 14.

(C) *Process of Synchronous Image Transmission.* The process of synchronous image transmission is the normal operation of the system. At this stage, firstly, through the first two phases, the network delay is estimated, based on which SDF sets a reasonable number of frames that a time-spatial image related to, which is defined as N in Section 3.1 and is related to the cycle of image transmission T . Secondly, IPU software maintains timers to control frame capture, generates synchronous time-spatial image with the period of N frames, and then transforms them timely to SDF with the cycle of T using TCP to achieve data fusion of multicameras in SDF server.

T is calculated by the formula below:

$$T = N \cdot f. \quad (18)$$

In formula (18), f is defined as camera's frame rate per second. In addition, as a secondary backup system, the IPUs support real-time media streaming directly to SDF server, in order to assist operators when an accident or other abnormal situation is reported. To confirm effective transmission, a new

independent port is defined to receive synchronous video frame information. So there will be 2 independent ports to intercept disparate information, which are defined as time-spatial image listening port and video frame listening port. The process of synchronous image transmission is briefly shown in Figure 15.

6. Experimental Results

6.1. Experimental Environment

6.1.1. *Camera Installation and the Set of IPUs.* The proposed method is evaluated using two IPUs, which are all running in an independent PC with an Intel i3-2310M 2.10 GHz and 2 GB RAM with the aid of visual studio 2008. Both of them are related to a PAL video camera grabbing videos at the highway of Beijing Jinsong Bridge. Both of the two cameras have a 720×576 resolution at a frame rate of 25 frames per second. The cameras are fixed on the overpasses: one is set on the left and the other is set on the right of the interesting region, as shown in Figure 16, which are defined as I_{left} and I_{right} , respectively.

Each of the testing video streaming has 90,000 frames in all, representing the peak hour from 12:00 to 13:00. The testing region, which has 4 lanes, is in crowded state during the former one-fourth frames, while the rest is in fluent state. Therefore, from 12:00 to 12:15 the testing region is in crowded state, and from 12:15 to 13:00 the testing region is in fluent state.

6.1.2. *Camera Calibration and the Settings of Virtual Detecting Lines in SDF.* The SDF is also operating on an independent PC with an Intel i3-2310M 2.10 GHz and 2 GB RAM with the aid of visual studio 2008. On the basis of video images about each IPU received by SDF, point and line correspondences concerning the image and the reference world coordinate systems [10, 11], the images are mapped into a reference world coordinate system corresponding to an approximate bird's eye view of the real scene, as shown in Figures 17 and 18.



FIGURE 25: EPIs of camera right (I_{right}) generated by video frames from frame 24,000 to frame 24,300.

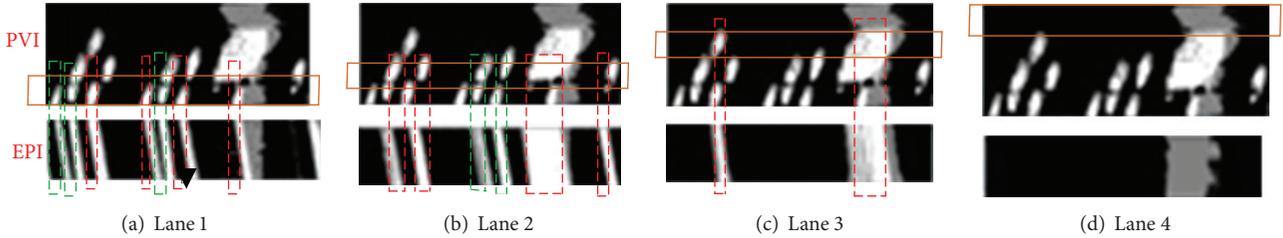


FIGURE 26: Vehicle existence detection of camera right generated by video frames from frame 24,000 to frame 24,300. In the figure, the meanings of color rectangular boxes are the same as those of Figure 23.

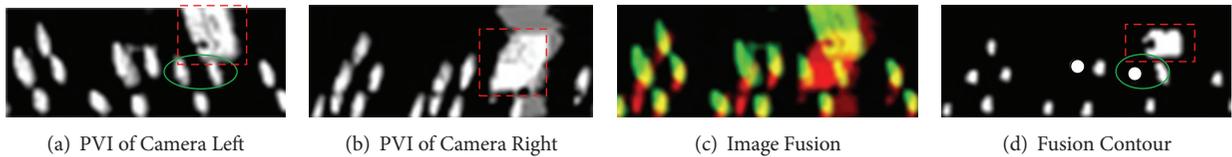


FIGURE 27: PVI images generated by video frames from frame 24,000 to 24,300. In the image (c), fusion map is shown, where the green, red and yellow region denote the vehicle contours in the camera left, camera right and the fusion image, respectively.

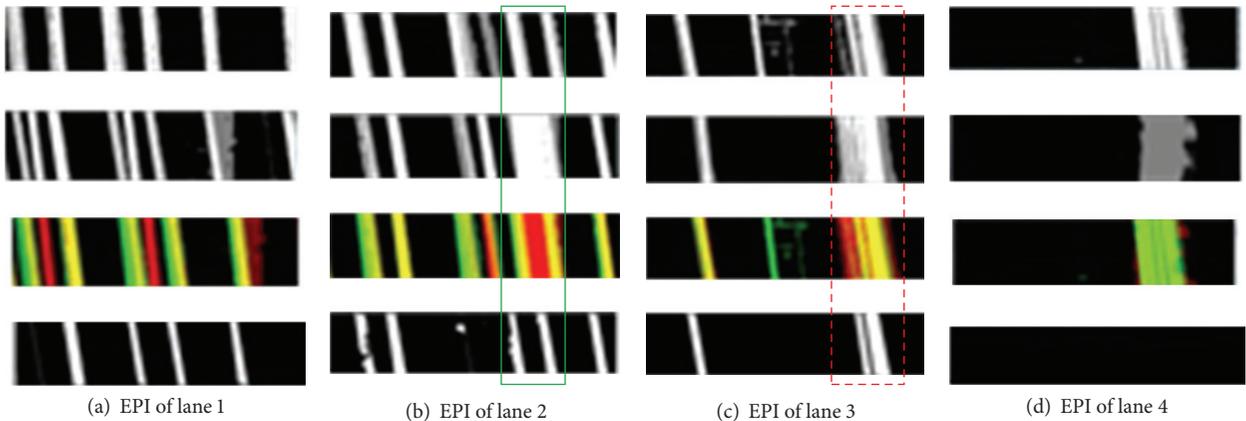


FIGURE 28: EPIs of each road generated by video frames from frame 24,000 to 24,300. The first and second row show the EPIs of each lane in camera left and camera right, respectively. In the third row, fusion map is shown, where the green, red and yellow region denote the vehicle contours in the camera left, camera right and the fusion image, respectively. The fusion contour is shown in the fourth row.

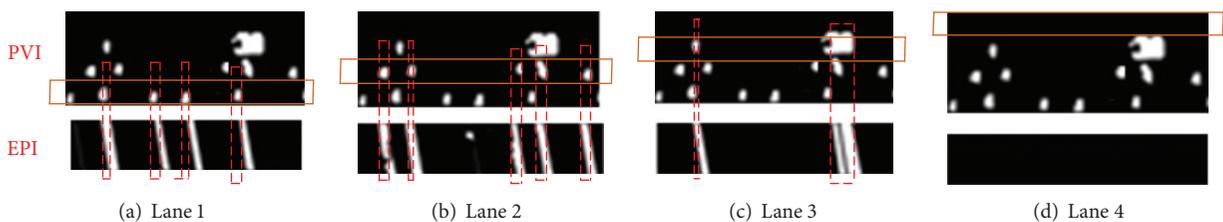


FIGURE 29: Vehicle existence detection with the method of PFM data fusion from frame 24,000 to frame 24,300. In the figure, the meanings of color rectangular boxes are the same as thos of Figure 23.



FIGURE 30: PVI of camera left (I_{left}) generated by video frames from frame 2,175 to frame 2,475.

Based on the results of calibration, virtual detecting and tracking lines are confirmed through SDF. Figures 19 and 20 show the results of camera calibration and virtual detecting lines settings. Between them, Figures 19(a) and 20(a) show the detecting region, which is surrounded by the pink color box. The thick red lines in Figures 19(b) and 20(b) are the settings of virtual detecting lines, of which the vertical and horizontal lines on each road denote the $dvdL$ and $tvdL$ s, respectively. Horizontal pink lines in Figures 19(b) and 20(b) denote traffic lane lines.

6.2. The Experimental Results. To show the effectiveness of the method proposed in this paper, experimental results about traffic volume detection concerning the method of a single IPU and PFM data fusion with SDF, running under the experimental environment introduced in Section 6.1 which consists of 2 typical traffic states, are analysed and contrasted.

6.2.1. Experimental Results in Different Traffic State. The experimental environment introduced in Section 4.1 of this paper consists of 2 typical traffic states, fluent and crowded traffic states. In fluent state, both of traffic volume and velocity are relatively larger, while in crowded traffic state, traffic volume is large, but the velocity is relatively slower. In this paper, we randomly select two different time quantum, respectively, representing fluent and crowded traffic states, to proof the environmental suitability of our method.

To obtain real-time road traffic information, the time of delay-line is usually set as 10 s or so in time-spatial image [10, 13]. In our experiment environment, we set the time of delay-line as 12 s, namely, PVI and EPI images are constituted by 300 consecutive frames, so the width of time-spatial image is 300 pixels.

(i) *Fluent Traffic State.* From 12:15 to 13:00, the detecting region is in fluent state. So this paper randomly selects 12 s in this period, from 12:16:00 to 12:16:12, corresponding to video frames from frame 24,000 to frame 24,300, to analyze the experimental results concerning the method of a single IPU and PFM data fusion.

(A) *Experimental Results about a Single IPU.*

(1) *Camera Left (I_{left}).* Figure 21 shows the PVI image of the detecting road, and Figure 22 shows the EPIs of each lane, all of which are generated by the IPU related to camera left

TABLE 1: Experimental result about the IPU related to *camera left* in fluent traffic state.

Lane index	Lane 1	Lane 2	Lane 3	Lane 4	Total
Actual traffic volume (unit: number of vehicles)	4	5	2	0	11
Number of vehicles detected by IPU	5	7	3	1	16
Number of missed vehicles	0	0	0	0	0
Number of false-segmented vehicles	1	2	1	1	5
Detection error rate	25%	40%	50%	—	45.6%
Statistics error rate	25%	40%	50%	—	45.6%
The average computing time	About 32.5 ms per fame				

TABLE 2: Experimental result about the IPU related to *camera right* in fluent traffic state.

Lane index	Lane 1	Lane 2	Lane 3	Lane 4	Total
Actual traffic volume (unit: number of vehicles)	4	5	2	0	11
Number of vehicles detected by IPU	7	6	2	0	15
Number of missed vehicles	0	1	0	0	1
Number of false-segmented vehicles	3	2	0	0	5
Detection error rate	75%	60%	0	0	54.5%
Statistics error rate	75%	20%	0	0	36.4%
The average computing time (unit: s)	About 32.5 ms per fame				

(I_{left}). In Figure 23 the process and result of vehicle existence detection of camera left are described.

To measure the accuracy, we define the detection error rate γ_d and the statistics error rate γ_s , shown in lines 6 and 7 of Tables 1, 2, 3, 4, 5, and 6 as follows:

$$\gamma_d = \frac{n_m + n_f}{n_a} \times 100\%, \quad (19)$$

$$\gamma_s = \left| \frac{n_a - n_d}{n_a} \right| \times 100\%.$$

According to Figure 23, experimental result about the IPU related to camera left in fluent traffic state is summarized in Table 1.

(2) *Camera Right (I_{right}).* Figure 24 shows the PVI image of the detecting road, and Figure 25 shows the EPIs of each lane, all of which are generated by the IPU related to camera right (I_{left}). In Figure 26, the process and result of vehicle existence detection of camera right are described.

According to Figure 26, experimental result about the IPU related to camera right in fluent traffic state is summarized in Table 2.

(B) *PFM Data Fusion with SDF.* In SDF, on the basis of time spatial images from each IPU and PVI about detecting road



FIGURE 31: EPIs of camera left (I_{left}) generated by video frames from frame 2175 to frame 2475.

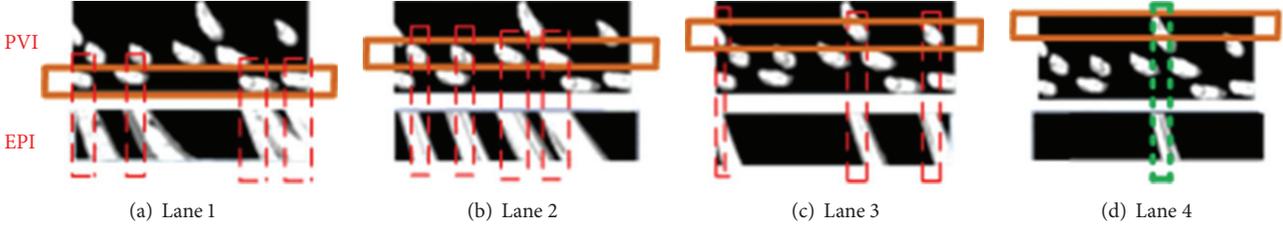


FIGURE 32: Vehicle existence detection of camera left generated by video frames from frame 2,175 to frame 2,475. In the figure, the meanings of color rectangular boxes are the same as those of Figure 23.

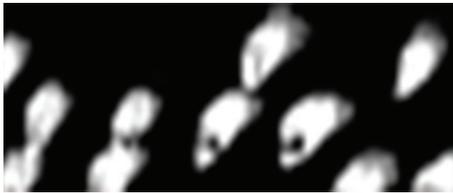


FIGURE 33: PVI of Camera Right (I_{right}) generated by video frames from frame 2,175 to 2,475.

TABLE 3: Experimental result with the method of PFM data fusion in fluent traffic state.

Lane index	Lane 1	Lane 2	Lane 3	Lane 4	Total
Actual traffic volume (unit: number of vehicles)	4	5	2	0	11
Number of vehicles detected by IPU	4	5	2	0	11
Number of missed vehicles	0	0	0	0	0
Number of false-segmented vehicles	0	0	0	0	0
Detection error rate	0	0	0	0	0
Statistics error rate	0	0	0	0	0
The average computing time	About 32.72 ms per fame				

and EPIs of each lane are revised with the method of PFM data fusion. Figure 27 described the process and the result of PFM data fusion about PVI of the detecting road, and Figure 28 described the process and the result of PFM data fusion about the EPIs of each lane.

In Figures 27 and 28, the dotted red boxes denote a bus in lane 3, but are mistakenly detected by camera left in lane 4 and by camera right in lane 2. And 2 cars in lane 2, which are surrounded by the green boxes or circles, are not detected by Camera Right. Through image fusion, all of the errors are revised.

TABLE 4: Experimental result about the IPU related to camera left in crowded traffic state.

Lane index	Lane 1	Lane 2	Lane 3	Lane 4	Total
Actual traffic volume (unit: number of vehicles)	4	4	3	0	11
Number of vehicles detected by IPU	4	4	3	1	12
Number of missed vehicles	0	0	0	0	0
Number of false-segmented vehicles	0	0	0	1	1
Detection error rate	0	0	0	—	9.1%
Statistics error rate	0	0	0	—	9.1%
The average computing time	About 32.5 ms per fame				

TABLE 5: Experimental result about the IPU related to camera right in crowded traffic state.

Lane index	Lane 1	Lane 2	Lane 3	Lane 4	Total
Actual traffic volume (unit: number of vehicles)	4	4	3	0	11
Number of vehicles detected by IPU	6	4	3	0	13
Number of missed vehicles	0	0	0	0	0
Number of false-segmented vehicles	2	0	0	0	2
Detection error rate	50%	0	0	0	18.2%
Statistics error rate	50%	0	0	0	18.2%
The average computing time (unit: s)	About 32.5 ms per fame				

According to Figure 29, experimental result with the method of PFM data fusion in fluent traffic state is summarized in Table 3.

(ii) *Crowded Traffic State.* From 12:00 to 12:15, the detecting region is in crowded state. So this paper randomly selects 12 s in



FIGURE 34: EPIs of camera right (I_{right}) generated by video frames from frame 2,175 to frame 2,475.

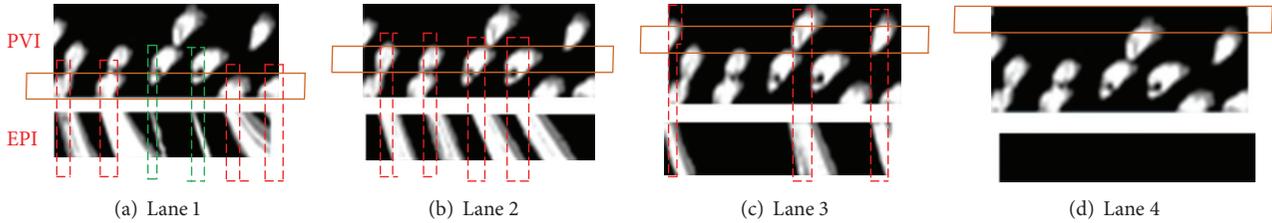


FIGURE 35: Vehicle existence detection of camera right generated by video frames from frame 2,175 to frame 2,475. In the figure, the meanings of color rectangular boxes are the same as those of Figure 23.

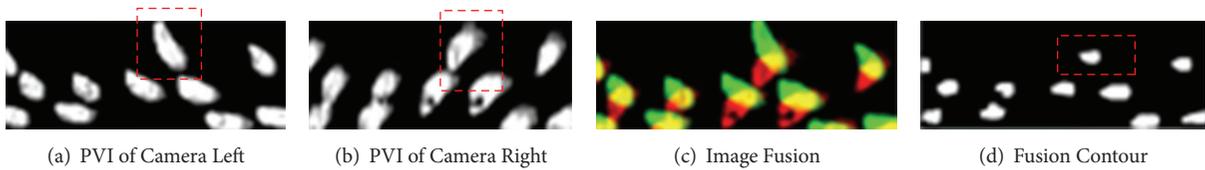


FIGURE 36: PVI generated by video frames from frame 2,175 to frame 2,475. In the figure, the meanings of color shapes are the same as those of Figure 27.

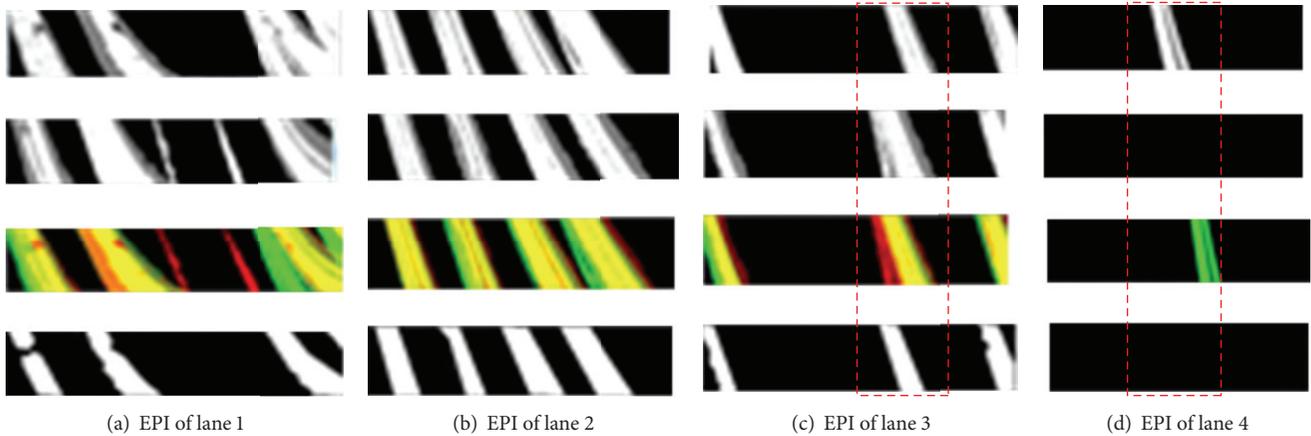


FIGURE 37: EPI of each road generated by video frames from frame 2,175 to frame 2,475. In the figure, the meanings of color shapes are the same as those of Figure 28.

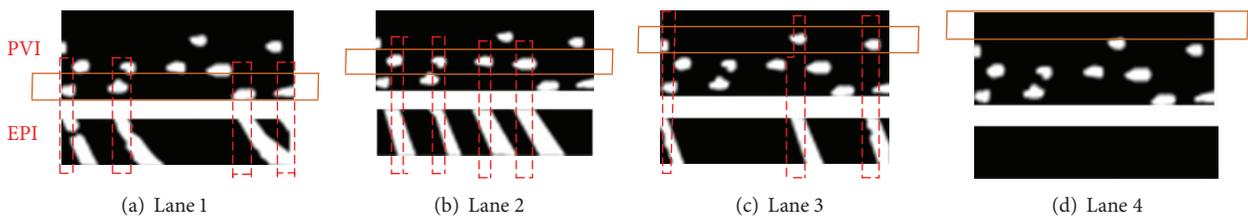


FIGURE 38: Vehicle existence detection with the method of PFM data fusion from frame 2,175 to frame 2,475. In the figure, the meanings of color rectangular boxes are the same as those of Figure 23.

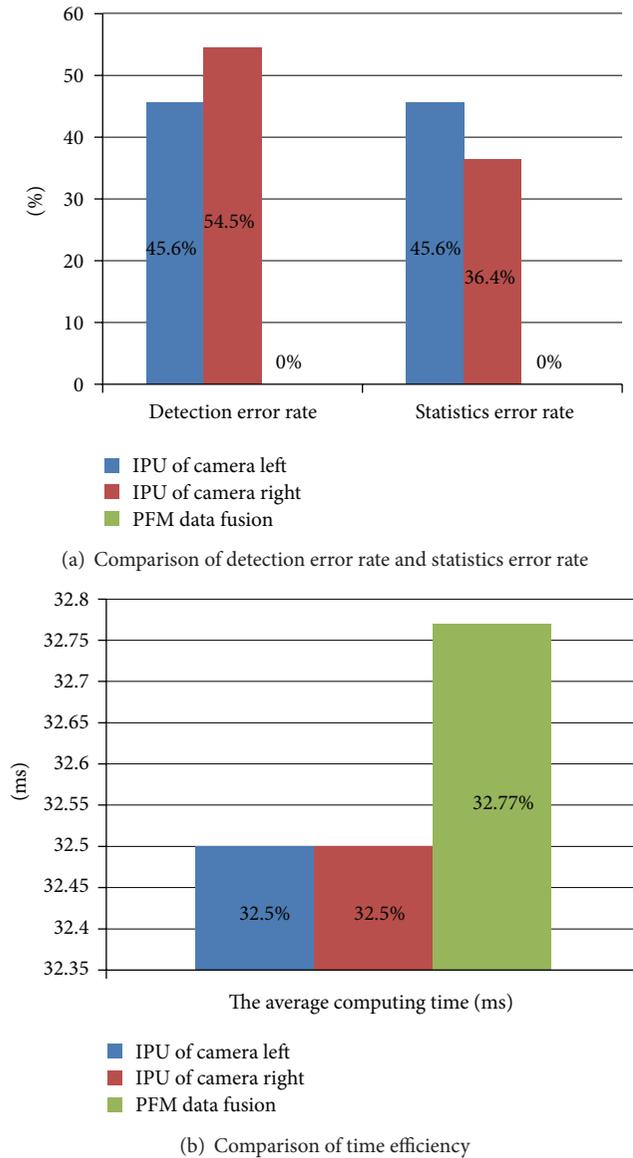


FIGURE 39: Comparisons of the methods about a single IPU of camera left, a single IPU of camera right, and PFM data fusion with SDF in fluent traffic state.

this period, from 12:01:27 to 12:01:39, corresponding to video frames from frame 2,175 to 2,475, to analysis the experimental results concerning the method of a single IPU and PFM data fusion.

(A) Experimental Results about a Single IPU.

(1) *Camera Left* (I_{left}). Figure 30 shows the PVI image of the detecting road, and Figure 31 shows the EPIs of each lane, all of which are generated by the IPU related to camera left (I_{left}). In Figure 32, the process and the result of vehicle existence detection of camera left are described.

According to Figure 32, experimental result about the IPU related to camera left in crowded traffic state is summarized in Table 4.

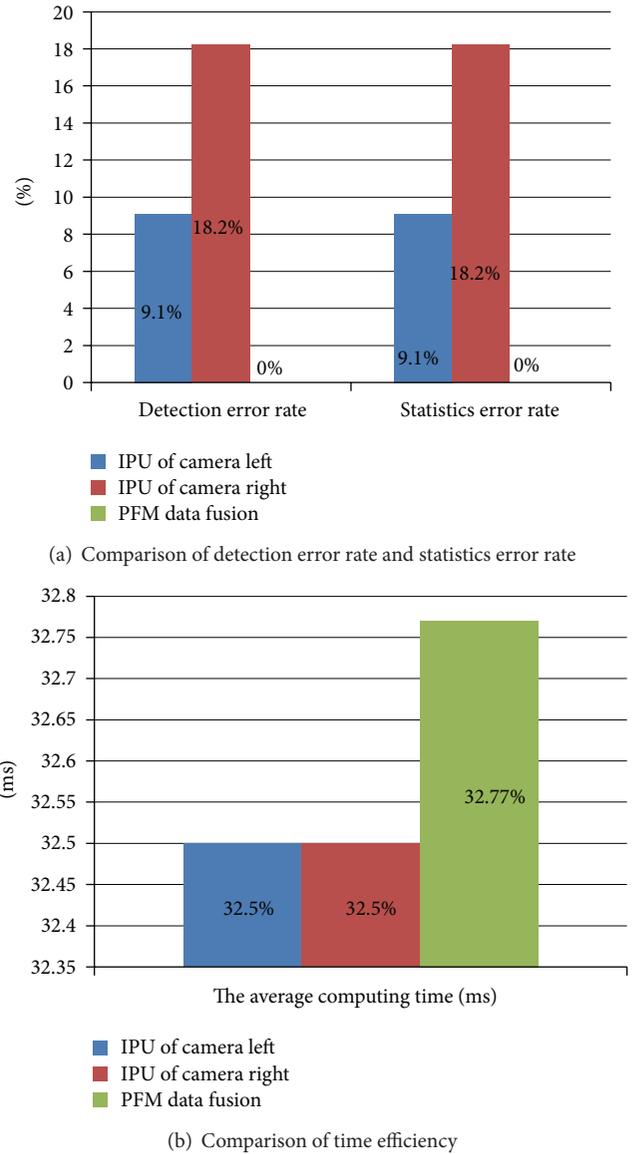


FIGURE 40: Comparisons of the methods about a single IPU of camera left, a single IPU of camera right, and PFM data fusion with SDF in crowded traffic state.

(2) *Camera Right* (I_{right}). Figure 33 shows the PVI image of the detecting road, and Figure 34 shows the EPIs of each lane, all of which are generated by the IPU related to camera right (I_{right}). In Figure 35, the process and result of vehicle existence detection of camera right are described.

According to Figure 35, experimental result about the IPU related to camera right in crowded traffic state is summarized in Table 2.

(B) *PFM Data Fusion with SDF*. Figures 36 and 37, respectively, described the process and the result of PFM data fusion about PVI of the detecting road, and the EPIs of each lane.

In Figures 36 and 37, the dotted red boxes denote a car in lane 3 but are mistakenly detected by camera left in lane 4. Through image fusion, the error is revised.

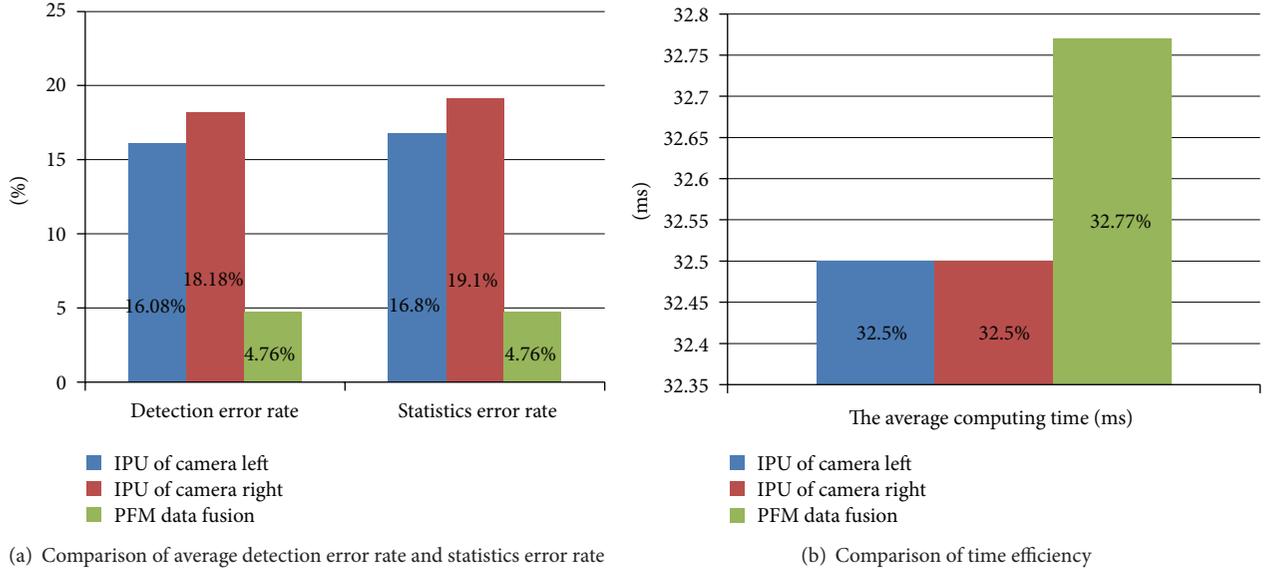


FIGURE 41: Comparisons of the methods about a single IPU of camera left, a single IPU of camera right, and PFM data fusion with SDF in crowded or in fluent traffic state.

TABLE 6: Experimental result with the method of PFM data fusion in crowded traffic state.

Lane index	Lane 1	Lane 2	Lane 3	Lane 4	Total
Actual traffic volume (unit: number of vehicles)	4	4	3	0	11
Number of vehicles detected by IPU	4	4	3	0	11
Number of missed vehicles	0	0	0	0	0
Number of false-segmented vehicles	0	0	0	0	0
Detection error rate	0	0	0	0	0
Statistics error rate	0	0	0	0	0
The average computing time	About 32.72 ms per fame				

According to Figure 38, experimental result with the method of PFM data fusion in crowded traffic state is summarized in Table 6.

(iii) *Analysis of Experimental Results in Different Traffic State.* Based on the Tables 1, 2, and 3, Figure 39 reflects the comparisons of detection error rate, statistics error rate, and time efficiency concerning the method of a single IPU and PFM data fusion with SDF in fluent traffic state. Similarly, based on the Tables 4, 5, and 6, Figure 40 reflects the comparisons in fluent traffic state.

To a large extent, the error happens because of the deviation of calibration, which depends on the vehicle height, as discussed in Section 2.1. And the proposed method weakens the effect of vehicle height and shadow greatly. So the error rate is approximately zero with a relatively high time efficiency using the proposed method in this paper, as shown in Figures 39 and 40.

6.2.2. *Analysis of Experimental Results.* With three Intel i3-2310M 2.10 GHz and 2 GB RAM PC, the average computing time is about 32.5 ms per fame in the IPU and 0.22 ms more in the data fusion process. In regard to vehicle existence detection in a different traffic state, through plenty of tests under the experimental environment introduced in Section 6.1 of this paper, the average detection error rate is 16.08%, 18.18%, and 4.76% in camera left, camera right, and the fusion images, respectively, and the average statistics error rate is 16.80%, 19.10%, and 4.76%, respectively, as shown in Figure 41. Therefore, the proposed method can be implemented quickly and accurately either in crowded state or in fluent state.

7. Conclusions

Based on the traffic monitoring system consisting of IPUs, network communication, and SDF, this paper proposes an approach for traffic parameters extraction in the highway, such as vehicle number, vehicle velocity, and vehicle type, based on time-spatial image and image fusion. The experimental results show that the proposed method can be implemented quickly and accurately either in crowded state or in fluent state. Further research to test the performance of the proposed method in more diversified circumstances, such as nighttime, and to optimize the algorithms of image processing and data fusion should be done. In addition, combined with pattern recognition and data mining technology, this method can be used to realize the traffic state identification and prediction, and to provide basis for section optimization.

Acknowledgments

This work was supported in part by Projects of International Cooperation and Exchanges of Natural Science Foundation of China (NSFC) under Grants 61111130119 and NSFC

60904069, the Doctoral Fund of Ministry of Education of China under Grant 20091103120008.

References

- [1] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in *Proceedings of the 11th ACM International Conference on Multimedia (MM '03)*, pp. 2–10, November 2003.
- [2] T. Semertzidis, K. Dimitropoulos, A. Koutsia, and N. Grammalidis, "Video sensor network for real-time traffic monitoring and surveillance," *IET Intelligent Transport Systems*, vol. 4, no. 2, pp. 103–112, 2010.
- [3] A. Koutsia, T. Semertzidis, K. Dimitropoulos, N. Grammalidis, and K. Georgouleas, "Intelligent traffic monitoring and surveillance with multiple cameras," in *Proceedings of International Workshop on Content-Based Multimedia Indexing (CBMI '08)*, pp. 125–132, June 2008.
- [4] K. Fujimura, K. Toshihiro, and K. Shunsuke, "Vehicle infrastructure integration system using vision sensors to prevent accidents in traffic flow," *IET Intelligent Transport Systems*, vol. 5, no. 1, pp. 11–20, 2011.
- [5] F. Lamosa, Z. Flu, and K. Uchimura, "Vehicle detection using multi-level probability fusion maps generated by a multi-camera system," in *Proceedings of IEEE Intelligent Vehicles Symposium*, pp. 452–457, June 2008.
- [6] Z. Zhu, G. Xu, B. Yang, D. Shi, and X. Lin, "VISATRAM: a real-time vision system for automatic traffic monitoring," *Image and Vision Computing*, vol. 18, no. 10, pp. 781–794, 2000.
- [7] N. C. Mithun, N. C. Rashid, and S. M. M. Rahman, "Detection and classification of vehicles from video using multiple time-spatial images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1215–1225, 2012.
- [8] N. K. Kanhere and S. T. Birchfield, "A taxonomy and analysis of camera calibration methods for traffic monitoring applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 441–452, 2010.
- [9] Y. Shang, Q. Yu, and X. Zhang, "Analytical method for camera calibration from a single image with four coplanar control lines," *Applied Optics*, vol. 43, no. 28, pp. 5364–5369, 2004.
- [10] D. Lee and Y. Park, "Measurement of traffic parameters in image sequence using spatio-temporal information," *Measurement Science and Technology*, vol. 19, no. 11, Article ID 115503, 2008.
- [11] K. Dimitropoulos, N. Grammalidis, D. Simitopoulos, N. Pavlidou, and M. Strintzis, "Aircraft detection and tracking using intelligent cameras," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, pp. 594–597, September 2005.
- [12] D. Yang, L. Xin, Y. Chen, Z. Li, and C. Wang, "A robust vehicle queuing and dissipation detection method based on two cameras," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC '11)*, pp. 301–307, 2012.
- [13] L. Li, L. Chen, X. Huang, and J. Huang, "A traffic congestion estimation approach from video using time-spatial imagery," in *Proceedings of the 1st International Conference on Intelligent Networks and Intelligent Systems (ICINIS '08)*, pp. 465–469, November 2008.
- [14] G. Litos, X. Zabulis, and G. Triantafyllidis, "Synchronous image acquisition based on network synchronization," in *Proceedings of IEEE Workshop on Three-Dimensional Cinematography*, June 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

