Robust Vehicle and Traffic Information Extraction for Highway Surveillance

Akio Yoneyama

Multimedia Communications Lab., KDDI R&D Laboratories Inc., 2-1-15 Ohara, Saitama 356-8502, Japan
Email: yoneyama@spg.kddlabs.co.jp

Chia-Hung Yeh

Department of Electrical Engineering and Integrated Media Systems Center, USC Viterbi School of Engineering,
University of Southern California, Los Angeles, CA 90089-2564, USA
Email: chyeh@sipi.usc.edu

C.-C. Jay Kuo

Department of Electrical Engineering and Integrated Media Systems Center, USC Viterbi School of Engineering,
University of Southern California, Los Angeles, CA 90089-2564, USA
Email: cckuo@sipi.usc.edu

A robust vision-based traffic monitoring system for vehicle and traffic information extraction is developed in this research. It is challenging to maintain detection robustness at all time for a highway surveillance system. There are three major problems in detecting and tracking a vehicle: (1) the moving cast shadow effect, (2) the occlusion effect, and (3) nighttime detection. For moving cast shadow elimination, a 2D joint vehicle-shadow model is employed. For occlusion detection, a multiple-camera system is used to detect occlusion so as to extract the exact location of each vehicle. For vehicle nighttime detection, a rear-view monitoring technique is proposed to maintain tracking and detection accuracy. Furthermore, we propose a method to improve the accuracy of background extraction, which usually serves as the first step in any vehicle detection processing. Experimental results are given to demonstrate that the proposed techniques are effective and efficient for vision-based highway surveillance.

Keywords and phrases: traffic monitoring, object tracking, moving cast shadow, occlusion, nighttime detection, background subtraction.

1. INTRODUCTION

Vision-based traffic monitoring systems are widely used in intelligent transportation systems (ITS). The goal of a traffic monitoring system is to extract traffic information, such as the vehicle count, traffic events, and traffic flow, which plays an important role for traffic analysis and traffic management. Several different types of devices, including loop detectors, sensors, and cameras, have been employed in traffic monitoring systems. Recently, vision-based analysis systems have become popular in transportation management due to their capability to extract a wide variety of information in comparison with the sensor-based system.

Vision-based systems have a good potential for highway surveillance applications [1, 2], and useful traffic information such as vehicle dimensions, lane changes, and other traffic-related information can be effectively extracted. However, it is challenging to maintain detection accuracy at all time since vision-based processing is sensitive to environmental factors such as lighting, shadow, and weather conditions. The following factors tend to result in the degradation of detection performance.

(i) Shadow

Moving region extraction is one of the fundamental steps in object detection and tracking from images. The moving cast shadow is often extracted with the vehicle of interest. The existence of the cast shadow in the moving object degrades extraction accuracy. For example, one solution to identify the vehicle region is based on model matching. The matching result can be severely affected by the existence of the cast shadow. There have been methods proposed to eliminate the cast shadow, say, to find different characteristics between shadow regions and nonshadow regions. This is however a difficult task by itself, and it may not always work properly.

(ii) Occlusion

Occlusion occurs when there is an overlapping of objects from the viewpoint of a camera. In this case, multiple vehicles

in the occlusion region are regarded as one object (vehicle). Generally speaking, occlusion prevents us from proper identification of each vehicle. The occlusion phenomena in a traffic scene can be divided into two classes: explicit occlusion and implicit occlusion. In the explicit occlusion case, multiple vehicles that enter a scene separately merge into a moving object region in the scene. In the implicit occlusion case, multiple vehicles enter the scene of being overlapped with each other from the beginning. Detection of explicit occlusion is easier than that of implicit occlusion since we can use the information of individual vehicles before occlusion occurs. In the case of implicit occlusion, we cannot obtain information of individual vehicles in the occluded region beforehand. It is possible to analyze the behavior of vehicles with the temporally reverse direction using the offline processing. This is however not suitable for real-time applications.

(iii) Lighting

The lighting condition changes by the weather, light sources, and other factors such as daytime or nighttime. It is difficult to cope with all kinds of situations with a single approach. Vehicle detection at nighttime is a complicated issue due to the lack of valuable visual information. Besides, there are disturbances caused by headlights, their corresponding reflections, and other light sources. In the real world, nighttime vehicle detection is usually assisted with some other techniques such as streetlights or infrared cameras. Conventional nighttime vehicle detection methods focus on finding pairs of headlights to estimate vehicle locations. However, they cannot extract other information such as the vehicle length. Moreover, results could be affected by the reflection of lights on the pavement.

In this work, we propose several new methods to cope with these three problems for a more robust highway surveillance system. First, we develop an algorithm to eliminate the moving cast shadow based on a simplified 2D vehicle-shadow model of six types projected to the 2D image plane. Second, images from multiple cameras for each scene are employed for occlusion detection. Third, for nighttime detection, we introduce the rear-view monitoring concept and develop a vehicle detection method to extract the vehicle information such as the vehicle width and length. In addition, we propose an enhanced method to quickly adjust the background according to the environmental change.

There are several advantages of our proposed techniques. They demand no prior information about the video capturing conditions such as the camera angle, the camera calibration, or the traffic scene description. Especially, a light source is not needed in the proposed shadow elimination algorithm. Edge detection and region segmentation are also not needed in the proposed occlusion detection algorithm. No additional equipment is required for nighttime vehicle detection and tracking. Moreover, the proposed enhanced background maintenance technique can provide high-accuracy background subtraction results. Finally, we show detection and tracking accuracy under various scene conditions to demonstrate the robustness of the proposed system with captured real-world data.

The rest of this paper is organized as follows. Related previous work is reviewed in Section 2. The basic vehicle model and the proposed vehicle detection algorithms are presented in Section 3. Simulation results are provided to demonstrate the robust performance of the proposed algorithms in Section 4. Finally, concluding remarks and future research directions are given in Section 5.

2. REVIEW OF PREVIOUS WORK

In this section, we review various video processing algorithms proposed for vision-based traffic monitoring systems. Background subtraction techniques are commonly used as the first step in separating foreground objects from the background region. Various issues related to background maintenance were discussed by Toyama et al. [3]. One major problem is the blending of foreground and background. Another one is delay of background update when there are changes in the background region. Processing tasks in [3] can be performed in the pixel, region, and frame levels. As to the background model, Wren et al. [4] applied the single Gaussian model to each pixel. The single Gaussian model is appropriate for scenes with almost static background. Haritaoglu et al. [5] employed the kernel density estimation method to cope with a varying background such as waving trees. Stauffer and Grimson [6] took a mixture Gaussian model to develop an alternative criterion for fast processing using temporal differences. Gutchess et al. [7] compared various background initialization algorithms for seven typical scenes.

There are several techniques proposed to differentiate moving cast shadows from background and vehicles. Background subtraction methods provide a simple, yet useful solution to moving region extraction since background is almost static for a short period of time. Moreover, solutions using a frequent update of the background image have been proposed to cope with gradual environmental change and noise [3]. However, extracted regions from the background subtraction method often include the cast shadow. It needs a further processing step to eliminate cast shadows from extracted moving regions. Cucchiara et al. [8] proposed a shadow suppression technique based on the characteristics of the Hue-saturation-value (HSV) color space. Another way to detect moving cast shadows was proposed by Fung et al. [9], where an indicator called the shadow confidence score was calculated in the luminance, chrominance, and gradient density. They also employed the Canny edge detector for edge detection and image segmentation, and extended their work to the visual-based dimension estimation by the 3D simple model in [10]. Prati et al. [11] presented the taxonomy of shadow detection methods according to gray level/color, local and static parameters, region and dynamic features, and performed a comparative study. It is obvious that moving shadow detection and elimination is one of the key research topics in vision-based highway surveillance systems. However, most methods have their shortcomings and
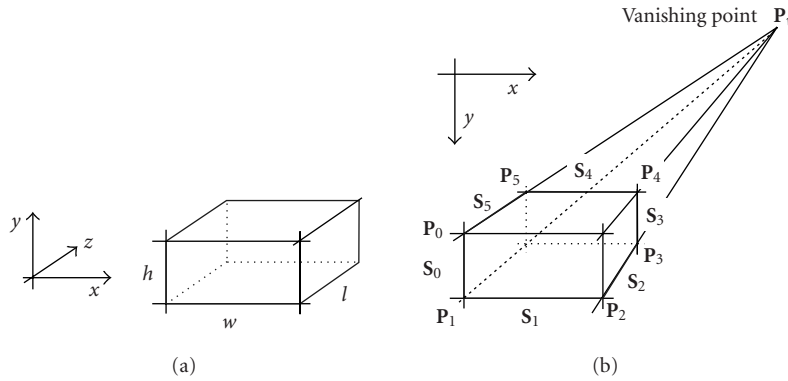
FIGURE 1: (a) The simplified vehicle model in the 3D space, and (b) the simplified vehicle model in the 2D image plane.

limitations. For algorithms that perform region segmentation and boundary detection in the pixel domain, it is sometimes difficult to discriminate the cast shadow from the self-shadow since the characteristics for the self-shadow region and the cast shadow region are very similar.

Cucchiara et al. [2] provided a traffic monitoring method by combining a low-level (pixel-based) processing technique and a high-level (knowledge-based) reasoning method, where occlusion reasoning helps estimate each vehicle in the occlusion region effectively. The reasoning rules explored vehicle features and expert rules such as "no change in the size," "inertia," and "traffic laws and rules." Another occlusion reasoning method was proposed by Jung and Ho [12]. The use of multiple cameras may be helpful to occlusion detection, for example, the stereo-vision method. With stereo vision, one can understand the 3D space using parallax. However, approaches using multiple cameras demand a way to establish the correspondence between images. Even though it is sometimes possible to find feature points among images taken from different cameras, to find the correspondence between images is generally a difficult task.

More recently, Gupte et al. [13] employed a monocular vision-based detection technique in three levels: the raw image level, the region level, and the vehicle level. Vehicles were modeled as rectangular patches. They described the interactive camera calibration tool, which demands only camera calibration parameters and the traffic direction to achieve the goal. Kato et al. [14] presented a thorough method to segment traffic-monitoring images into three types, namely, "foreground," "background," and "cast shadow." Since the intensity distributions of these three types have a large amount of overlap, it is difficult to learn these distributions separately. They applied the 1D HMM (where the 1D means the temporal direction) to each subblock in the image, which is a constrained HMM that may prohibit specific transitions. Wavelet coefficients in high-frequency bands are used to determine "foreground," "background," or "shadow region." They also took context-dependence among nearby regions into account.

Traffic monitoring systems have to function properly any time and under any weather conditions. Nighttime traffic monitoring is however very different from that of daytime.

For instance, the background subtraction technique cannot be used to extract moving vehicle regions due to the effects of headlights and their reflections on the pavements. A possible solution is to detect headlight positions and estimate vehicle locations [15, 16, 17]. Generally speaking, detection accuracy at nighttime is lower than that at daytime due to various illumination disturbances and the lack of useful visual information.

3. PROPOSED VISION-BASED SURVEILLANCE TECHNIQUES AND SYSTEMS

In this work, we investigate techniques to enhance the performance robustness of highway surveillance systems in the areas of cast shadow elimination, occlusion detection, and nighttime detection. The first topic is the 2D joint vehicle-shadow model establishment for moving cast shadow elimination, which is explained in Section 3.1. The second topic is to use the multiple-camera system to deal with occlusion problems, which is described in Section 3.2. The third topic is to overcome the nighttime detection via the rear-view monitoring technique, which is presented in Section 3.3. Section 3.4 describes an enhanced background maintenance technique to improve detection accuracy. Finally, the overall system is summarized in Section 3.5.

3.1. Moving cast shadow elimination

We first consider the simplified cuboid model $\mathbf{M}$ in the 3D space that is composed of the width, height, and length of a vehicle as shown in Figure 1a. Figure 1b illustrates how model $\mathbf{M}$ is transformed into $\mathbf{M}'$ in the 2D image plane. The model has six vertices $\mathbf{M}' = \{p_k \mid k = 0, \ldots, 5\}$. In some setting of the camera axis, the model in the 2D image plane can be approximated using the following features:

$$S0\|S3, \qquad S1\|S4, \qquad S2\|S5. \tag{1}$$

It is desirable to capture the long range of the road so that we can extract the behavior of vehicles such as lane changes. Consequently, the vertical angle of the camera axis needs to be set to near the horizontal direction. Figure 2 shows the vertical angle of the camera axis and the horizontal axis of
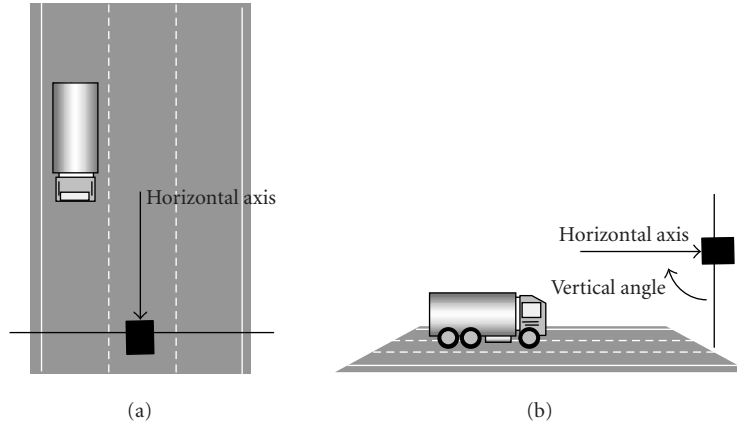
FIGURE 2: (a) The vertical angle of the camera axis, and (b) the horizontal axis of the camera.

the camera. Under this consideration, we can set our vehicle model to have a single vanishing point as shown in **Figure** 1b. The position of the vanishing point $\mathbf{p}_v$ is defined as the intersection of $\mathbf{S}_2$ and $\mathbf{S}_5$ in the 2D image plane. Other sides in the model have the following feature:

$$S0\|S3\| \quad y\text{-axis}, \qquad S1\|S4\| \quad x\text{-axis}. \tag{2}$$

This feature can be obtained by setting the horizontal axis of the camera perpendicular to the traffic flow. Thus, the traffic flow direction is along the z-direction as shown in **Figure** 1a. This approximation simplifies the system while maintaining detection accuracy. Shadow elimination must be done in advance to vehicle detection in most existing methods.

Being different from other approaches, the shadow model is included in the 2D image plane as shown in **Figure** 3a in our framework. The shadow of an object (e.g., the vertical pole) is decomposed into the direction of its width and length. In the shadow model, we use three vectors $\mathbf{s}_h$, $\mathbf{s}_l$, and $\mathbf{s}_w$ to represent "the height of an object," "the length component of the shadow," and "the width component of the shadow," respectively. According to the location of the cast shadow on the 2D image plane, we can divide the shadow model into six types as shown in **Figure** 3b. The orientation of three lines in **Figure** 3b is the same as the orientation of $\mathbf{s}_h$, $\mathbf{s}_l$, and $\mathbf{s}_w$ in the shadow model as given in **Figure** 3a.

Figure 4a shows the image of the joint vehicle-shadow model on the 2D image plane. The outer bounding box in each type represents the fitted six-vertex model for a vehicle that includes the cast shadow. The inner bounding box in each type represents the fitted six-vertex model for a vehicle without the cast shadow. It is clear that there is at least one side whose location and length are not changed by the existence of the cast shadow. We can distinguish the model type using the location of the nonchanging side. We denote this side by $\mathbf{S}_f$, where $\mathbf{S}_f$ is shown by a thick line for each type in **Figure** 4a. Since both the vehicle model and the shadow model are on the 2D image plane, the joint vehicle-shadow

model is also on the 2D image plane. **Figure** 4b shows the relationship between the lengths of sides of the joint model and those of the cast shadow model in **Figure** 1a.

As illustrated in **Figure** 4b, the relationship between sides of $\mathbf{M}'$ for the 2D vehicle model only and $\mathbf{M}'_J$ for the 2D joint vehicle-shadow model can be expressed as

$$\begin{aligned}
\left|\mathbf{S}_{f+}(\mathbf{M}')\right| &= \left|\mathbf{S}_{f+}(\mathbf{M}'_J)\right| - r_+\left|\mathbf{S}_f(\mathbf{M}')\right|, \\
\left|\mathbf{S}_{f-}(\mathbf{M}')\right| &= \left|\mathbf{S}_{f-}(\mathbf{M}'_J)\right| - r_-\left|\mathbf{S}_f(\mathbf{M}')\right|,
\end{aligned} \tag{3}$$

where

  (i) $\mathbf{S}_{f+}(\mathbf{M}')$ is the side that is adjacent to $\mathbf{S}_f$ of 2D model $\mathbf{M}'$ in the ascending order,
 (ii) $\mathbf{S}_{f-}(\mathbf{M}')$ is the side that is adjacent to $\mathbf{S}_f$ of 2D model $\mathbf{M}'$ in the descending order,
(iii) r_+ is the ratio of shadow model parameters between $|\mathbf{S}_{f+}|$ and $|\mathbf{S}_f|$,
 (iv) r_- is the ratio of shadow model parameters between $|\mathbf{S}_{f-}|$ and $|\mathbf{S}_f|$.

We take Type-0 in **Figure** 4b as an example, the vehicle height can be found as $\mathbf{S}_f$. If we know r_+ and r_- on the 2D image plane, the vehicle width and length can be obtained via $\mathbf{S}_{f+}(\mathbf{M}')$ and $\mathbf{S}_{f-}(\mathbf{M}')$, respectively. Equation (3) implies that we can determine the vehicle model $\mathbf{M}'$, if (1) the model type is determined, (2) the ratios r_+ and r_- of shadow model parameters are known, and (3) the parameters of the joint vehicle-shadow model can be successfully measured using the background subtraction method. We examine conditions (1) and (2) in the following.

The six model types depend on the relative locations of the light source, the vehicle, and the camera as shown in **Figure** 4a. The location of $\mathbf{S}_f$ in the model indicates the light source direction on the 2D image plane. By utilizing the feature that the side has neither self-shadow nor cast shadow, we can determine $\mathbf{S}_f$ by examining the luminance in each side of the six-vertex model and finding the side that has the highest
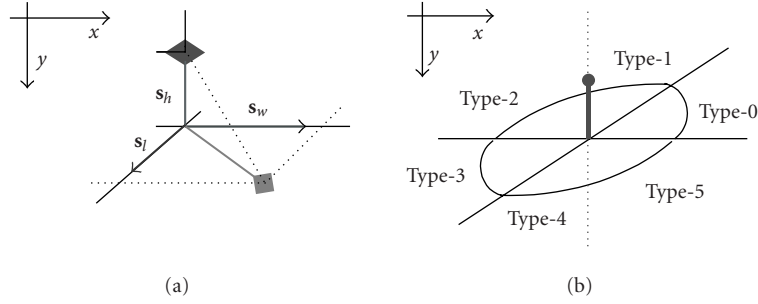
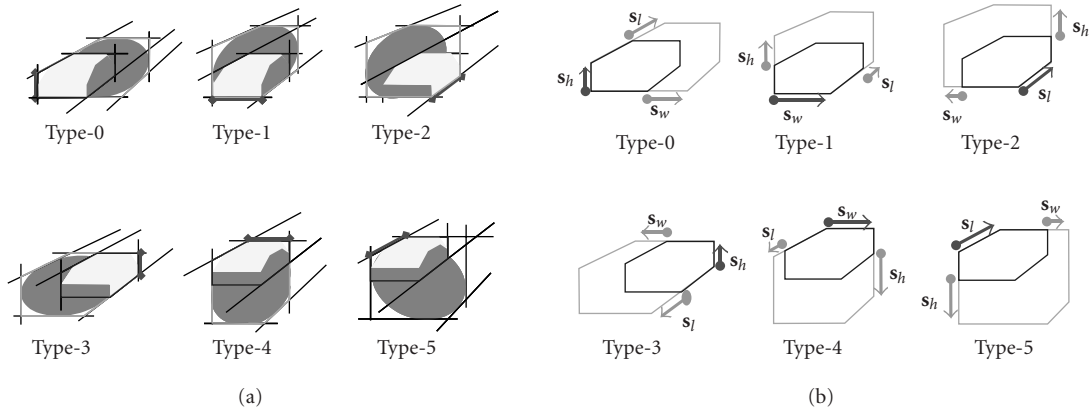FIGURE 3: (a) The shadow model, and (b) six cast shadow types according to different lighting directions.

FIGURE 4: (a) Six joint vehicle-shadow types, and (b) their simplified configurations.

luminance in average. This process can be expressed mathematically as

$$L_n = \frac{1}{N} \sum_{k=0}^{N-1} \text{lum}_k, \quad \{k \mid (x, y) \in \mathbf{S}_{j,n}\},$$
$$L_{\max} = \max (L_n : \mathbf{S}_{j,n} \in \mathbf{M}'_J), \tag{4}$$

where N denotes the number of pixels in $\mathbf{S}_{j,n}$ and lum_k represents the luminance value of position (x, y) in the captured image. Shadow model parameters $\mathbf{s}_l$, $\mathbf{s}_h$, and $\mathbf{s}_w$ can be determined by the length of the cast shadow. Although it is difficult to obtain the accurate cast shadow area from all passing vehicles due to various environmental factors, it is possible to roughly differentiate the cast shadow region and the vehicle with luminance differences. One such process is shown in Figure 5a. Then, ratios r_+ and r_- among the three lengths $\mathbf{s}_l$, $\mathbf{s}_h$, and $\mathbf{s}_w$ can be obtained.

There are several key issues to be addressed in the proposed algorithm such as "determination of the vanishing point," "determination of the moving object contour," and "determination of the shadow model parameters." They are discussed in detail below.

Determination of the vanishing point

A simple approach to obtain the vanishing point $\mathbf{p}_v$ is to extract the lane orientation in the background image and edge detection provides a solution to this problem. Another way to achieve this is through camera calibration. However, the approach may not be applicable if there is no obvious lane painting detectable from the image. To cope with such situations, we use the feature that most vehicles pass the scene parallel to the lanes to determine the vanishing point since we do not assume the availability of any prior information about the camera and the light source. Even though, we cannot determine the cast shadow region because the shadow model in the 2D image plane is not fixed yet, we can still obtain the centroid of the six-vertex joint vehicle-shadow model. Then, we can track vehicles by connecting the centroid in sequential images. It should be noted that the existence of the cast shadow region does not affect the calculation of the vanishing point. At first, we obtain the tracking information in the 2D image plane using multiple detected vehicles, then the position of the vanishing point $\mathbf{p}_v$ can be set to the intersection of these tracking lines.

Determination of the moving object contour

In Section 3.1, the joint-model-type can be obtained by checking the average luminance of six sides of model $\mathbf{M}'_J$. However, most pixel values on the six-vertex model are in the background region. They may not lie along the contour of the extracted moving object region as shown in Figure 5b. Thus, it is desirable to find the appropriate pixel on the contour of the extracted moving object region to replace the value of pixels of the sides of the six-vertex model to calculate
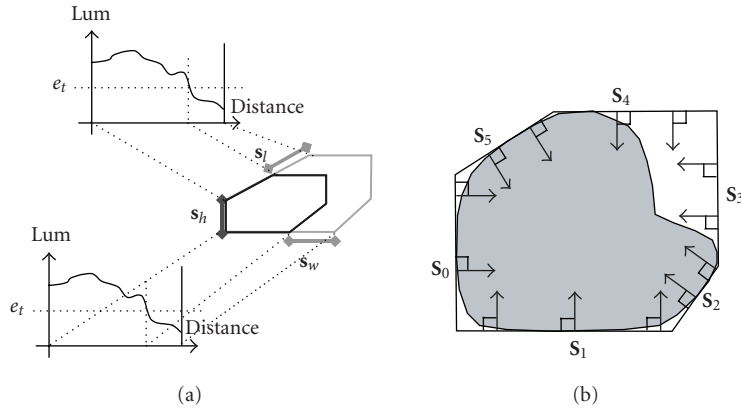
FIGURE 5: (a) Determination of shadow model parameters, and (b) obtaining the pixel on the object contour.

the average luminance. For each pixel on a side $\mathbf{S}_n$ of joint vehicle-shadow model $\mathbf{M}'_j$, we search an appropriate pixel (x, y), which is located nearest to the boundary and along the direction perpendicular to $\mathbf{S}_n$ as shown in Figure 5b.

Determination of the shadow model parameters

In the process of shadow parameter determination, the first step is to find appropriate moving objects from the scene. It is difficult to perform shadow region segmentation from the image with low contrast. To obtain a vehicle from images of higher contrast, we pick up vehicles whose average luminance of $\mathbf{S}_f$ is higher than a predetermined threshold e_p. The next step is to find the length of the shadow along each direction of the image. The basic idea is given in Figure 5a, where we need a mechanism to determine two threshold values to separate the cast shadow region from the vehicle region along two dimensions. In this work, the threshold value e_t is set to the average value of the side. We find point p_b whose luminance crosses threshold e_t, and then calculate the distance between p_b and the end of the side, which is equal to $\mathbf{s}_w$ or $\mathbf{s}_l$. In comparison to conventional approaches [11] that compare pixel values in the current image with those in the background image, our proposed scheme employs only the current image. By picking up vehicles with bright colors and using the characteristics of the vehicle-shadow model shown in Figure 5a, we can easily obtain lengths $\mathbf{s}_l$, $\mathbf{s}_w$, $\mathbf{s}_h$ and the ratio $\mathbf{s}_l : \mathbf{s}_w : \mathbf{s}_h$. Thus, with our method, it is unnecessary to compare the pixel values in the current frame with those in the reference background.

3.2. Occlusion detection

In this section, we focus on the detection of implicit occlusion. A temporal reversed processing could be a solution to the implicit occlusion problem. However, not every implicit occlusion can be solved by this approach, for example, the occluded situation may continue throughout the scene. Here, we propose a solution to implicit occlusion detection using multiple cameras. The proposed method is based on a perspective projection of multiple images in the 2D road surface plane. For simplicity, we only project the centers, which are

obtained from both images $\mathbf{I}_p$ and $\mathbf{I}_s$, to the projected plane $\mathbf{I}_r$ as shown in Figure 6a. We can detect implicit occlusion from the mismatch between the representative centers. With this approach, the projected road regions in both images $\mathbf{I}_p$ and $\mathbf{I}_s$ must be identical so that we can determine occlusion from the mismatch in the projected plane $\mathbf{I}_r$.

The perspective projection $\mathbf{P}$ from points $p_p = (x_p, y_p)$ in the 2D image plane $\mathbf{I}_p$ to their corresponding points $p_r = (x_r, y_r)$ in the road surface image plane $\mathbf{I}_r$ can be represented by $\mathbf{I}_r = \mathbf{P}(\mathbf{I}_p)$ and realized by

$$x_r = \frac{c_{00}x_p + c_{01}y_p + c_{02}}{c_{20}x_p + c_{21}y_p + 1}, \qquad y_r = \frac{c_{10}x_p + c_{11}y_p + c_{12}}{c_{20}x_p + c_{21}y_p + 1}. \quad (5)$$

In the perspective projection, at least four pair coordinates of the corresponding points between the 2D image plane and the projected road surface plane are needed to determine all parameters $\{\mathbf{C} : c_{00}, c_{01}, \ldots, c_{21}\}$ as shown in (5). From now on, the projection from $\mathbf{I}_p$ to $\mathbf{I}_r$ is denoted by $\mathbf{P}_{pr}$, and the projection from $\mathbf{I}_s$ to $\mathbf{I}_r$ is denoted by $\mathbf{P}_{sr}$. The coefficients for projections $\mathbf{P}_{pr}$ and $\mathbf{P}_{sr}$ are denoted by $\mathbf{C}_{pr}$ and $\mathbf{C}_{sr}$, respectively.

For each image plane given in Figure 6a, the four feature points and a vanishing point are represented by $\mathbf{p}_{k1}$, $\mathbf{p}_{k2}$, $\mathbf{p}_{k3}$, $\mathbf{p}_{k4}$, and $\mathbf{p}_{kv}$ in Figure 6b. The first subscript k represents the image plane ($\mathbf{I}_k : k = p, r, s$) and the second subscript, with values equal to 1, 2, 3, 4, and v, stands for the feature point ID. These feature points are often specified with road lane paintings. However, the approach may not be suitable if the camera cannot detect any clear lane painting. Since our goal is to detect occlusion by matching from images $\mathbf{I}_p$ and $\mathbf{I}_s$, it is enough to find three corresponding feature points for consistence check between projected image regions. The three pairs are chosen to be $\mathbf{p}_{p1}$, $\mathbf{p}_{s1}$ (the near left point), $\mathbf{p}_{p2}$, $\mathbf{p}_{s2}$ (the near right point), and $\mathbf{p}_{pv}$, $\mathbf{p}_{sv}$ (the vanishing point). The other coordinates for the setting of camera positions can be obtained as

$$\mathbf{P}_{k3} = (1 - a) \cdot \mathbf{p}_{k1} + a \cdot \mathbf{p}_{kv}, \qquad \mathbf{P}_{k4} = (1 - a) \cdot \mathbf{p}_{k2} + a \cdot \mathbf{p}_{kv}. \quad (6)$$
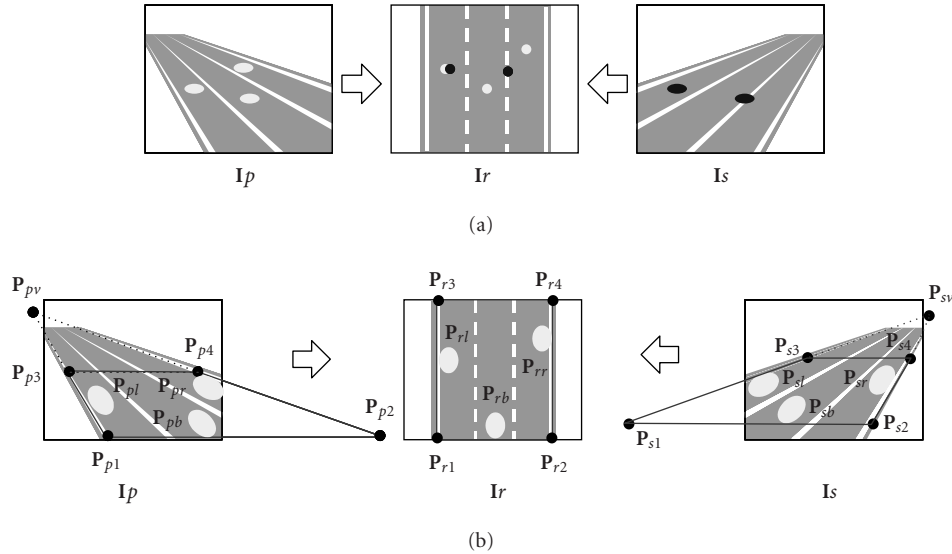
(a)

(b)

FIGURE 6: (a) The concept of matching in a projected image plane, and (b) the correspondence of feature points in images.
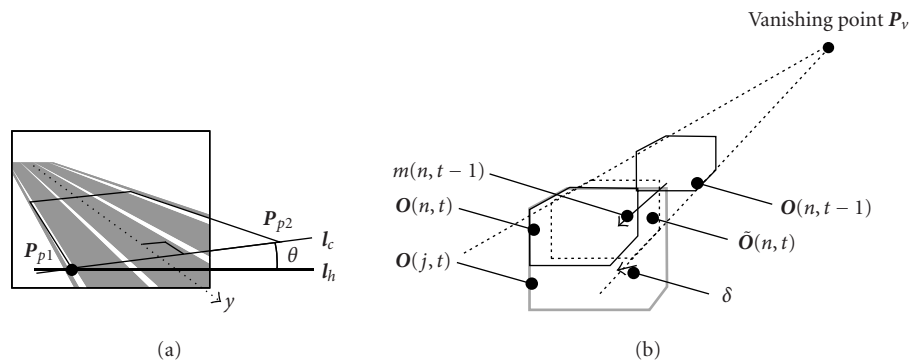

(a)

(b)

FIGURE 7: (a) The angle difference between a horizontal line $\mathbf{l}_h$ and a cross-traffic line $\mathbf{l}_c$ in image $\mathbf{I}_p$, and (b) explicit occlusion detection.

Equation (6) indicates that positions of $\mathbf{p}_{k3}$ and $\mathbf{p}_{k4}$ can be derived from positions of $\mathbf{p}_{k1}$, $\mathbf{p}_{k2}$, $\mathbf{p}_{kv}$, and parameter "a." Once these three corresponding points $\mathbf{p}_{k1}$, $\mathbf{p}_{k2}$, and $\mathbf{p}_{kv}$ in images $\mathbf{I}_p$ and $\mathbf{I}_s$ are determined, we can match these two images in the same $\mathbf{I}_r$, which has predefined positions of four feature points $\mathbf{p}_{r1}$, $\mathbf{p}_{r2}$, $\mathbf{p}_{r3}$, and $\mathbf{p}_{k4}$. Parameter "a" can be set arbitrarily between 0 and 1. When this value is set near to 1, the projected road range becomes longer in the direction of traffic. It is desirable for long-range object tracking. However, the region near the vanishing point is susceptible to noise. The value of 0.75 is adopted in our experiments.

The vanishing points $\mathbf{p}_{pv}$ and $\mathbf{p}_{sv}$ in the captured images $\mathbf{I}_p$ and $\mathbf{I}_s$ are determined by vehicle tracking. We can track vehicles by connecting the center of the detected moving object in the image sequence. Please note that the cast shadow region does not affect the calculation of the vanishing point. At first, we obtain the tracking information on the 2D image plane using multiple detected vehicles. Then, the location of vanishing point $\mathbf{p}_{kv}$ can be obtained by the intersection of

these tracking lines. The next step is the determination of feature points $\mathbf{p}_{k1}$ and $\mathbf{p}_{k2}$ in images $\mathbf{I}_p$ and $\mathbf{I}_s$. Figure 7a represents the relationship between the traffic lane angle and the image plane. A quadrilateral in the image represents the region to be projected. The line $\mathbf{l}_c$ represents a perpendicular line to the traffic flow on the road surface plane. The line $\mathbf{l}_h$ represents a horizontal line in the original image. Parameter θ is the angle difference between $\mathbf{l}_c$ and $\mathbf{l}_h$.

It is difficult for $\mathbf{p}_{p1}$ and $\mathbf{p}_{p2}$ to have the same coordinates with respect to the traffic axis (i.e., the y-axis) in Figure 7a from the image since the relationship between the left side and the right side of the road in the image is unspecified. Thus, the approximation $\mathbf{y}_{p1} \approx \mathbf{y}_{p2} \approx \mathbf{y}_{pb}$ is assumed for the determination of $\mathbf{p}_{p1}$ and $\mathbf{p}_{p2}$. The above assumption means that the angle difference θ is regarded as zero. The effect of this assumption can be ignored when the vanishing point locates within the 2D image plane. Yet, we still cannot determine feature points $\mathbf{p}_1$ and $\mathbf{p}_2$ from the image directly. We assume that $\mathbf{p}_{pl}$, $\mathbf{p}_{pr}$, and $\mathbf{p}_{pb}$ are the leftmost, the rightmost,

TABLE 1: The occlusion detection algorithm.

Conditions	Occlusion detection algorithm
Case 1: If $(d_1 < \text{Th}_{r1})$	Determined object location on $\mathbf{I}_r$: $\mathbf{o}_{pr}(n)$ Object locations to be deleted on $\mathbf{I}_r$: $\mathbf{o}_{sr}(m_1)$
Case 2: Else if $(d_1 > \text{Th}_{r2})$	Determined object location on $\mathbf{I}_r$: $\mathbf{o}_{pr}(n)$
Case 3: Else If $(d_1/d_2 > r_t)$	Determined object location on $\mathbf{I}_r$: $\mathbf{o}_{sr}(m_1)$, $\mathbf{o}_{sr}(m_2)$ Object locations to be deleted on $\mathbf{I}_r$: $\mathbf{o}_{pr}(n)$
Case 4: Else	Determined object location on $\mathbf{I}_r$: $(\mathbf{o}_{pr}(n) + \mathbf{o}_{sr}(m_1))/2$, $\mathbf{o}_{sr}(m_1)$.

and the bottommost of vehicles in a traffic scene in image $\mathbf{I}_p$, respectively, as shown in Figure 6b. The coordinates of $\mathbf{p}_{p1}$ and $\mathbf{p}_{p2}$ can be obtained via

$$
\begin{aligned}
x_{p1} &= x_{pv} + a_l(x_{pl} + x_{pv}), \quad a_l = \frac{y_{pb} - y_{pv}}{y_{pl} - y_{pv}}, \\
x_{p2} &= x_{pv} + a_r(x_{pr} + x_{pv}), \quad a_r = \frac{y_{pb} - y_{pv}}{y_{pr} - y_{pv}}.
\end{aligned}
\tag{7}
$$

When the four coordinates $\mathbf{p}_{p1}$, $\mathbf{p}_{p2}$, $\mathbf{p}_{p3}$, and $\mathbf{p}_{p4}$ are determined, we can obtain all coefficients of $\mathbf{C}_{pr}$ in (6) for image source $\mathbf{I}_p$. The coefficients of $\mathbf{C}_{sr}$ for image source $\mathbf{I}_s$ can be obtained with (7) similarly.

Implicit occlusion can be detected by checking the spatial difference between centers in the projected image. At first, background subtraction is used to extract the moving region and an identification number n is assigned to each region. Then, a bounding box and the center position of the bounding box are obtained for each detected moving object region. Let $\mathbf{o}_p(n)$ be the center position of bounding box n in image plane $\mathbf{I}_p$. Consequently, the projected center location $\mathbf{o}_{pr}(n)$ in image plane $\mathbf{I}_r$ is represented by $\mathbf{P}_{pr}(\mathbf{o_p}(n))$. For each $\mathbf{o}_{pr}(n)$, we find the first and the second nearest center positions $\mathbf{o}_{sr}(m1)$ and $\mathbf{o}_{sr}(m2)$ to $\mathbf{o}_{pr}(n)$ from image plane $\mathbf{I}_s$, and calculate the distances d_1 and d_2, where $d_1 \leq d_2$, via

$$
\begin{aligned}
d_1 &= ||\mathbf{o}_{sr}(m_1) - \mathbf{o}_{pr}(n) + \mathbf{d}_{\text{offset}}||_2, \\
d_2 &= ||\mathbf{o}_{sr}(m_2) - \mathbf{o}_{pr}(n) + \mathbf{d}_{\text{offset}}||_2.
\end{aligned}
\tag{8}
$$

The vector $\mathbf{d}_{\text{offset}}$ in (8) is used to adjust the deviation between projected centers from images $\mathbf{I}_p$ and $\mathbf{I}_s$. The value of $\mathbf{d}_{\text{offset}}$ is dependent on the three extracted feature points $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_v$ as described above. An effective selection of this value is to take the average difference between images. The occlusion detection algorithm is summarized in Table 1, where parameters r_t, Th_{r1}, and Th_{r2} are used for detection, where r_t is used to determine the occlusion decision threshold, Th_{r1} represents the permissible mismatch for the identification of the object, and Th_{r2} represents the maximum limit for occlusion detection. It is clear that $\text{Th}_{r1} \leq \text{Th}_{r2}$.

In Table 1, the statement "object locations to be deleted" in Case 1 means object locations to be removed from the vehicle recognition process. Object positions not mentioned in each case are finally regarded as the detected object locations. For Case 1, detected position pairs from images $\mathbf{I}_p$ and $\mathbf{I}_s$ are

identified as a single object. In this case, $\mathbf{o}_s(m_1, t)$ is removed from the detected vehicle position. For Case 2, an object detected in image $\mathbf{I}_p$ is not detected correctly in image $\mathbf{I}_s$, or the position is overdetected in image $\mathbf{I}_p$. For Case 3, there exist occlusion instances in image $\mathbf{I}_p$. In this case, the erroneously detected position $\mathbf{o}_p(n, t)$ is removed from the detected vehicle position. Finally, for Case 4, there exists some probability of occlusion. It occurs mainly when a large object enters the scene and more than two vehicles are located in the same occluded region. Since we cannot identify every vehicle from the scene, we estimate the mean of the two points as the main object location.

For explicit occlusion detection, it is possible to use the information already in previous images to estimate object locations in the current image since the size, dimension, and motion of each vehicle do not change in a short-time period. Figure 7b shows the idea of explicit occlusion detection in the original image sequence. The estimated location of an occluded vehicle is modified to match the result of background subtraction in the current image. Then, the information about the vehicle is updated. The steps for explicit occlusion detection are given below.

(1) Prediction of vehicles in the current image from the previous image.

We use $\mathbf{m}(k, t)$, $\mathbf{s}(k, t)$, and $\mathbf{o}(k, t)$ to denote the motion, size, and center position information of vehicle k in the 2D image plane at time t. With the inertia principle, the location and the size of the object in the 2D image plane can be predicted using vanishing point $\mathbf{p}_v$, that is, we have

$$
\begin{aligned}
\tilde{\mathbf{o}}(n, t) &= \mathbf{o}(n, t - 1) + \mathbf{m}(n, t - 1), \\
\tilde{\mathbf{s}}(n, t) &= a \cdot \mathbf{s}(n, t - 1), \\
\text{where} \quad a &= \frac{||\mathbf{p}_v - \tilde{\mathbf{o}}(n, t)||_2}{||\mathbf{p}_v - \mathbf{o}(n, t - 1)||_2}.
\end{aligned}
\tag{9}
$$

(2) Matching between the predicted location and detected results in the current image.

The predicted location of object n is matched with the region of detected object j whose location and size are denoted by $\mathbf{s}(j, t)$ and $\mathbf{o}(j, t)$, respectively.

(3) Modification of the current vehicle location from the matching result.

Let $\mathbf{O}(n, t)$ be the bounding box of object n at time t. If the predicted region $\mathbf{O}(n, t)$ of object n obtained from $\tilde{\mathbf{o}}(n, t)$ and $\mathbf{s}(n, t)$ is located within region $\mathbf{O}(j, t)$ of object j, the center location of object n is set to $\tilde{\mathbf{o}}(n, t)$. If region $\mathbf{O}(n, t)$ deviates from region $\mathbf{O}(n, t - 1)$ significantly, we modify the center location of object n so that it does not deviate from the predicted center more than δ, that is,

$$
\begin{aligned}
\mathbf{o}(n, t) &= \tilde{\mathbf{o}}(n, t - 1) + \delta, \\
\mathbf{s}(n, t) &= \tilde{\mathbf{s}}(n, t).
\end{aligned}
\tag{10}
$$

(4) Finally, the vehicle motion information is updated via

$$
\mathbf{m}(n, t) = \mathbf{o}(n, t) - \mathbf{o}(n, t - 1).
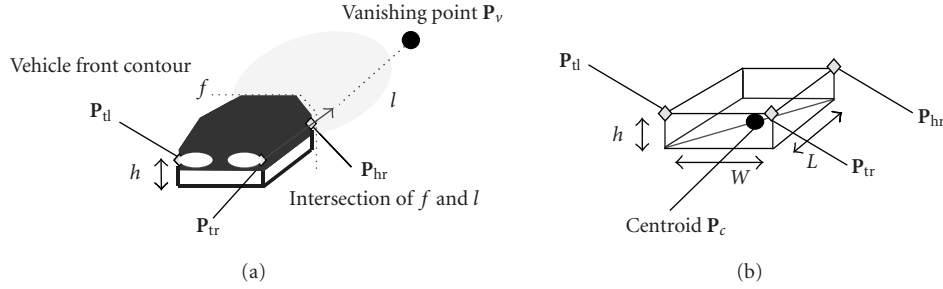\tag{11}
$$

FIGURE 8: (a) The rear-view vehicle model observed at nighttime, and (b) the relationship between feature points.

3.3. *Nighttime detection with rear-view monitoring*

Traffic monitoring systems are required to function properly under any time and weather conditions. Nighttime traffic monitoring imposes a great challenge, since most daytime detection methods have much poorer performance when being directly applied to nighttime detection. In previous work [15, 16, 17], vehicle detection at nighttime has been based on finding headlight locations and pairing the headlights. To find headlight locations, one has to distinguish headlights from other disturbances such as reflected lights on the pavement. The headlights are paired according to their spatial relationship.

In this section, we propose a rear-view detection approach, which has several advantages as compared with the conventional front-view detection approach. First, the rear-view approach improves nighttime detection performance due to the light disturbance reduction and utilization of the headlight information in another way. Second, the rear-view approach achieves superior tracking accuracy by exploiting the opposite traffic direction in video. Owing to the detection of both sides (front and tail), the rear-view monitoring can easily extract the vehicle length, which is one of the important traffic parameters especially for nighttime detection. Moreover, the rear-view system can reduce the disturbances caused by reflected headlights on the pavement because the camera axis does not face the strong headlights of oncoming vehicles.

Vehicle detection at nighttime using the rear-view monitoring technique is detailed below. The vehicle model under this setting is shown in Figure 8a. The detection process is to determine not only taillights but also the front end of each vehicle. With these two features, we can determine the vehicle length. The most salient feature of each vehicle at nighttime is its taillights. Since this is similar to that of front-view headlights, we can detect the positions of taillights using the same manner. Furthermore, we should recognize the vehicle front contour as a boundary of the dark region (i.e., the vehicle body) and the bright region (i.e., the lighted pavement).

As shown in Figure 8b, we define the following parameters for the vehicle model at nighttime.

(1) $\mathbf{p}_{tl}$ and $\mathbf{p}_{tr}$. They are the left- and right-ends of taillights. The distance between them represents the vehicle width. This model has vertical offset h between taillights and the road surface plane.

(2) $\mathbf{p}_{hr}$, which is the rightmost point of the vehicle contour. It represents the intersection between line l, which connects $\mathbf{p}_{tr}$ and the vanishing point $\mathbf{p}_v$, and f, which has a large luminance difference between both sides. The vehicle length can be extracted from the distance between $\mathbf{p}_{tr}$ and $\mathbf{p}_{hr}$ in the image plane. We assume that parameter $\mathbf{p}_{hr}$ has the same vertical offset h from the road surface plane as $\mathbf{p}_{tl}$ and $\mathbf{p}_{tr}$.

Vehicle detection is based on finding feature points $\mathbf{p}_{tl}$, $\mathbf{p}_{tr}$, and $\mathbf{p}_{hr}$. We can extract the width parameter W, the length parameter L, and the centroid point $\mathbf{p}_c$ in the road plane from these feature points as follows.

Detection of vehicle tail positions $\mathbf{p}_{tl}$ and $\mathbf{p}_{tr}$

Taillight locations are found using the conventional light location detection method. That is, vehicle tail locations can be estimated by pairing detected taillights. This pairing operation exploits the following characteristics. First, the vertical position for a pair of taillights is almost the same because of the camera setting. Second, the distance between each pair of taillights in the real world has almost the same value. Consequently, we can determine possible distance between taillights according to the light location in the image plane. Finally, the location of feature points $\mathbf{p}_{tl}$ and $\mathbf{p}_{tr}$ can be determined as the leftmost and the rightmost points of each detected vehicle tail.

Detection of vehicle front position $\mathbf{p}_{hr}$

The detection of $\mathbf{p}_{hr}$ is achieved by finding the intersection between boundary f and line l in the vehicle model given by Figure 8a. This approach demands the knowledge of vanishing point $\mathbf{p}_v$, which can be obtained using vehicles' trajectory information. When we investigate the luminance value on line l from $\mathbf{p}_{tr}$ to $\mathbf{p}_v$, there is a significant change from a low value to a high value around $\mathbf{p}_{hr}$. Thus, we can determine the location of $\mathbf{p}_{hr}$ from the luminance value of line l as well. This detection is based on the assumption that the difference between spatially neighboring pixels in the vehicle body region is smaller than that at boarder f as shown in Figure 8.

Vehicle position extraction

The vehicle location can be estimated from results of the above steps. We can obtain the following parameters: width W, length L, and centroid position $\mathbf{p}_c$ for each vehicle as

shown in Figure 8b. Please note that we do not have the height value h from the image yet. It is however possible to use the empirical value to estimate parameter h.

3.4. Enhanced background maintenance

In this section, we examine ways to improve the background maintenance technique. Since background subtraction is an essential step to any vehicle detection algorithm, it demands high accuracy. For the modeling of each pixel on the background image, the Gaussian model would be sufficient if pixels on the background image belong to a particular surface of an object under the same lighting condition. Here, we treat each pixel in the background image as a random variable with an adaptive Gaussian distribution. It is assumed that the background status does not change much within a short period of time. Mathematically, for the background pixel at position (x, y) and time t, its intensity is modeled by the Gaussian probability function

$$B(x, y, t) = G(B_{\text{rep}}(x, y, t), \sigma^2(x, y, t)), \quad (12)$$

where B_{rep} and σ represent the mean and the standard deviation, respectively. Generally speaking, the model should be updated according to the latest image to maintain the accuracy of background subtraction results. Sometimes, there could be a global luminance change in the latest input image, which may be caused by automatic color/white balance control of the camera or a sudden change of weather conditions such as the effect of rains and clouds. In this case, it is preferred not to segment any region to be foreground and background since the change is not caused by object motion.

The background update may introduce a blending problem if the background model is updated using pixels that do not belong to the background region in the latest image. Consequently, the following two principles should be followed as much as possible. First, the update should be performed frequently enough to avoid potential overdetection of vehicles. Second, we should avoid pixels that belong to the moving region in the latest image in the update. In terms of mathematics, the background model is updated via

$$
\begin{aligned}
B_{\text{rep}}(x, y, t+1) &= (1 - \alpha M_B(x, y, t)) B_{\text{rep}}(x, y, t) \\
&\quad + \alpha M_B(x, y, t) I(x, y, t), \\
\sigma^2(x, y, t+1) &= (1 - \alpha M_B(x, y, t)) \sigma^2(x, y, t) \\
&\quad + \alpha M_B(x, y, t)(I(x, y, t) - B_{\text{rep}}(x, y, t))^2,
\end{aligned}
$$
$$(13)$$

where $0 \leq \alpha \leq 1$ is a weighting parameter, $M_B(x, y, t)$ is a binary mask to represent the background region, $(M_B(x, y, t) = 1$ means that the pixel on the input image at position (x, y) and time t belongs to the background region), $I(x, y, t)$ represents the input pixel value at position (x, y) at time t. To meet the first principle, we can control adaptation delay by selecting a proper value of α. Faster adaptation is achieved by setting α near 1. As to the second principle, the setting of M_B may help control the blending. With the

conventional approach, mask M_B is obtained by

$$M_B(x, y, t) = \begin{cases} 1 & \text{if } |I(x, y, t) - B_{\text{rep}}(x, y, t)| < T_1 \sigma(x, y, t), \\ 0 & \text{otherwise.} \end{cases}$$
$$(14)$$

Note that M_B is set to zero at all positions when there is a fast monotonic change in the image such as automatic brightness/white balance control or sudden brightness change due to the weather. It means that the background model is not updated at all, since the difference between the latest pixel value and the corresponding pixel value of the background model is larger than $T_1 \sigma$. In this case, the whole image is regarded as one moving object region when the background subtraction technique is applied. To solve this problem, we calculate mask M_B with a new scheme. Instead of using (14), we determine M_B according to the relation between the difference in consecutive frames and the deviation α. At first, we calculate the absolute difference between consecutive frames as

$$D(x, y, t) = |I(x, y, t) - I(x, y, t-1)|. \quad (15)$$

Then, we calculate mask M_B via

$$M_B(x, y, t) = \begin{cases} 1 & \text{if } D(x, y, t) < T_2 \sigma(x, y, t), \\ 0 & \text{otherwise,} \end{cases}$$
$$(16)$$

where T_2 is another parameter.

Since the mask update formula in (16) does not depend on the pixel value of the background model B_{rep} as given in (14), the current pixel value is included in the background model update as far as the temporal absolute difference D is equal to or lower than $T_2 \sigma$. Thus, (16) includes pixels whose change in consecutive frames is not significant in the update of the background model. With this modification, we can update the background model quickly according to the latest image when there is a change caused by the automatic color/white balance control of the camera or the sudden weather change. Finally, the foreground region is determined by

$$M_F(x, y, t) = \begin{cases} 1 & \text{if } |I(x, y, t) - B_{\text{rep}}(x, y, t)| < T_3 \sigma(x, y, t), \\ 0 & \text{otherwise.} \end{cases}$$
$$(17)$$

This new background maintenance technique is called the fast adaptive background maintenance method.

3.5. Summary

Figure 9 shows the block diagram of a vision-based traffic monitoring system that consists of the proposed three techniques: shadow elimination, occlusion detection, and nighttime detection. The proposed shadow elimination algorithm has several unique features. First, parameters of vehicle and shadow models can be estimated from input video without the light source and camera calibration information.
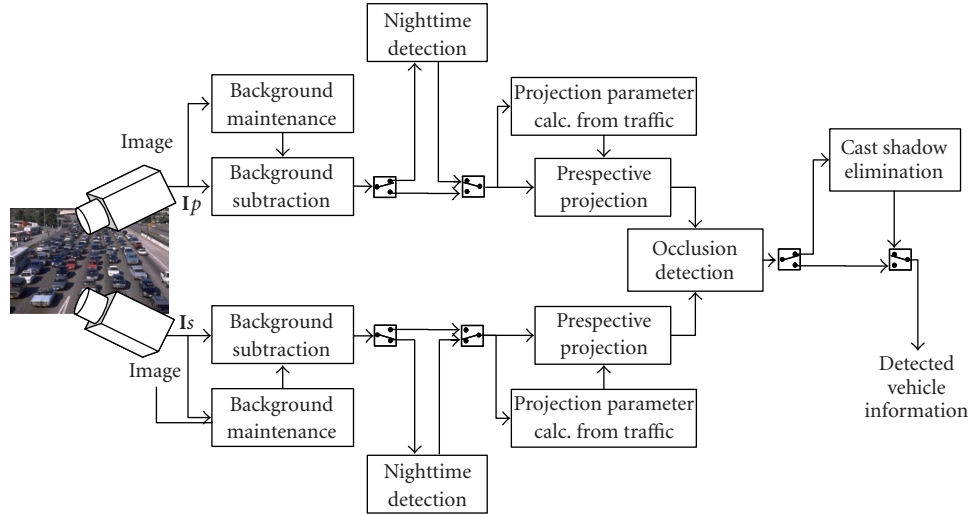
FIGURE 9: The block diagram of the proposed vision-based traffic monitoring system.

Second, this algorithm does not perform 3D image analysis operations so that the complexity is low. Third, we distinguish the cast shadow region from the vehicle itself via the proposed joint model rather than two separate models and this simplifies the algorithm greatly. For occlusion detection, images from multiple cameras for each scene are employed. Our approach has several advantages in comparison with conventional approaches. First, this approach does not need edge detection or region segmentation operations that are sensitive to environmental factors. Second, our method does not need camera calibration in advance. Third, previous work with multiple-camera solutions has difficulties in finding the corresponding feature points among captured images. No matching is required for the proposed method since we only focus on the centroid of each vehicle instead of all pixels. For nighttime detection, we introduce the rear-view monitoring concept to overcome the problem arising from the poor lighting condition. Finally, we present a background maintenance method that can adjust the background model according to the environmental change quickly.

Vehicle tracking accuracy depends on the direction of traffic flow and the camera angle. Vehicles that are far from the camera have a lower resolution as compared to vehicles that are near to the camera. For vehicles at the far end of the scene with their low spatial resolutions, it is difficult to identify them. Furthermore, the associated motion information along the road may not have a sufficient resolution. In contrast, it is easier to detect and track vehicles with a higher resolution if they are near the camera. There is another factor worth consideration, that is, when vehicles are far from the camera, there is a higher probability of occlusion, which degrades detection accuracy.

4. EXPERIMENTAL RESULTS

We evaluated the performance of a vision-based highway surveillance system by implementing the proposed shadow elimination, occlusion detection, vehicle nighttime detection, and enhanced background maintenance techniques using various highway traffic scenes captured with a fixed camera position. The size of the image is 320×240 pixels and the frame rate is 30 fps. Each color component is quantized to 8 bits.

4.1. Cast shadow elimination

Figures 10a and 10b show results of the vehicle-model-type decision from the scenes. In these two images, the side $\mathbf{S}_f$ and other sides of the vehicle model are drawn in thick and thin lines, respectively. In these two vehicles, the extracted shadow parameters are normalized with the length of $\mathbf{S}_f$. The parameters r_+ and r_- can be obtained according to the shadow model type. Although contours of the extracted moving object regions do not include the whole area of each vehicle, the fitting of the vehicle model and the determination of the shadow model work very well.

Figure 11a shows the fitted joint models $\mathbf{M}'_J$ with cast shadow regions and the calculated vehicle models $\mathbf{M}'$ without cast shadow regions. The tracking of each vehicle in the latest 0.5 second is also drawn. From the result, vehicles in both outbound and inbound lanes are tracked successfully. Figure 11b shows the result of the proposed method applied to images from another scene. As we can see from the figure, the moving cast shadow is longer than that in Figure 11a. The left image shows the result of the detected joint vehicle-shadow model (in the dark box) and the six-vertex vehicle model (in the light box). The right image shows the result of moving region extraction using background subtraction. The detected vehicle model information as well as the track information of each detected vehicle are also displayed in both images. The scene condition of Figure 11b is different from that of Figure 11a in terms of the light source condition and the camera location. We see that the proposed method can detect vehicles well under both conditions.
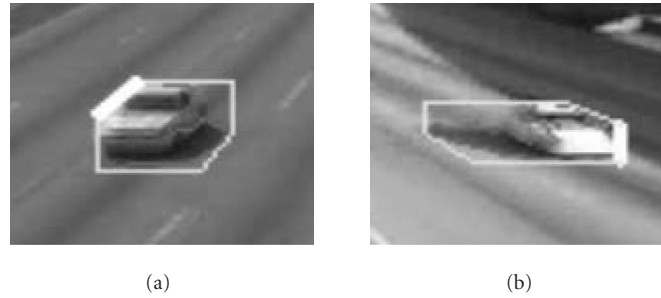
(a)   (b)

FIGURE 10: (a) Vehicle-model-type determination for shadow elimination with two different cases.

(a)   (b)

FIGURE 11: (a) Detection and tracking results for two scenarios.

4.2. Occlusion detection

Figure 12a was captured by a camera above the center lane of the highway and Figure 12b was captured above the leftmost lane of the highway. Figures 12c and 12d are masks obtained using the background subtraction operation. It is clear that we cannot distinguish either the cast shadow region or the occurrence of occlusion only from the mask information as shown in Figure 12c.

Figures 13a and 13b are results of the projected image on the road surface plane. Figure 14 shows the detected occlusion. The small hollow circle within the large white thick circle represents the obtained center from the moving object region in image $\mathbf{I}_p$. Since the region in each white thick circle is regarded as a single moving object region in image $\mathbf{I}_p$, the system cannot recognize that there are actually two vehicles in this image. The solid white circles in the images represent the obtained center of moving regions in image $\mathbf{I}_s$. By the detection algorithm in Table 1, we can determine these regions to be Case 3. Consequently, the solid white green circles are obtained as the centers of the vehicles in the occlusion region.

The performance of the proposed occlusion detection algorithm is compared with that of single-camera solutions in Table 2. We compare the result by checking the identified vehicles passing the middle of the projected image. With a single image source, the rate of the identified vehicle in the occlusion region is 53.9% and 25.3%, respectively. In particular, the detection ratio in image $\mathbf{I}_s$ has a low score. This is due to the poor camera angle in image capturing. The larger

the horizontal angle difference between the camera axis and the traffic flow is, the higher the probability of the occluded scene becomes. On the contrary, the proposed method can identify vehicles in the occlusion region up to 76.3%. We see that the vehicle detection rate is improved significantly by the proposed occlusion detection system.

4.3. Vehicle nighttime detection

Figures 15a and 15b show results of nighttime vehicle location detection using the conventional front-view technique and the proposed rear-view monitoring technique, respectively. From these figures, the proposed method successfully detects the vehicle length as well as the vehicle location at nighttime. Detection with the front-view monitoring has two primary disadvantages. First, there are more cases of headlight pair misdetection. Second, the vehicle length cannot be extracted for some vehicles.

Figures 16a and 16b show results using the front-view monitoring and the rear-view monitoring, respectively. Here, we picked up three typical vehicles as examples. The x-axis represents the distance along line l (see Figure 8a) with respect to point $\mathbf{p}_{tr}$, in the direction opposite to the vanishing point, and the y-axis represents the pixel values on line l in the model given by Figure 8a. In Figure 16a, there are no significant features around the vehicle tail region in the front-view image. In contrast, in Figure 16b, the vehicle region on line l is represented with an arrow line and there are significant changes in the pixel value around the vehicle head region for vehicles in the rear-view image. Consequently, it is
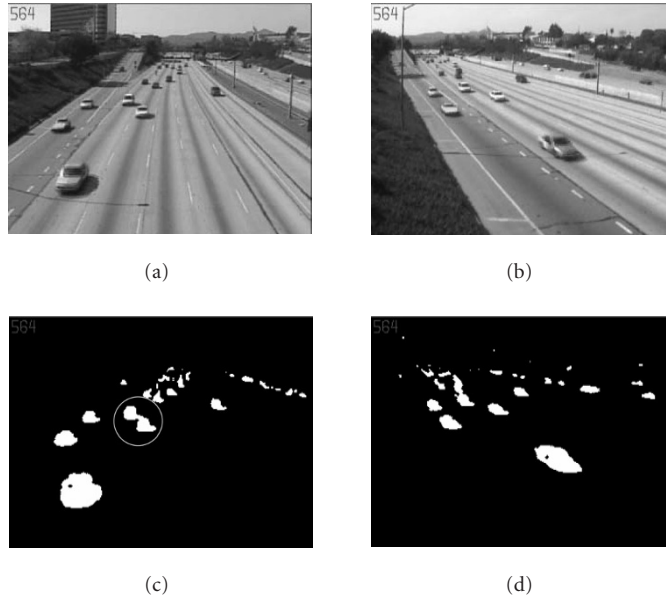
(a)

(b)

(c)

(d)

FIGURE 12: Simultaneously captured image for a scene: (a) captured from above the center lane $\mathbf{I}_p$ and (b) captured from the left $\mathbf{I}_s$; and detected moving regions from the output of background subtraction: (c) extracted moving region from (a), and (d) the extracted moving region from (b).
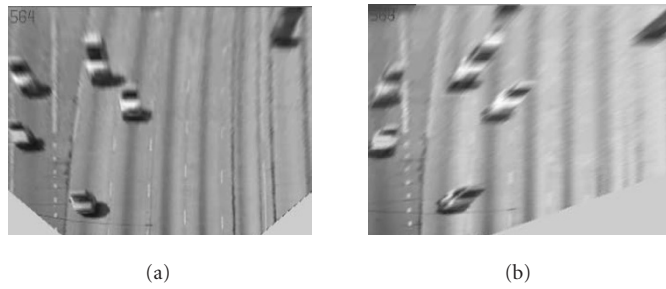

(a)

(b)

FIGURE 13: The projected images to the road surface plane: (a) projected image $\mathbf{I}_p$ and (b) projected image $\mathbf{I}_s$.
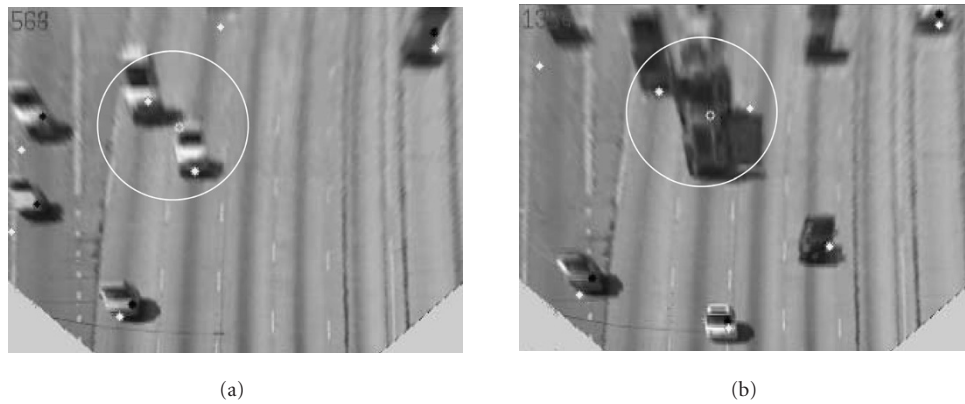

(a)

(b)

FIGURE 14: The results of occlusion detection using the proposed algorithm.

TABLE 2: Results of occlusion detection.

Item	$\mathbf{I}_p$	$\mathbf{I}_s$	Proposed	Ground truth
Identified vehicles	359 (91.1%)	318 (80.7%)	376 (95.4%)	394
Identified vehicles in occlusion	41 (53.9%)	23 (25.3%)	58 (76.3%)	76
Misdetection	35	78	23	—
Overdetection	0	2	5	—

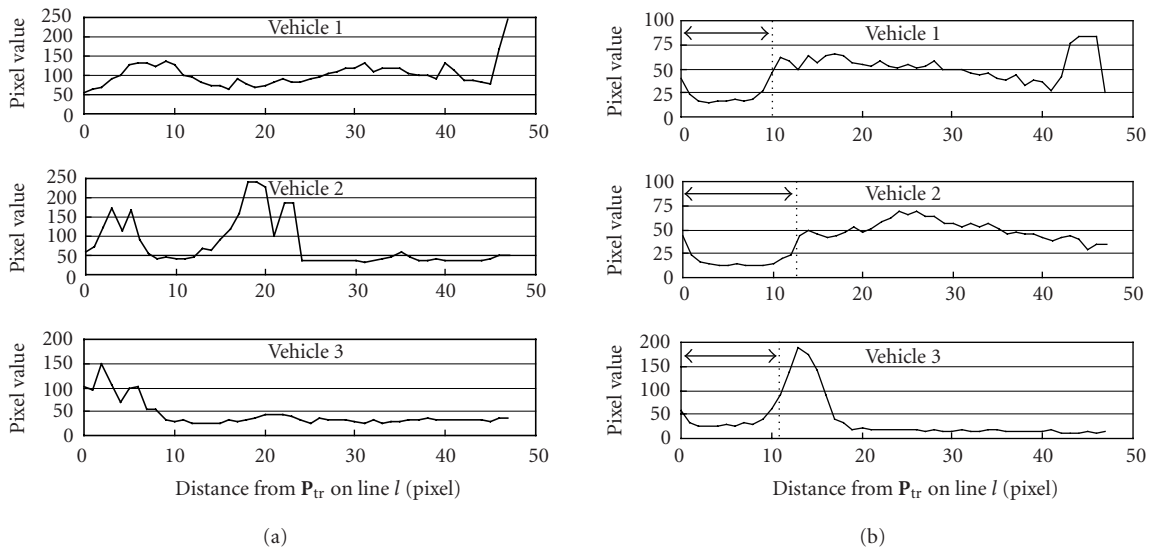(a)   (b)

FIGURE 15: Nighttime vehicle detection result using (a) the front-view monitoring technique, and (b) the proposed rear-view monitoring technique.

(a)   (b)

FIGURE 16: Luminance transition from $\mathbf{p}_{tr}$ toward $\mathbf{p}_v$ along line l in the model (see Figure 7a) of (a) the front view and (b) the rear view of the monitoring system.

easy to determine the third feature point $\mathbf{p}_{fr}$ from the rear-view images as compared to that of the front-view images.

We compare detection accuracy using the front and the rear views at daytime and nighttime in Table 3. The accuracy is evaluated by checking the centroid position of vehicles when they pass the vertically middle of the image plane. For each condition, the number of passing vehicles was set to be

320. It is clear that detection using the rear-view technique achieves higher accuracy than the front-view technique in both daytime and nighttime. There is a remarkable miss rate with the front-view monitoring at nighttime. The degradation is due to the difficulty in finding the third feature point $\mathbf{p}_{hr}$ in the front-view video. The miss at daytime is mainly caused by occlusion. The miss at nighttime is relatively high

TABLE 3: Comparison of detection performance with different conditions.

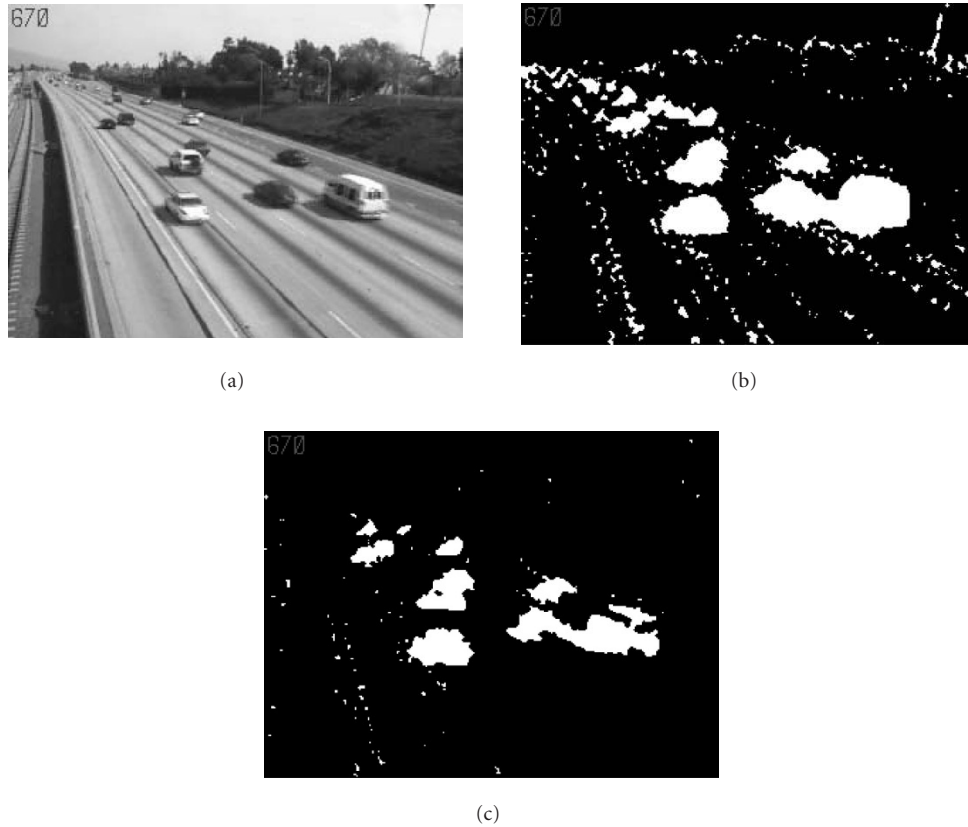| Conditions	Miss	False alarm	Ground truth (no. of vehicles)
Daytime, rear view	11 (3.4%)	3 (0.9%)	320
Nighttime, rear view	18 (5.6%)	7 (2.2%)	320
Daytime, front view	18 (5.6%)	7 (2.2%)	320
Nighttime, front view	163 (50.9%)	11 (3.4%)	320

(a)

(b)

(c)

FIGURE 17: Results of background subtraction with two different background maintenance schemes: (a) the original image, (b) the detected moving object region by the conventional method, and (c) the detected moving object region by the proposed method.

as compared to that at daytime due to the poor lighting condition. Detection with the nighttime image is based on detecting locations of light sources and pairing them according to their spatial relation. Thus, misdetection at nighttime occurs for objects such as motorcycles or vehicles with one of their taillights off. Moreover, the false alarm is caused by disturbances such as reflected headlights. Among these four monitoring conditions, the daytime monitoring with the rear-view monitoring performs the best.

4.4. Results of enhanced background maintenance algorithm

Finally, we compare the performance of the conventional and the proposed fast adaptive background maintenance techniques. The conventional approach adopts (14) to update mask M_b in the background model while the proposed method uses (16) for the update and (17) for determining

the moving objects. Figures 17a, 17b, and 17c show the background subtraction results from a daytime image sequence. Figure 17a shows one input image, where there is change in the pixel value caused by automatic white balance control of the camera. Figures17b and 17c show results of background subtraction using the conventional background maintenance and the proposed fast adaptation background maintenance, respectively. In Figure 17b, detection accuracy degrades because of overdetection with the conventional background maintenance. Detection accuracy is improved by the proposed fast adaptation background maintenance.

5. CONCLUSION

In this paper, we proposed some solutions to solve three crucial problems, that is, moving cast shadow elimination, occlusion detection, and nighttime detection, in a vision-based

monitoring system for highway traffic surveillance. A 2D joint vehicle-shadow model was proposed to obtain the vehicle and the shadow information from luminance analysis. A multiple-camera system was employed to detect occlusion based on image matching using perspective projection. A rear-view monitoring technique was used to improve the vehicle detection at nighttime. Moreover, a new fast adaptive background maintenance was developed to adapt the environmental change quickly. As demonstrated by experimental results, the proposed system can detect various kinds of vehicle information robustly. In the near future, we would like to extract various kinds of traffic events based on vehicle detection results for highway surveillance system as an extension of this work.

REFERENCES

[1] B. L. Tseng, C. Y. Lin, and J. R. Smith, "Real-time video surveillance for traffic monitoring using virtual line analysis," in *Proc. IEEE International Conference on Multimedia and Expo (ICME '02)*, vol. 2, pp. 541–544, Lausanne, Switzerland, August 2002.

[2] R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, no. 2, pp. 119–130, 2000.

[3] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Proc. 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 225–261, Corfu, Greece, September 1999.

[4] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 7, pp. 780–785, 1997.

[5] I. Haritaoglu, D. Harwood, and L. S. Davis, "A fast background scene modeling and maintenance for outdoor surveillance," in *Proc. IEEE International Conference on Pattern Recognition (ICPR '00)*, vol. 4, pp. 179–183, Barcelona, Spain, September 2000.

[6] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 2, pp. 246–252, Fort Collins, Colo, USA, June 1999.

[7] D. Gutchess, M. Trajkovic, E. Cohen-Solal, D. Lyons, and A. K. Jain, "A background model initialization algorithm for video surveillance," in *Proc. IEEE International Conference on Computer Vision (ICCV '01)*, vol. 1, pp. 733–740, Vancouver, British Columbia, Canada, July 2001.

[8] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with HSV color information," in *Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC '01)*, pp. 334–339, Oakland, Calif, USA, August 2001.

[9] G. S. K. Fung, N. H. C. Yung, G. K. H. Pang, and A. H. S. Lai, "Effective moving cast shadow detection for monocular color image sequences," in *Proc. 11th International Conference on Image Analysis and Processing (ICIP '01)*, pp. 404–409, Palermo, Italy, September 2001.

[10] A. H. S. Lai, G. S. K. Fung, and N. H. C. Yung, "Vehicle type classification from visual-based dimension estimation," in *Proc. IEEE Conference on Intelligent Transportation System (ITSC '01)*, pp. 201–206, Oakland, Calif, USA, August 2001.

[11] A. Prati, I. Mikic, C. Grana, and M. M. Trivedi, "Shadow detection algorithm for traffic flow analysis: a comparative study," in *Proc. IEEE Conference on Intelligent Transportation System (ITSC '01)*, pp. 340–345, Oakland, Calif, USA, August 2001.

[12] Y. K. Jung and Y. S. Ho, "Traffic parameter extraction using video-based vehicle tracking," in *Proc. IEEE Conference on Intelligent Transportation Systems (ITSC '99)*, pp. 764–769, Tokyo, Japan, October 1999.

[13] S. Gupte, O. Masoud, R. F. K. Martin, and N. Papanikolopoulos, "Detection and classification of vehicles," *IEEE Trans. Intell. Transport. Syst.*, vol. 3, no. 1, pp. 37–47, 2002.

[14] J. Kato, T. Watanabe, S. Joga, J. Rittscher, and A. Blake, "An HMM-based segmentation method for traffic monitoring movies," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 9, pp. 1291–1296, 2002.

[15] D. Koller, J. Weber, T. Huang, et al., "Towards robust automatic traffic scene analysis in real-time," in *Proc. IEEE International Conference on Computer Vision and Image Processing (CVPR '94)*, pp. 126–131, Jerusalem, Israel, October 1994.

[16] I. Lee, H. Ko, and D. K. Han, "Multiple vehicle tracking based on regional estimation in nighttime CCD images," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '02)*, vol. 4, pp. 3712–3715, Orlando, Fla, USA, May 2002.

[17] D. Gao, J. Zhou, and L. Xin, "SVM-based detection of moving vehicles for automatic traffic monitoring," in *Proc. IEEE International Conference on Intelligent Transportation Systems (ITSC '01)*, pp. 745–749, Oakland, Calif, USA, August 2001.

Akio Yoneyama received the B.E. and the M.E. degrees in electrical engineering from Keio University, Japan, in 1992 and 1994, respectively. He joined KDD in 1994. Since 1996, he was with KDDI R&D Laboratories Inc. From 2002 to 2003, he was a Visiting Researcher at University of Southern California. His current research interests include video compression and image processing.

Chia-Hung Yeh was born in Taiwan, on January 10, 1975. He received the B.S. and Ph.D. degrees from National Chung Cheng University (CCU), Chiayi, Taiwan, in 1997 and 2002, respectively, all from the Department of Electrical Engineering. He received Postdoctoral Research Fellowship Award from Professor C.-C. Jay Kuo's research group at the Department of Electrical Engineering-Systems, (USC) Viterbi School of Engineering, University of Southern California (USC), from August 2002 to December 2004. He was a Technical Consultant of ALi Cooperation from August 2003 to August 2004. In the summer of 2004, he joined MAVs Laboratory Inc. in the USA and has been the Vice President and Chief Technology Officer in MAVs Laboratory Inc., Taiwan, since 2005. His research interests are bioinformatics, image/audio/video signal processing, image/video database management, and optical information processing/computing. He has served as a reviewer for international journals and conferences and was an invited speaker at conferences. In 2002, he received the Outstanding Student Award from National Chung Cheng University. He is a Member of the Institute of Electrical and Electronics Engineers (IEEE) and International Society for Optical Engineering (SPIE).

C.-C. Jay Kuo received the B.S. degree from the National Taiwan University in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology in 1985 and 1987, respectively. He is with the Department of Electrical Engineering, the Signal and Image Processing Institute (SIPI), and the Integrated Media Systems Center (IMSC) at the University of Southern California (USC) as a Professor of electrical engineering and mathematics. His research interests are in the areas of digital image/video analysis and modeling, multimedia data compression, communication and networking and multimedia database management. Dr. Kuo has guided about 60 students to their Ph.D. degrees and supervised 15 postdoctoral research fellows. Currently, his research group at USC consists of around 40 Ph.D. students and 5 postdoctors. Dr. Kuo is a Fellow of IEEE and SPIE. He is the Editor-in-Chief for the *Journal of Visual Communication and Image Representation.* Kuo received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively. He received the Northrop Junior Faculty Research Award from the USC Viterbi School of Engineering in 1994. He received the Best Paper Award from the multimedia communication Technical Committee of the IEEE Communication Society in 2005.