Highly accurate texture-based vehicle segmentation method

William Wai Leung Lam
Clement Chun Cheong Pang
Nelson H. C. Yung
University of Hong Kong
Department of Electrical and Electronic Engineering
Room 524, Chow Yei Ching Building
Pokfulam, Hong Kong
E-mail: wllam@eee.hku.hk

Abstract. In modern traffic surveillance, computer vision methods have often been employed to detect vehicles of interest because of the rich information content contained in an image. Segmentation of moving vehicles using image processing and analysis algorithms has been an important research topic in the past decade. However, segmentation results are strongly affected by two issues: moving cast shadows and reflective regions, both of which reduce accuracy and require postprocessing to alleviate the degradation. We propose an efficient and highly accurate texture-based method for extracting the boundary of vehicles from the stationary background that is free from the effect of moving cast shadows and reflective regions. The segmentation method utilizes the differences in textural property between the road, vehicle cast shadow, reflection on the vehicle, and the vehicle itself, rather than just the intensity differences between them. By further combining the luminance and chrominance properties into an OR map, a number of foreground vehicle masks are constructed through a series of morphological operations, where each mask describes the outline of a moving vehicle. The proposed method has been tested on real-world traffic image sequences and achieved an average error rate of 3.44% for 50 tested vehicle images.
[DOI: 10.1117/1.1645849]

Subject terms: vehicle segmentation; shadow detection; texture analysis; visual traffic surveillance; intelligent transportation systems.

Paper 030267 received Jun. 6, 2003; revised manuscript received Oct. 9, 2003; accepted for publication Oct. 10, 2003.

1 Introduction

With the rapid advancement of computer and communication technologies, the research and development of intelligent transportation systems (ITS) methods are becoming more and more ambitious and substantial.[1] Among the many facets of ITS, visual traffic surveillance plays an important role in traffic survey, data capturing, incident detection, and safety management in general. Comparing it with traditional detection technologies, it is able to convey a comprehensive content of information that is easy for human interpretation. Such information can, of course, be interpreted by machines if appropriate algorithms are available. The ability to interpret visual information with computers not only improves operation efficiency, but also the overall ''intelligence'' of the ITS. For this reason, much research effort has been directed to finding methods that can automatically analyze and interpret the content of an image or image sequence. To do that, features of vehicles in an image must first be extracted or segmented, from which mean speed, flow rate, and incidents, among other traffic parameters, are determined. As such, vehicle segmentation has always been one of the major components of visual traffic surveillance.[2–11]

Segmentation algorithms that extract the vehicle of interest from an image background in a single-image frame or an image sequence have recently been actively pursued.[2,7,9,10] In many of these algorithms, the detection of vehicles is mainly based on their motion property, luminance, chrominance, edges, and/or corner features. Unfortunately, these approaches suffer from the major drawback that when they are applied to outdoor scenes, undesired natural phenomena such as cast shadows on the road, or reflection on the usually shiny vehicle exterior, interfere with the desired features. As a result, segmentation becomes inaccurate. This also creates a series of problems associated with pseudo-occlusion if the cast shadows are not detected and eliminated. To be more accurate in extracting vehicle(s) from a given scene, numerous shadow detection methods have been proposed,[9,10,12–15] most of which are based on features such as intensity and color. However, they all experienced a number of limitations, such as specific weather conditions, that make them ineffective in practical outdoor environments.

On the other hand, texture analysis can be potentially effective in solving the problem. A proof of the role of textural information in outdoor object recognition was performed by comparing classification correctness.[16] Haralick[16] showed that if textural information was used to classify an outdoor object, 99% accuracy was achieved. Conversely, spectral information-based classification achieved only 74% accuracy. Therefore, in this research we propose a very accurate and effective algorithm that is based on texture analysis for segmenting vehicles. We assume that the textural features of moving vehicles are significantly different from the background. However, the

studies of texture analysis of road conditions and vehicles are limited.[17] For that reason, a literature review of vehicle segmentation is given in the next section, and the analysis of the problem of using a texture analysis technique is briefly described in Sec. 3. Following that, the proposed methodology is presented in Sec. 4. Simulation results and discussion are given in Sec. 5. Finally, conclusions are drawn in Sec. 6.

2 Related Works

Due to the importance of vehicle segmentation, numerous extraction methods have been proposed in the past.[2–11] Their main purpose is to remotely acquire traffic image sequences from roadside surveillance cameras and interpret them into traffic parameters and vehicle behavior by analyzing the image sequences. Among them, background subtraction is a common approach in extracting moving vehicles from a stationary reference background that is estimated over a number of image frames. However, most of these approaches suffer from a major drawback. In outdoor daylight scenes, shadows cast by moving vehicles are often detected as part of the vehicles, since shadows move in accordance with the movement of the vehicle. Moreover, a vehicle exterior is normally reflective, which may reflect images of nearby objects, and change the homogeneity of vehicle regions. When the detected vehicles contain shadows and/or reflective regions, large errors may result in the computation of their shape, speed, and other parameters. This also creates a multitude of problems associated with occlusion.

Although numerous shadow detection methods have been proposed,[9,10,12–15] some of them are limited to indoor environment only, while for those that can be used outdoor, environment information is usually required. Basically, these methods can be classified into two major categories: single-frame approach and the reference-frame approach.

2.1 *Single-Frame Approach*

The single-frame approach only utilizes the information within a single-image frame. Traditionally, vehicle segmentation algorithms have been mostly developed from the single frame concept.[13,14] As there are limited visual characteristics that can be extracted from a single-input frame, authors tend to make stricter assumptions as such.

Funka-Lea and Bajcsy[13] presented an active shadow recognition method by combining color and geometric properties of the image. They suggested a number of cues that together point toward the identification of a shadow. One of the cues is that the intensity, hue, and saturation changes due to shadows tend to be predicable. Therefore, an image can be segmented by color image segmentation methods that recover single material surfaces as single image regions, regardless of whether the surface is partially in shadow. The penumbra and umbra of shadows are then recovered based on the illumination. To recover the geometric property of the scene, such as the location of the light sources, an extendable probe is used to actively obtain shadows in the scene. Both outdoor and indoor scenes have been tested, and shadows have been reasonably detected. However, due to the use of the linear color cluster assumption, their method is limited to a relatively simple scene. Additionally, the umbra and penumbra properties of shadow are hardly maintained in a complex outdoor scene.

Salvador, Cavallaro, and Ebrahimi[14] presented a method that is based on the use of invariant color models to identify and classify shadows in color images. The candidate shadow regions are first extracted by searching the edge map in the dark regions of the image. After color conversion to the invariant color model, the candidate shadow pixels are classified as self shadow points or as cast shadow points based on the detected color edge of the image. The method has been applied to a number of indoor scenes with one or two simple objects and one light source. From their results, shadows have been correctly identified and classified. Similarly to other single frame approach methods, the application of their method is restricted by assumptions that shadows are cast on a flat and nontextured surface, objects are uniformly colored, and a single light source illuminates the scene. Of course, none of these are valid in outdoor scenes.

2.2 *Reference-Frame Approach*

Essentially, the reference-frame approach[7,9,15] is a simple and efficient approach. This approach utilizes a number of previous frames to estimate a reference frame for comparison with the current image frame, which is suitable for detecting cast shadows that are associated with moving objects captured by a static camera.

Fung et al.[7] proposed a methodology to detect moving vehicles by eliminating the vehicles' cast shadows and stationary background in an image sequence. In principle, the input image is subtracted from the background image in the luminance, chrominance, and gradient density domains. By mapping through various shadow score functions, these shadow scores in different domains are combined and transformed into the overall shadow confidence score, which provides an indication of the likelihood of the pixels belonging to the cast shadow region. In this segmentation algorithm, the edge pixels that belong to the vehicle are obtained through a threshold filtering by their shadow confidence scores. The convex hull of these vehicle edges is then calculated and is used to define the vehicle mask, whereas the remaining pixels become the shadow region. Although it works well with many different outdoor scenes, this method does not preserve the concavity of the vehicle, and higher segmentation accuracy is only achievable for vehicles with lighter color and larger dimension: otherwise the error rate rapidly increases.

Mikic et al.[9] presented an algorithm that statistically classifies pixels into shadow, object, and background classes. In their approach, the color response of the camera is statistically predetermined as a background verse, shadowed-color-component diagonal matrix. Based on the given *a priori* probabilities of the pixel belonging to different classes, each pixel is classified by maximizing the *a posteriori* probability of the class membership. A spatial smoothing filter is imposed for postprocessing the noisy shadow detection results computed by the previous stage. Their algorithm has been successfully tested on a traffic scene with long shadows. However, computation of the diagonal matrix is highly dependent on the camera settings, and may lead to performance degradation if there are changes in scene conditions.

Stauder, Mech, and Osterman,[15] proposed a detection method for ideal indoor cast shadow. Their algorithm works well under the assumptions that there is a flat background and the light source is of nonnegligible size and intensity. According to their simulation results on three test sequences, their algorithm is able to detect single or multiple moving cast shadows in indoor video sequences with spotlights and cast shadows on the background. If the shadows are weak, their algorithm may fail due to the assumption on penumbra and umbra properties of shadow, which does not hold anymore.

2.3 *Summary*

The reference-frame approach utilizes multiple previous frames to determine estimation over the temporal domain, from which the properties of the regions under shadow and not under shadow can be extracted. Therefore, in most practical cases, the reference-frame approach achieves higher accuracy and robustness compared to the single-frame approach. However, most of the current reference-frame approaches only consider a specific aspect of the shadow, and do not fully utilize the spectrum of features that may be useful to the eventual classification of vehicle and shadow. Therefore, it is our intention to include those unique shadow features that we have reviewed, and integrate them to give a fresh proposal to segment vehicles and eliminate cast shadows in practical outdoor scenes.

3 Problem Analysis

Conventionally, vehicles are extracted based on their appearance and/or motion from an image or image sequence. A common first step in most recently proposed algorithms relies on subtracting a background reference image from its input image.[11,18] Subsequently, unwanted regions such as the cast shadow have to be removed by postprocessing techniques. To accurately extract vehicles from an image, we aim to eliminate the cast shadow first.

Broadly, cast shadow can be defined as the darkened region on the background of an image that is due to the foreground objects blocking the light source. The luminance values of the cast shadow pixels are normally lower than similar pixels in the background image. On the other hand, chrominance values of the cast shadow pixels are similar to similar pixels in the background image. In addition, we observed that the textural feature of a cast shadow is also very similar to similar pixels in the background image. In other words, if the textural features of the background and foreground are substantially different, such difference is not affected or altered by the cast shadow itself. To understand this better, let us consider this textual property in depth.

Generally, texture spatial organization is often described by the correlation coefficients that evaluate linear spatial relationships between primitives. In an autocorrelation model, a single pixel is considered a texture primitive, where primitive tone property is the gray level. If the texture primitives are relatively large, the autocorrelation function value decreases slowly with increasing distance, while it decreases rapidly if the texture consists of small primitives. If primitives are placed periodically in an image, the autocorrelation function is also periodic. Texture description of an image block is commonly calculated using the following autocorrelation function R,[19]

$$R(u,v)=\frac{(2M+1)(2M+1)}{(2M+1-u)(2N+1-v)} \times\frac{\sum_{m=0}^{2M-u}\sum_{n=0}^{2N-v}p(m,n)p(m+u,n+v)}{\sum_{m=0}^{2M}\sum_{n=0}^{2N}p^2(m,n)} \quad \begin{matrix}0\leqslant u\leqslant 2M\\ 0\leqslant v\leqslant 2N,\end{matrix} \tag{1}$$

where u, v are the position displacements in the m, n direction, $2M+1$, $2N+1$ are the dimensions of the image block p, and $p(m,n)$ represents the intensity value at (m,n). Based on the assumption that the image block is periodic in the spatial domain, the autocorrelation function R can be determined in the frequency domain from the image power spectrum,

$$R=\mathcal{F}^{-1}\{|F|^2\}, \tag{2}$$

where $\mathcal{F}$ is the Fourier transform. Alternatively, the autocorrelation function R can be rewritten in the spatial domain using the following equation,

$$R(u,v) =\frac{\sum_{m=0}^{2M}\sum_{n=0}^{2N}p(m,n)p[(2M+1)\mathrm{mod}(m+u),(2N+1)\mathrm{mod}(n+v)]}{\sum_{m=0}^{2M}\sum_{n=0}^{2N}p^2(m,n)} \quad \begin{matrix}0\leqslant u\leqslant 2M\\ 0\leqslant v\leqslant 2N.\end{matrix} \tag{3}$$

Essentially, the textural relationship between image frames can be evaluated by the autocorrelation function R, and the autocorrelation difference d_R between two image blocks can be calculated by the square difference of two autocorrelation functions R to compare their similarities,

$$d_R(u,v)=[R_i(u,v)-R_j(u,v)]^2, \tag{4}$$

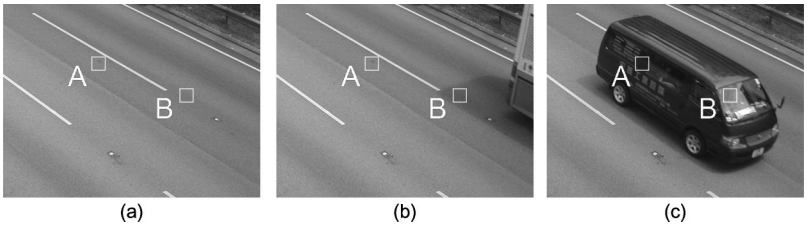


Fig. 1 (a) Image frame *f*1. (b) Image frame *f*2. (c) Image frame *f*3.

Table 1 Texture differences d_T between frames f_1, f_2, and f_3 for image blocks A and B.

Texture difference d_T	Image block A	Image block B
Frames f_1 and f_2	2.89×10^{-6}	1.72×10^{-3}
Frames f_1 and f_3	9.27	1.92×10^{-2}
Frames f_2 and f_3	9.27	1.07×10^{-2}

where R_i, R_j are the autocorrelation functions R of two different image blocks. As a final point, the texture difference d_T between two image blocks can be simply calculated by taking the mean square difference of the two autocorrelation functions, as shown in

$$d_T=\frac{1}{(2M+1)(2N+1)}\sum_{u=0}^{2M}\sum_{v=0}^{2N}[R_i(u,v)-R_j(u,v)]^2, \quad (5)$$

where R_i, R_j are the autocorrelation functions R of two different image blocks.

Figure 1 depicts three image frames f_1, f_2, and f_3, taken by a static camera where the image blocks A and B of size $M=N=16$ denote the same locations across the three image frames. The texture differences d_T of image block A and image block B between these frames are summarized in Table 1. The texture difference d_T between frame f_1 and frame f_2 for image block A is extremely small, as there is no moving object. The texture difference d_T between frame f_1 and frame f_2 for image block B is also relatively low, even though B is inside the cast shadow of a bus in frame f_2. On the other hand, in frame f_3, a dark colored vehicle is at the center of the image, and both image blocks A and B now cover part of the vehicle. The texture differences d_T for image block A and image block B of frame f_3 are drastically different from those of frame f_1 and f_2, even though the luminance values are lower in image block A of frame f_3 than image block B of frame f_2. Therefore, our observation is valid that textural features in input images are only slightly different from those of the corresponding pixels in the background image, disregarding the luminance difference.

4 Methodology

4.1 *Assumptions*

In our segmentation methodology, four assumptions are made with respect to the extraction of vehicles. First, the camera is assumed to be stationary. The background is assumed to be stationary, too, and contains texture primitives, such as the road surface. In reality, the background may appear to be slightly moving when the camera vibrates in its position. Such slight movement is tolerable in our proposed method. Second, the light source is assumed to be single and strong, thus illumination difference between the shadow and background is reasonable in intensity. Third, the texture of the road is assumed to be homogenous within the field of view. Illumination changes due to a moving cast shadow are smooth. Fourth, same elements have similar texture across frames, and therefore textural features of the different vehicle components are significantly different.

4.2 *Method Overview*

Based on these four assumptions, the moving vehicles of an input frame f_i can be reasonably extracted from the background, where the background reference frame f_b is estimated by the scoreboard algorithm.[18] Lai and Yung[18] adopt a scoreboard to astutely select from a running mode or a running average algorithm a background reference frame. However, simple subtraction between image sequences and background reference frames contains a large amount of error due to the shadow. In our proposed extraction algorithm as depicted in Fig. 2, three likelihood maps, T map, L map, and C map, are initially computed according to the differences in texture, luminance, and chrominance between the input frame f_i and background reference frame f_b, respectively. An OR-map is then constructed by performing a logical OR operation of the likelihood maps. Finally, a foreground mask is refined by performing morphological operations. This method has an inherent advantage that cast shadow regions are automatically removed, as they have the same textural property as the background. The extracted vehicle can be simply generated based on the foreground mask, and the vehicle exterior can be created by subtracting the morphological erosion from the foreground mask.

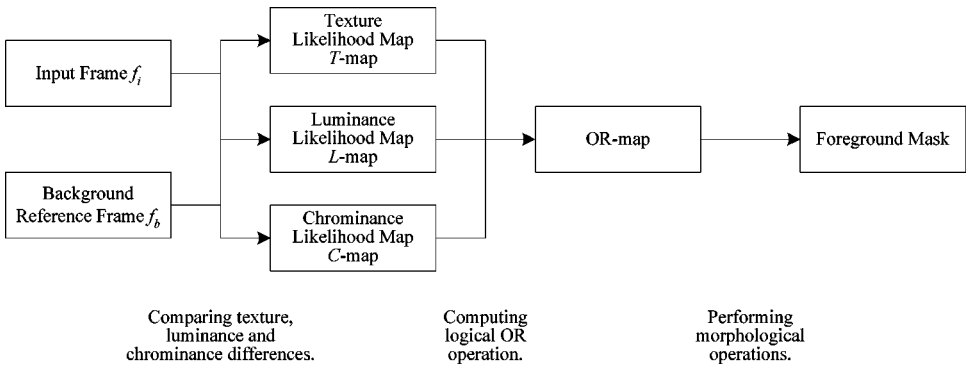


Fig. 2 Overview of proposed vehicle extraction method.

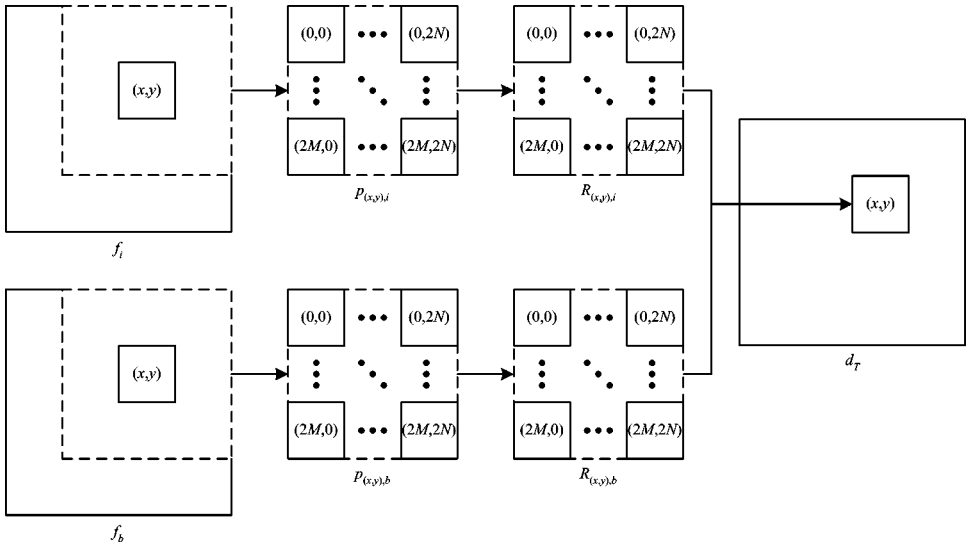


Fig. 3 Computing the texture difference d_T.

4.3 *Details of the Method*

The first task of vehicle segmentation is to construct a texture likelihood map, T map. In this step, each pixel with its neighborhood from an input frame f_i are transformed into an input image block p_i, as given by,

$$p_{(x,y)}(m,n)=f(x+m-M,y+n-N) \quad \begin{matrix} 0\leqslant m\leqslant 2M \\ 0\leqslant n\leqslant 2N \end{matrix}, \tag{6}$$

and its same neighborhood from a background reference frame f_b are transformed similarly into a background image block p_b. A T map is constructed according to the texture difference d_T,

$$d_T(x,y)=\frac{1}{(2M+1)(2N+1)}\sum_{u=0}^{2M}\sum_{v=0}^{2N}[R_{(x,y),i}(u,v) - R_{(x,y),b}(u,v)]^2 \quad \begin{matrix} M\leqslant x\leqslant X-M-1 \\ N\leqslant y\leqslant Y-N-1 \end{matrix}, \tag{7}$$

which is the mean square difference of two autocorrelation functions R of each input image block p_i with the same location of background image block p_b, as shown in Fig. 3. The T map is finally computed by comparing the texture threshold τ_T with the texture difference map d_T,

$$T\text{-map}(x,y)=\begin{cases} 1 & d_T(x,y)>\tau_T \quad M\leqslant x\leqslant X-M-1 \\ 0 & \text{otherwise} \quad N\leqslant y\leqslant Y-N-1 \end{cases}. \tag{8}$$

The second task is then to construct the L and C maps. In this step, the color model YCbCr is used to separate the luminance and chrominance components of the images. There are numerous basic color models.[14,20] Kumar, Sengupta, and Lee[20] presented a fundamental unbiased study of different color representation such as RGB, HSV, and YCbCr for detecting foreground objects and their shadows in image sequences. For RGB, all three values are very sensitive to varying illumination. For HSV, H becomes very unstable when S is small. For YCbCr, it is widely used, and the luminance channel Y is clearly separated from the chrominance channels Cb and Cr.

A luminance difference d_Y between the input frame f_i and the background reference frame f_b is computed according to the following equation,

$$d_Y(x,y)=\begin{cases} Y_i(x,y)-Y_b(x,y) & Y_i(x,y)-Y_b(x,y)>0 \quad M\leqslant x\leqslant X-M-1 \\ 0 & \text{otherwise} \quad N\leqslant y\leqslant Y-N-1 \end{cases}, \tag{9}$$

and a chrominance difference d_C between input frame f_i and background reference frame f_b is computed according to the following equation,

$$d_C(x,y)=[Cb_i(x,y)-Cb_b(x,y)]^2+[Cr_i(x,y) - Cr_b(x,y)]^2 \quad \begin{matrix} M\leqslant x\leqslant X-M-1 \\ N\leqslant y\leqslant Y-N-1 \end{matrix}. \tag{10}$$

In our method, we use summation of square differences in the Cb and Cr channels to determine the value. There could be many other methods to calculate the chrominance difference between two pixels, but summation of square differences is direct and simple for our application.

Both the L and C maps are then calculated by comparing the luminance threshold τ_L and chrominance threshold τ_C with the luminance difference map d_Y and luminance difference map d_C, respectively,

$$L \text{ map}(x,y)=\begin{cases} 1 & d_Y(x,y)>\tau_Y \quad M\leqslant x\leqslant X-M-1 \\ 0 & \text{otherwise} \quad N\leqslant y\leqslant Y-N-1 \end{cases}, \tag{11}$$

$$C \text{ map}(x,y)=\begin{cases} 1 & d_C(x,y)>\tau_C \quad M\leqslant x\leqslant X-M-1 \\ 0 & \text{otherwise} \quad N\leqslant y\leqslant Y-N-1 \end{cases}. \tag{12}$$

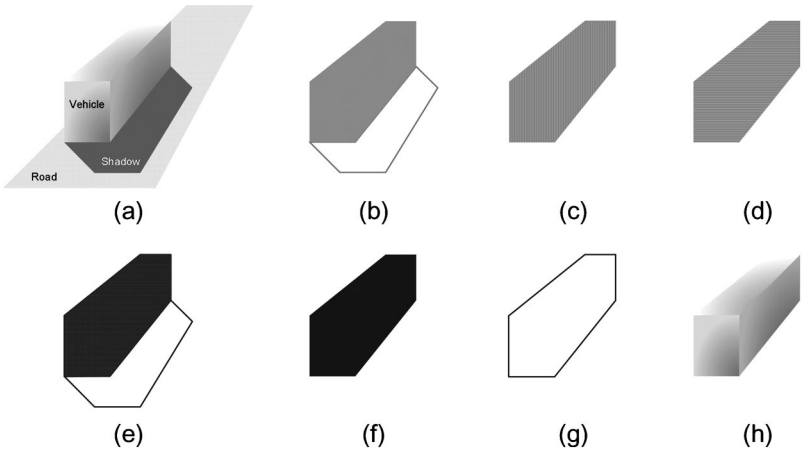


Fig. 4 Illustration of the proposed methodology: (a) input frame f_i; (b) *T* map; (c) *L* map; (d) *C* map; (e) OR map; (f) morphological opening and closing; (g) contour; and (h) extracted vehicle.

The OR map is then computed by a basic logical OR operation of the *T*, *L*, and *C* maps,

$$\text{OR map}(x,y)=T \text{ map}(x,y)+L \text{ map}(x,y)+C \text{ map}(x,y)\begin{matrix}M\leqslant x\leqslant X-M-1\\ N\leqslant y\leqslant Y-N-1\end{matrix}. \quad (13)$$

Finally, the OR map is then operated on by a morphological opening to smooth the contours and remove undesired features, such as background noise, and boundaries between the shadowed and unshadowed regions of the road. Afterward, morphological closing is employed to fuse narrow breaks, long thin gulfs, and eliminate small holes. The contour of the vehicle can be created by subtracting the morphological erosion from the vehicle mask.

To illustrate the proposed methodology, we use Fig. 4 as a demonstration. Figure 4(a) depicts a typical outdoor traffic scene with a vehicle under bright sunlight. In Fig. 4(b), the shaded gray region illustrates the *T* map, which indicates that there is no significant textural difference in the cast shadow region, and that the boundary between the shadowed and unshadowed regions has significant textural difference. In Figs. 4(c) and 4(d), the shaded gray regions illustrate the *L* and *C* maps, respectively. Figure 4(e) illustrates the OR map. The OR-map then undergoes morphological opening and closing to produce a vehicle mask O_V, as shown in Fig. 4(f). Finally, the contour of the vehicle is created by subtracting the morphological erosion from the vehicle mask, as show in Fig. 4(g), and the extracted vehicle is illustrated in Fig. 4(h).

4.4 *Selection of Thresholds*

There are many algorithms for selecting optimal thresholds, such as the Isodata algorithm, background-symmetry algorithm, and triangle algorithm.[21] However, there is no universal approach for threshold selection that is guaranteed to work for all images. In this work, the optimal setting of every parameter τ_T, τ_Y, and τ_C is individually determined by the Isodata algorithm, as it is simple, automatic, and the error rate (ER) is low when compared with human justification. According to Eqs. (7), (9), and (10), the values of d_T, d_Y, and d_C are between 0 and 1, where the highest intensity and lowest intensity of an image frame are 0 and 1, respectively. Therefore, the technique for choosing thresholds for d_T, d_Y, and d_C are the same.

The Isodata algorithm is based on an iterative technique. The values d_T, d_Y, and d_C are first quantized into the number of levels 2^B, where *B* is any positive integer. The histogram of each variable is then constructed. Each histogram is segmented into two parts using a starting threshold value such as $\tau_0=2^{B-1}$, half of the maximum dynamic range. The sample mean $m_{f,0}$ of the difference values associated with the foreground pixels and the sample mean $m_{b,0}$ of the difference values associated with the background pixels are computed by

$$m_{b,0}=\frac{1}{\sum_{i=0}^{2^{B-1}-1}}\sum_{i=0}^{2^{B-1}-1}h[i]i,$$

$$m_{f,0}=\frac{1}{\sum_{i=2^{B-1}}^{2^B-1}h[i]}\sum_{i=2^{B-1}}^{2^B-1}h[i]i, \quad (14)$$

where $h[i]$ is the number of pixels at position *i*. A new threshold value τ_1 is computed as the average of these two sample means. The process is repeated, based on the new threshold,

$$\tau_k=\frac{m_{f,k-1}+m_{b,k-1}}{2}, \quad (15)$$

until the threshold value does not change any more, i.e., $\tau_k=\tau_{k-1}$.

After finding the parameters τ_T, τ_Y, and τ_C by the Isodata algorithm, the *T*, *L*, and *C* maps can be constructed according to Eqs. (8), (11), and (12). Each map may contain background noise or other undesired features, and so morphological erosion and dilation are applied to reduce any false positive in each map, as mentioned in the last section.

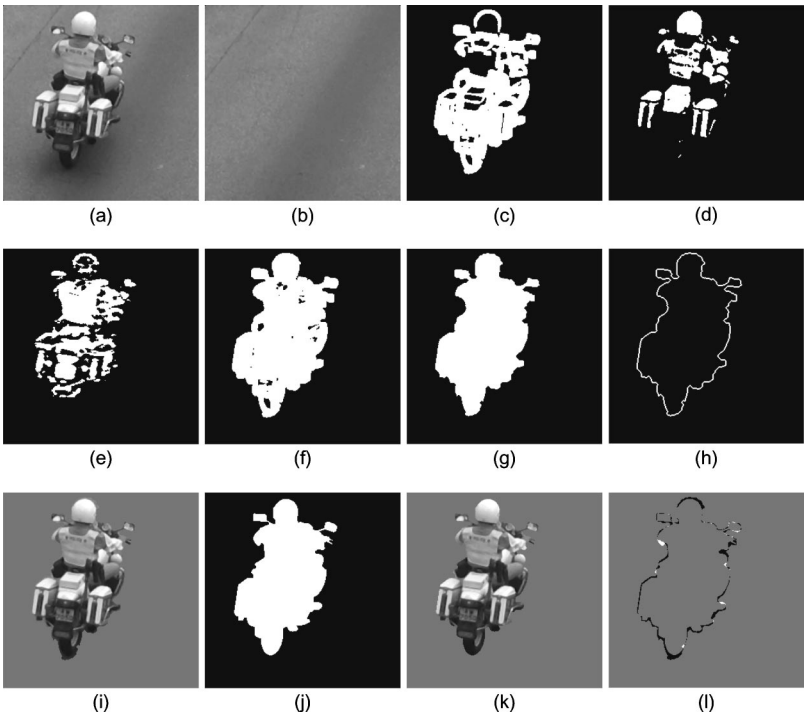


Fig. 5 Police motorcycle sample: (a) input frame f_i; (b) background reference frame f_b; (c) T map; (d) L map; (e) C map; (f) OR map; (g) computed vehicle mask O_V; (h) contour; (i) extracted vehicle; (j) reference vehicle mask O_R; (k) reference vehicle; and (l) error difference mask O_d.

5 Simulation Results and Discussions

Some typical outdoor traffic image sequences on different roads have been captured to test the accuracy and robustness of the proposed method. The image sequences were captured under different lighting conditions, including cloudy, sunny, and different times of day with the camera position either overhead or by the roadside. The proposed method was tested under different lighting conditions, viewing angles, vehicle sizes, and colors. Out of all the images tested, 50 were selected to illustrate the proposed method.

For evaluation purposes, reference vehicle masks O_R of the vehicles without their cast shadows are required for subsequent calculation of classification errors. A reference vehicle mask O_R for each vehicle sample is defined manually by combining the visual observation on the images and the knowledge about the vehicle. The ER of each vehicle sample is calculated according to the number of pixels of the error difference mask O_d and the number of pixels of the reference vehicle mask O_R as

$$\mathrm{ER}=\frac{\Sigma_{x=M}^{X-M-1}\Sigma_{y=N}^{Y-N-1}|O_d(x,y)|}{\Sigma_{x=M}^{X-M-1}\Sigma_{y=N}^{Y-N-1}O_R(x,y)}, \tag{16}$$

where O_d indicates the difference between the reference vehicle mask O_R and the computed vehicle mask O_V as

$$O_d(x,y)=O_R(x,y)-O_V(x,y). \tag{17}$$

Among those 50 vehicle samples, the simulation results of 13 prominent cases are further illustrated in Figs. 5 through 17, and these 13 vehicle samples are further classified under two weather conditions: cloudy and sunny. Under cloudy weather conditions, five vehicle samples have been selected: police motorcycle (Fig. 5), four wheel drive (Fig. 6), taxi (Fig. 7), security car (Fig. 8), and ambulance (Fig. 9). Under sunny weather conditions, eight vehicle samples are selected: sedan (Fig. 10), van (Fig. 11), truck (Fig. 12), mini bus (Fig. 13), bus (Fig. 14), two motorcycles (Fig. 15), three vehicles (Fig. 16), and vehicle occlusion (Fig. 17).

5.1 *Cloudy Weather Conditions*

In the first sample, an input frame f_i of a police motorcycle is shown in Fig. 5(a). In Fig. 5(b), a background reference frame f_b was generated by the background estimation algorithm.[18] In Figs. 5(c) through 5(f), the results of the T, L, C and OR maps are shown, respectively. The background noise can be initially eliminated by performing morphological operations, and the inner boundaries can be removed by performing morphological closing to produce vehicle mask O_V, as shown in Fig. 5(g). The contour can be created by subtracting the morphological erosion from the vehicle mask O_V, as shown in Fig. 5(h). Finally, the extracted vehicle can be bounded by the vehicle mask, as shown in Fig. 5(i). The region that is not extracted is indicated in gray color. A reference vehicle mask O_R is manually defined as shown in Fig. 5(j), and the error difference mask O_D is depicted in Fig. 5(k). The black region indicates that the algorithm mistakenly identifies the background as part of the vehicle, while the white region con-

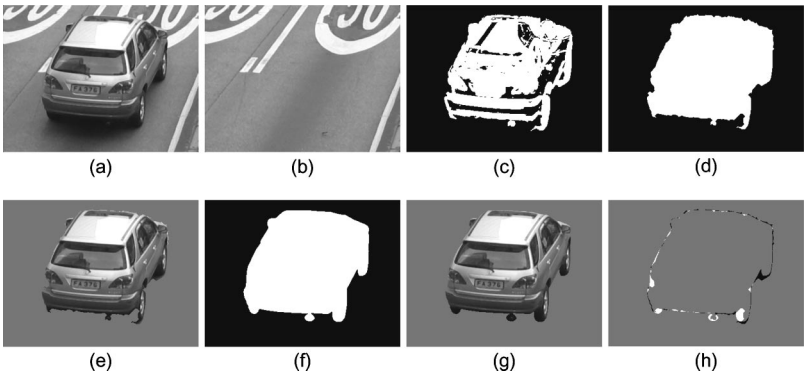


Fig. 6 Four wheel drive sample: (a) f_i; (b) f_b; (c) OR map; (d) O_V; (e) extracted vehicle; (f) O_R; (g) reference vehicle; and (h) O_d.

versely indicates that the algorithm identifies the vehicle as the background. The ER of the police motorcycle sample can then be calculated as

$$\mathrm{ER}=\frac{\Sigma_{x=M}^{X-M-1}\Sigma_{y=N}^{Y-N-1}|O_d(x,y)|}{\Sigma_{x=M}^{X-M-1}\Sigma_{y=N}^{Y-N-1}O_R(x,y)}=\frac{1{,}372}{17{,}150}=8.00\%. \quad (18)$$

The moving motorcycle can be segmented reasonably well with the proposed method while preserving the concavity of the vehicle, as no convex hull is applied to the vehicle mask. Furthermore, most parts of the region corresponding to the black tires can also be extracted without being classified as the cast shadow. However, a narrow background region near to the rear tire is classified as a foreground region. As a whole, a motorcycle is considered to be the most difficult to segment, compared to other vehicles because of the huge amount of detail and concavity, and hence, 8% error.

The proposed method also works well with vehicle color similar to the road, as shown in Fig. 6(a). The final segmented vehicle is depicted in Fig. 6(e). All parts of the gray colored vehicle can be extracted. However, the area of the umbra region behind the vehicle and the rear tires cannot be distinguished well, as depicted in Fig. 6(h), and the ER of the four wheel drive sample is 4.26%.

In the third sample, the results of segmenting a taxi are shown in Fig. 7. The final segmented vehicle is depicted in Fig. 7(e). All parts of the taxi are extracted successfully and the umbra region is automatically removed. The ER of the taxi sample is only 2.91%, as the color of the taxi is red, which is significantly different from the road.

On the road, there are many different types of vehicles. In Figs. 8 and 9, the results of applying our method to a security car and ambulance samples are depicted, respectively. In these two samples, the bodies of the vehicles are white in color with large illumination difference, and so the error differences of both samples are low as illustrated in Fig. 8(d) and Fig. 9(d), respectively. The ER of the security car and the ambulance samples are 2.81 and 1.88%, respectively.

5.2 *Sunny Weather Conditions*

Under sunny weather conditions, an input frame f_i of a sedan sample is shown in Fig. 10(a). In Fig. 10(b), a back-

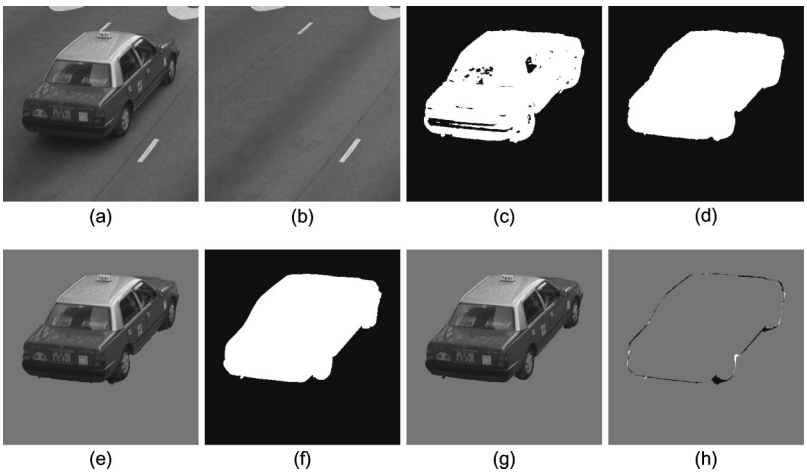


Fig. 7 Taxi sample: (a) f_i; (b) f_b; (c) OR map; (d) O_V; (e) extracted vehicle; (f) O_R; (g) reference vehicle; and (h) O_d.

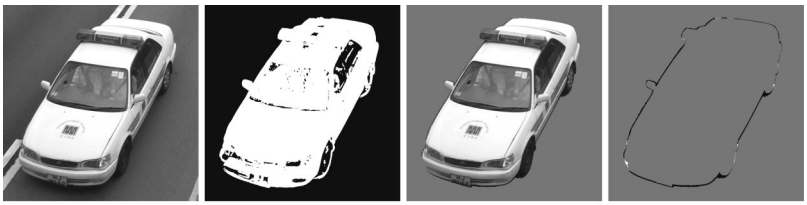


Fig. 8 Security car sample: (a) f_i; (b) OR map; (c) extracted vehicle; and (d) O_d.

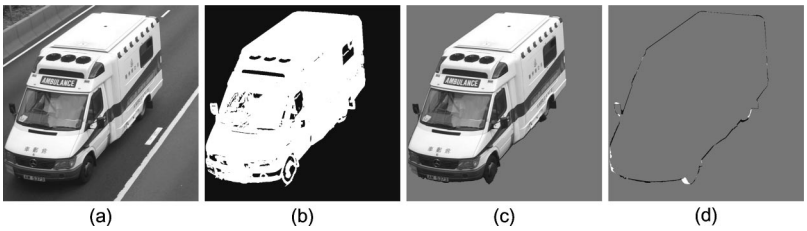


Fig. 9 Ambulance sample: (a) f_i; (b) OR map; (c) extracted vehicle; and (d) O_d.

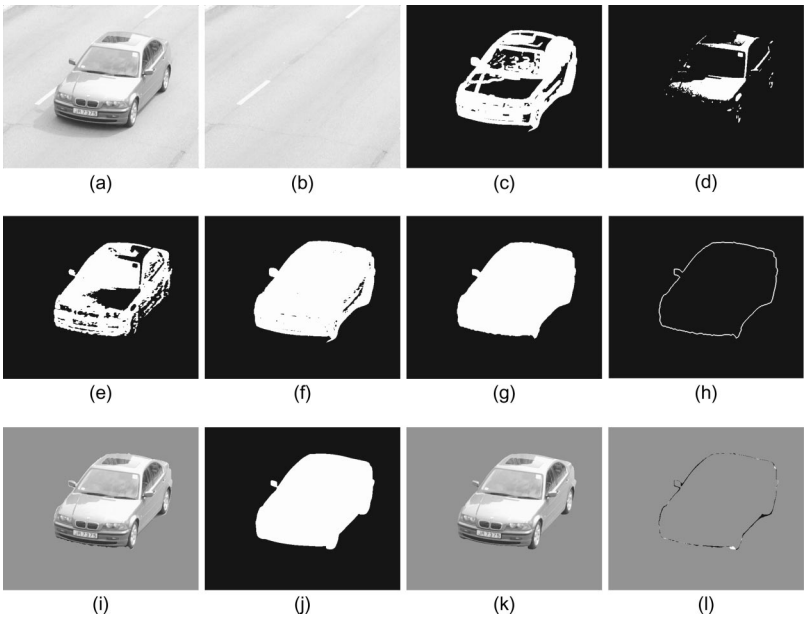


Fig. 10 Sedan sample: (a) f_i; (b) f_b; (c) T map; (d) L map; (e) C map; (f) OR map; (g) morphological opening; (h) O_V; (i) extracted vehicle; (j) O_R; (k) reference vehicle; and (l) O_d.

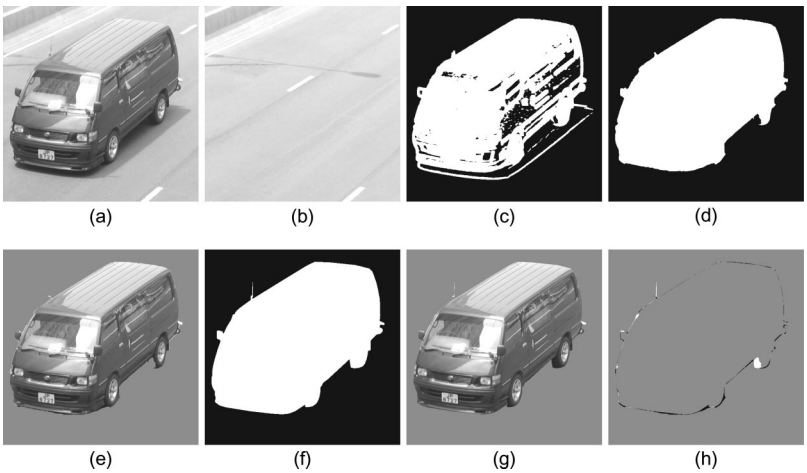


Fig. 11 Van sample: (a) f_i; (b) f_b; (c) OR map; (d) O_V; (e) extracted vehicle; (f) O_R; (g) reference vehicle; and (h) O_d.

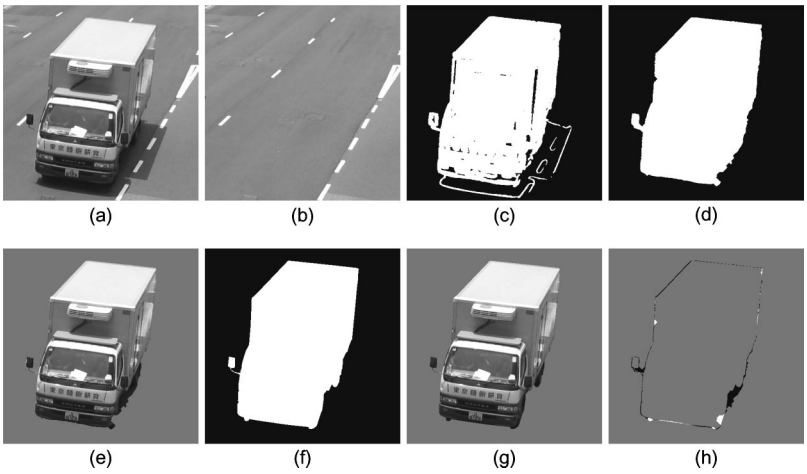


Fig. 12 Truck sample: (a) f_i; (b) f_b; (c) OR map; (d) O_V; (e) extracted vehicle; (f) OR; (g) reference vehicle; and (h) O_d.

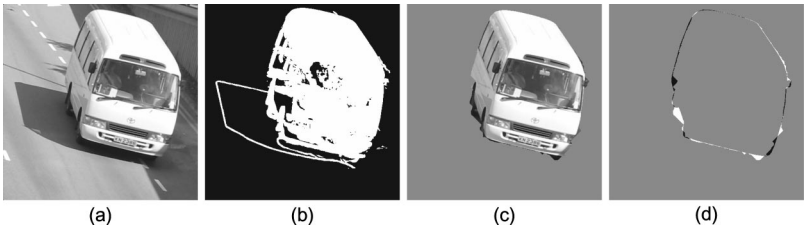


Fig. 13 Minibus sample: (a) f_i; (b) f_b; (c) OR map; (d) O_V; (e) extracted vehicle; (f) O_R; (g) reference vehicle; and (h) O_d.

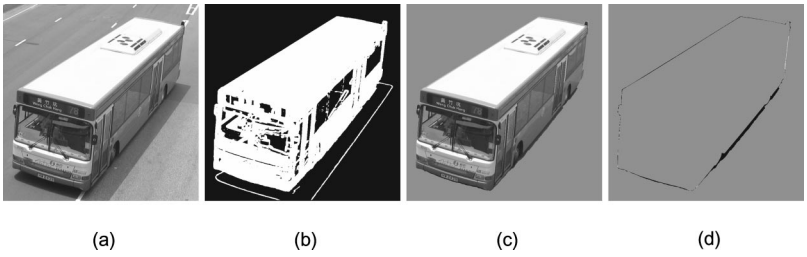


Fig. 14 Bus sample: (a) f_i; (b) OR map; (c) extracted vehicle; and (d) O_d.

Fig. 15 Two motorcycles sample: (a) f_i; (b) OR map; (c) O_V; (d) extracted vehicle; and (e) O_d.

Fig. 16 Three vehicles sample: (a) f_i; (b) OR map; (c) extracted vehicle; and (d) O_d.

ground reference frame f_b was generated by the background estimation algorithm.[18] In Figs. 10(c) through 10(f), the results of the *T*, *L*, *C*, and OR map are shown, respectively. It can be observed that the luminance of the input image is always lower than the background image in the cast shadow region. The vehicle mask and the contour can be determined by performing the morphological operations as shown in Figs. 10(g) and 10(h), respectively. Finally, the segmented vehicle can be bounded by the vehicle mask as shown in Fig. 10(i). The region that is not extracted is indicated in gray color. Again, a reference vehicle mask O_R is manually defined as shown in Fig. 10(j), and the error difference mask O_D is depicted in Fig. 10(k), where the ER of the sedan sample is 2.34%. The cast shadow can be well separated by our method while preserving the concavity of the vehicle. However, a narrow background region at the side of the vehicle is also detected as a foreground region.

In the van sample, a dark blue van is under consideration as shown in Fig. 11(a). Even the luminance values of the vehicle are lower than the cast shadow. The final segmented vehicle is depicted in Fig. 11(e). All parts of the dark blue vehicle can be extracted, and the cast shadow of the vehicle was successfully removed. However, the region of the rear tire cannot be extracted as part of the vehicle and the ER of van sample is 1.63%.

In the next sample, the results of segmenting a truck are shown in Fig. 12. The final segmented vehicle is depicted in Fig. 12(e). All parts of the truck are extracted successfully, and the cast shadow region is automatically removed. However, the background region at the side of the vehicle is mistaken as the foreground region. The ER of the truck sample is 3.12%.

In Figs. 13 and 14, the results of applying our method to a minibus and bus sample are depicted, respectively. Our proposed method can easily detect less concave vehicles. In the minibus case, there is a side road building. The shadow is reflected on the minibus, as shown in Fig. 13(a), and so the ER of the minibus sample is 3.44%. The number of pixels of the segmented bus is relatively large, and the error difference is consequently low, as illustrated in Fig. 14(d), respectively. The ER of the bus sample is 2.21%.

In the next sample, the results of segmenting two motorcycles are shown in Fig. 15. The final segmented vehicle is depicted in Fig. 15(d). Both motorcycles contain concavity, and so the error rates of the top left and bottom right motorcycles are 7.35% and 6.33%, respectively.

In Fig. 16, a situation showing several vehicles per images as in current traffic situations was analyzed. Our proposed method is capable of segmenting multiple vehicles. In this sample, three vehicles were segmented based on our methods. The error rates of the top left four wheel drive, bottom left sedan, and right taxi are 5.84, 6.14, and 4.62%, respectively.

In the last sample, the result of applying our method to a vehicle occlusion sample is depicted in Fig. 17. Potentially, our proposed method is capable of detecting occluded vehicles. However, the partition of two vehicles is out of the scope of this work. In this sample, the error differences is low, as illustrated in Fig. 17(d), and the ER is 2.24%.

5.3 *Summary*

The error rates of all 50 cases based on the number of pixels of the reference vehicles are statistically plotted in Fig. 18. A number of observations can be made. First, as the number of pixels of the reference vehicle increase, the error rate decreases. Second, most of the error points cluster between 2 and 4%. Third, the error points corresponding to large errors are those smaller vehicles usually with plenty of details, concavity, and complex vehicle outlines associated with them. They have the reputation of being difficult to be segmented. The average ER of 50 different tested vehicle samples is 3.44%.

6 Conclusions

We present a highly accurate texture-based segmentation method for extracting moving vehicles, which can effectively separate the cast shadow from the vehicle under dif-

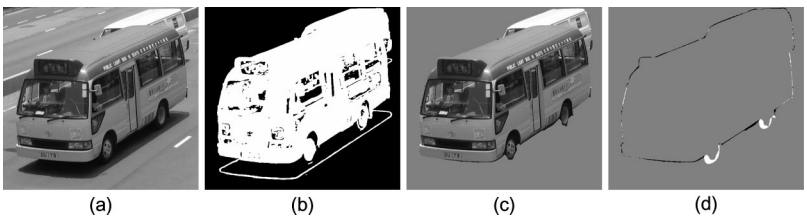


Fig. 17 Vehicle occlusion sample: (a) f_i; (b) OR map; (c) extracted vehicle; and (d) O_d.

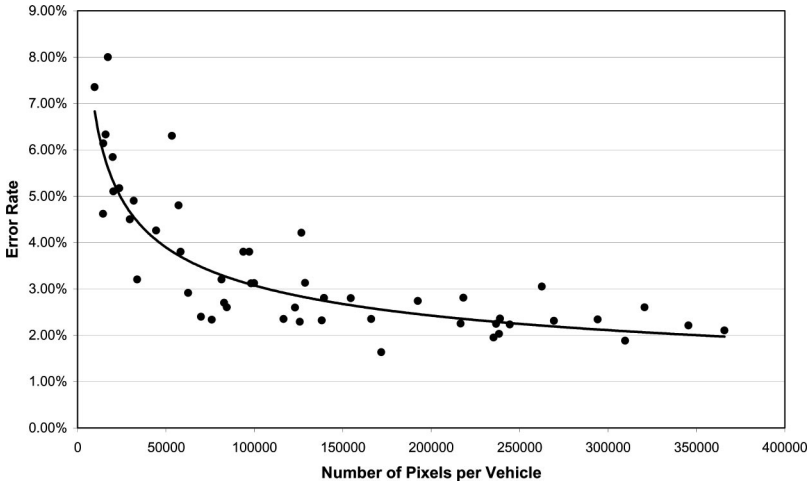


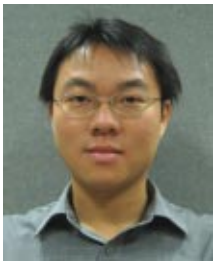
Fig. 18 Error rate versus number of pixels of reference vehicle based on 50 tested vehicle samples.

ferent environments and vehicle colors. In this method, texture difference is the key to differentiating the shadow from the vehicle, as well as associating the shadow and the road. Used in conjunction with the luminance and chrominance difference, an OR map can be calculated, representing a consensus between these features. A foreground mask can be subsequently computed by performing the morphological operations of the OR map. We have tested our proposed method on different vehicle samples under typical outdoor scenes. Our proposed method is shown to be successful for various outdoor daylight conditions and vehicles. The simulation demonstrated that moving vehicles can be segmented very accurately while preserving vehicle concavity. It also has an advantage that cast shadow regions are automatically removed, as they have the same textural property as the background reference frame. From our results and analysis on various vehicle samples, we have found that our proposed method is reasonably robust in different outdoor daylight conditions and for different vehicles; the average error rate of 50 different vehicle samples is 3.44%. This figure compares well with any existing segmentation methods that work on similar types of images.

References

1. N. Yung and A. Lai, ''A system architecture for visual traffic surveillance,'' *5th World Congress Intell. Transport Syst.*, ITS Congress Association (1998).
2. R. Cucchiara, C. Crana, M. Piccardi, A. Prati, and S. Sirotti, ''Improving shadow suppression in moving object detection with HSV color information,'' *IEEE Conf. Intell. Transportation Syst.*, pp. 334–339 (2001).
3. R. Cucchiara, M. Piccardi, and P. Mello, ''Image analysis and rule-based reasoning for a traffic monitoring system,'' *IEEE Trans. Intell. Transportation Syst.* **1**(2), 119–130 (2000).
4. R. Cucchiara, M. Piccardi, A. Prati, and N. Scarabottolo, ''Real-time detection of moving vehicles,'' *Intl. Conf. Image Anal. Process.*, pp. 618–623 (1999).
5. N. D. Doulamis, A. D. Doulamis, Y. Avrithis, and S. D. Kollias, ''A stochastic framework for optimal key frame extraction from MPEG video databases,'' *IEEE 3rd Workshop Multimedia Sig. Process.*, pp. 141–146 (1999).
6. A. Elgammal, D. Harwood, and L. Davis, ''Non-parametric model for background subtraction,'' *6th European Conf. Computer Vis.*, pp. 751–767 (2000).
7. 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 traffic image sequences,'' *Opt. Eng.* **41**(6), 1425–1440 (2002).
8. A. H. S. Lai, G. S. K. Fung, and N. H. C. Yung, ''Vehicle type classification from visual-based dimension estimation,'' *IEEE Conf. Intell. Transportation Syst.*, pp. 201–206 (2001).
9. I. Mikic, P. C. Cosman, G. T. Kogut, and M. M. Trivedi, ''Moving shadow and object detection in traffic scenes,'' *Proc. Intl. Conf. Patt. Recog.*, pp. 321–324 (2000).
10. A. Prati, I. Mikic, C. Grana, and M. M. Trivedi, ''Shadow detection algorithms for traffic flow analysis: a comparative study,'' *Proc. IEEE Intell. Transportation Syst.*, pp. 340–345 (2001).
11. A. N. Rajagopalan and R. Chellappa, ''Vehicle detection and tracking in video,'' *Intl. Conf. Image Process.* **1**, 351–354 (2000).
12. A. Branca, G. Attolico, and A. Distante, ''Cast shadow removing in foreground segmentation,'' *16th Intl. Conf. Patt. Recog.* **1**, 214–217 (2002).
13. G. Funka-Lea and R. Bajcsy, ''Combining color and geometry for the active, visual recognition of shadows,'' in *Proc. Intl. Conf. Computer Vis.*, pp. 203–209 (1995).
14. E. Salvador, A. Cavallaro, and T. Ebrahimi, ''Shadow identification and classification using invariant color models,'' *IEEE Intl. Conf. Acoustics, Speech, Sig. Process.* **3**, 1545–1548 (2001).
15. J. Stauder, R. Mech, and J. Ostermann, ''Detection of moving cast shadows for object segmentation,'' *IEEE Trans. Multimedia* **1**(1), 65–76 (1999).
16. R. Haralick, ''Statistical and structural approaches to texture,'' *Proc. IEEE* **67**(5), 786–804 (1979).
17. M. Yamada, K. Ueda, I. Horiba, and N. Sugie, ''Discrimination of the road condition toward understanding of vehicle driving environments,'' *IEEE Trans. Intell. Transportation Syst.* **2**(1), 26–31 (2001).
18. A. H. S. Lai and N. H. C. Yung, ''A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence,'' *Proc. IEEE Intl. Symp. Circuits Syst.* **4**, 241–244 (1998).
19. M. Sonka, V. Hlavac, and R. Boyle, ''Texture,'' Chap. 14 in *Image Processing, Analysis, and Machine Vision*, pp. 646–678, PWS Publishing, Pacific Grove, CA (1999).
20. P. Kumar, K. Sengupta, and A. Lee, ''A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system,'' *Proc. IEEE Intell. Transportation Syst.*, pp. 100–105 (2002).
21. A. Rydberg and G. Borgefors, ''Integrated method for boundary delineation of agricultural fields in multispectral satellite images,'' *IEEE Trans. Geosci. Remote Sens.* **39**(11), 1678–1680 (2000).

William Wai Leung Lam received his BEng degree with distinction in electrical engineering from McMaster University, Canada, and his MPhil degree in electrical and electronic engineering from the Hong Kong University of Science and Technology. He is currently a PhD candidate in the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include digital image processing, texture synthesis, intelligent transportation systems, and computer networks.



Clement Chun Cheong Pang received the BEng and MEng degrees in electrical engineering from McMaster University, Ontario, Canada, in 1999 and 2001, respectively. He is currently pursuing the PhD degree at the University of Hong Kong. His research interests include visual traffic surveillance, electrocardiogram (ECG) signal processing, as well as the application of artificial neural networks on electroencephalogram (EEG) signals.

Nelson H. C. Yung received his BSc and PhD degrees from the University of Newcastle-Upon-Tyne. He was a lecturer at the same university from 1985 to 1990. From 1990 to 1993, he worked as a senior research scientist at the Department of Defence, Australia. He joined the University of Hong Kong in late 1993 as an associate professor. He leads a research team in digital image processing and intelligent transportation systems. He is the founding Director of the Laboratory for Intelligent Transportation Systems Research at Hong Kong University (HKU), and also Deputy Director of HKU's Institute of Transport Studies. He has coauthored a computer vision book, and has published more than 100 journal and conference papers in the areas of digital image processing, parallel algorithms, visual traffic surveillance, and autonomous vehicle navigation and learning algorithms.