<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Vehicle feature extraction by patch-based sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Lam, WWL; Pang, CCC; Yung, NHC</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Proceedings Of Spie - The International Society For Optical Engineering, 2003, v. 5022 II, p. 921-932</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2003</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/46361">http://hdl.handle.net/10722/46361</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; S P I E - the International Society for Optical Engineering. Copyright © S P I E - International Society for Optical Engineering.; Copyright 2003 Society of Photo-Optical Instrumentation Engineers. This paper was published in Image and Video Communications and Processing, Electronic Imaging: Science and Technology Proceedings, Santa Clara, California, USA, 21-24 January 2003, v. 5022, p. 921-932 and is made available as an electronic reprint with permission of SPIE. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.</td>
</tr>
</tbody>
</table>
Vehicle Feature Extraction by Patch-Based Sampling

William W. L. Lam*, Clement C. C. Pang, Nelson H. C. Yung
Department of Electrical & Electronic Engineering
The University of Hong Kong, Hong Kong SAR

ABSTRACT

In modern traffic surveillance, computer vision methods are often employed to detect vehicles of interest because of the rich information content contained in an image. In this paper, we propose an efficient method for extracting the boundary of vehicles free from their moving cast shadows and reflective regions. The extraction method is based on the hypothesis that regions of similar texture are less discriminative, disregarding intensity differences between the vehicle body and the cast shadow or reflection on the vehicle. In this novel algorithm, a united likelihood map that based on the relationship of texture, luminance and chrominance of each pixel is initially constructed. Subsequently, a foreground mask is constructed by applying morphological operations. Vehicles can be successfully extracted and different vehicle components can be efficiently distinguished by the related autocorrelation index within the vehicle mask.

Keywords: visual traffic surveillance, vehicle extraction, object segmentation, shadow detection, texture analysis

1. INTRODUCTION

With the rapid development of computer and communication technologies, the research and development of Intelligent Transportation Systems (ITS) techniques are becoming more and more ambitious and substantial [19]. Among the many facets of ITS, visual traffic surveillance plays an important role in traffic data capturing, incident detection and safety management in general. It is able to convey a comprehensive content of information that is easy for human interpretation. Such information can also be interpreted by machines if appropriate algorithms are available. Being able to interpret visual information by machines not just 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 interpret the content of an image or image sequence. In order to do that, features of vehicles in an image must first be extracted for computing mean speed, flow rate, incidents, among other traffic parameters. As such, vehicle extraction has always been one of the major components of visual traffic surveillance recently [5, 6].

Segmentation algorithms that extract the vehicle of interest from the image background in an image sequence have recently been actively pursued [9, 17]. In many of these algorithms, the detection of vehicles is mainly based on their speed, dimension, luminance, chrominance and edge features from an image or image sequence [3, 4, 10, 13]. Unfortunately, these approaches suffer a major drawback that when they are applied to outdoor scenes, undesired object features, such as tree branches, cast shadows on the road, are extracted together with the vehicles. This also creates a series of problems associated with occlusion if the cast shadows are not detected and eliminated. In order to accurately detect the vehicle, numerous shadow detection methods have been proposed [1, 2, 7, 12, 15, 16]. They all suffer from a number of limitations such as specific weather conditions are required that make them ineffective in practical outdoor environments or some of them are limited to indoor environments only.

Texture analysis is an alternative way to extract vehicles effectively. A proof of the role of textural information in outdoor object recognition was done by comparison of classification correctness [8]. If textural information was used to classify an outdoor object, 99% of accuracy was achieved. Conversely, spectral information-based classification achieved only 74% of accuracy. Therefore, we propose an efficient algorithm that based on texture analysis for extracting vehicles in this paper. We assume that the textural features of moving objects are significantly different from the background. However, the studies of texture analysis of road condition and vehicles are limited [18], for that reason, a problem analysis of using texture analysis technique is briefly described in the next section. Following that, the proposed methodology is introduced in section 3. Simulation results and discussion are given in section 4. Finally, the conclusion is drawn in section 5.

*willam@eee.hku.hk; phone 852 2857-8414
2. 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 [3, 4, 9, 10, 17]. Subsequently, unwanted objects such as the cast shadow have to be removed by post-processing techniques. In order 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 those of the corresponding pixels in the background image and the chrominance values of the cast shadow pixels on the other hand, are identical or only slightly different from those of the corresponding pixels in the background image. Besides, we observed that textural feature of cast shadow is only slightly different from those of the corresponding pixels in the background image. In other words, the textural property of the background is substantially altered by the cast shadow of a foreground object. To understand this better, let us consider this textural 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 texture consists of small primitives. If primitives are placed periodically in an image, the autocorrelation function is also periodic. Texture description of an image patch is commonly calculated using the following autocorrelation function \( R(u,v) \),

\[
R(u,v) = \frac{(2M + 1)(2N + 1)}{(2M + 1 - u)(2N + 1 - v)} \sum_{m=0}^{2M} \sum_{n=0}^{2N} \sum_{m=0}^{2M} \sum_{n=0}^{2N} R_{ij} \cdot R_{ij+m+n}
\]

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

\[
R = \mathcal{F}^{-1} \{ |F|^2 \},
\]

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

\[
R(u,v) = \frac{\sum_{m=0}^{2M} \sum_{n=0}^{2N} p(m,n) \cdot p((2M + 1) \mod (m+u), (2N + 1) \mod (n+v))}{\sum_{m=0}^{2M} \sum_{n=0}^{2N} p^2(m,n)}
\]

Essentially, the textural relationship between image frames can be evaluated by the autocorrelation function \( R \) as shown in Figure 1. Figure 1 depicts three image frames \( f_1, f_2, f_3 \), taken by a static camera and the autocorrelation results for image patches A and B of size \( M = N = 16 \) denoting the same locations of three image frames \( f_1, f_2, f_3 \). The profiles of the autocorrelation function \( R \) of frame \( f_1 \) and frame \( f_3 \) for image patch A are very similar as there is no moving object. Moreover, the profiles of the autocorrelation function \( R \) of frame \( f_1 \) and frame \( f_2 \) for image patch B are also very similar even the cast shadow of a bus is projected on the road of frame \( f_2 \). In frame \( f_3 \), a dark colored vehicle is at the center of the image and both image patches A and B now cover part of the vehicle. The profiles of the autocorrelation function \( R \) for image patch A and image patch B of frame \( f_3 \) are drastically different from those of frame \( f_1 \) and frame \( f_2 \) even the luminance values are lower in image patch A of frame \( f_1 \) than image patch B of frame \( f_2 \).

The autocorrelation difference \( d_R \) between two image patches can be calculated by the square difference of two autocorrelation functions \( R \) to compare their similarities,

\[
d_R(u,v) = \left[ R(u,v) - R_1(u,v) \right]^2.
\]
Figure 1: (a) Image frame $f_1$; Autocorrelation function $R$ of frame $f_1$ for: (b) image patch A and (c) image patch B. (d) Image frame $f_2$; Autocorrelation function $R$ of frame $f_2$ for: (e) image patch A and (f) image patch B. (g) Image frame $f_3$; Autocorrelation function $R$ of frame $f_3$ for: (h) image patch A and (i) image patch B.

Figure 2: Autocorrelation differences $d_R$ in image patch A between: (a) frame $f_1$ and frame $f_2$, (b) frame $f_1$ and frame $f_3$, (c) frame $f_2$ and frame $f_3$. Autocorrelation differences in image patch B between: (d) frame $f_1$ and frame $f_2$, (e) frame $f_1$ and frame $f_3$, (f) frame $f_2$ and frame $f_3$. 

Proc. of SPIE Vol. 5022 923
The texture difference $d_T$ between two image patches can be simply calculated by mean square difference of two autocorrelation functions $R$ to compare their similarities,

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

where $R_i, R_j$ are the autocorrelation functions $R$ of two different image patches. The texture differences $d_T$ of image patch A and image patch B between those frames $f_1, f_2, f_3$ are summarized in Table 1. The texture difference $d_T$ between frame $f_1$ and frame $f_2$ for image patch 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 patch B is also relatively low even the cast shadow of a bus is projected on the road of frame $f_2$. Therefore, our observation is satisfied: textural features are only slightly different from those of the corresponding pixels in the background image.

<table>
<thead>
<tr>
<th>Texture Difference $d_T$</th>
<th>Image Patch A</th>
<th>Image Patch B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame $f_1$ and Frame $f_2$</td>
<td>$2.890 \times 10^{-6}$</td>
<td>$1.724 \times 10^{-3}$</td>
</tr>
<tr>
<td>Frame $f_1$ and Frame $f_3$</td>
<td>$9.272$</td>
<td>$1.923 \times 10^{-2}$</td>
</tr>
<tr>
<td>Frame $f_2$ and Frame $f_3$</td>
<td>$9.265$</td>
<td>$1.071 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

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

3. METHODOLOGY

3.1 Assumptions
In our extraction methodology, four assumptions are made with respect to the extraction of vehicles. First, the camera is assumed to be stationary and the background is assumed to be stationary too and contains texture primitives, such as the road surface. Second, the light source is assumed to be single and strong, thus illumination difference between the shadow and background is large 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. Forth, same elements have similar texture, and therefore textural features of the different vehicle components are significantly different.

3.2 Method Overview
Under these four assumptions, the moving objects of an input image frame $f_i$ can be reasonably extracted from the background, where the background reference frame $f_b$ is estimated by the running mode and running average algorithms [11]. In the proposed extraction algorithm as depicted in Figure 3, three likelihood maps $L_T, L_Y, L_C$ are initially created according to the differences in texture, luminance and chrominance between the input image frame $f_i$ and background reference frame $f_b$. A united likelihood map $L_{ut}$ is then constructed by performing a logical OR operation of those likelihood maps $L_T, L_Y, L_C$. Finally, a foreground mask is constructed 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.

Figure 3: Overview of proposed vehicle extraction method.
After extracting the foreground mask, different components within the mask can be further segmented by simply lowering the number of quantization levels of a foreground object. Many segmentation methods have been proposed [9, 17]. However, their segmented regions cannot be further categorized. With the description of textural feature in each region, different components can be further categorized based on the similarities between segmented regions as depicted in Figure 4.

**Figure 4:** Categorizing vehicle components from an extracted mask.

### 3.3 Details of the Method

The first task of vehicle extraction is to construct a texture likelihood map $L_T$. In this step, each pixel with its neighborhood from an input image frame $f_i$ is transformed to be defined an input image patch $p_i$, and its same neighborhood from a background reference frame $f_b$ are transformed to be defined a background image patch $p_b$.

$$ p_{i(x,y)} (m,n) = f(x+m-M, y+n-N) \quad 0 \leq m \leq 2M $$
$$ \quad 0 \leq n \leq 2N. \quad (6) $$

A map of texture difference $d_T$ is then constructed as shown in Figure 5 according to the mean square difference of two autocorrelation functions $R$ of each input image patch $p_i$ with the same location of background image patch $p_b$.

$$ d_T(x,y) = \frac{1}{(2M+1)(2N+1)} \sum_{u=0}^{2M} \sum_{v=0}^{2N} \left[ R_{i(x,y),i} (u,v) - R_{i(x,y),b} (u,v) \right]^2 $$
$$ \quad M \leq x \leq X - M - 1 $$
$$ \quad N \leq y \leq Y - N - 1. \quad (7) $$

**Figure 5:** Constructing a texture difference map $d_T$. 

---

Proc. of SPIE Vol. 5022     925
A texture likelihood map $L_T$ is then computed by comparing the texture threshold $\tau_T$ with the texture difference map $d_T$.

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

(8)

The second task of vehicle extraction is to construct a luminance likelihood map $L_Y$ and a chrominance likelihood map $L_C$. In this step, the color model YCbCr is used to separate the luminance and chrominance components of the images [15]. A luminance difference map $d_Y$ between the input image frame $f_i$ and the background reference frame $f_b$ is constructed according to the following equation,

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

(9)

and a chrominance difference map $d_C$ between input image frame $f_i$ and background reference frame $f_b$ is constructed according to the following equation,

$$
d_C(x, y) = \left[ C_b(x, y) - C_b(x, y) \right]^2 + \left[ C_r(x, y) - C_b(x, y) \right]^2
$$

\[ M \leq x \leq X - M - 1 \]

\[ N \leq y \leq Y - N - 1. \]

(10)

Both the luminance likelihood map $L_Y$ and chrominance likelihood map $L_C$ are then calculated by comparing the luminance threshold $\tau_L$ and chrominance threshold $\tau_C$ with the luminance difference map $d_Y$ and luminance difference map $d_C$ respectively,

$$
L_Y(x, y) = \begin{cases} 
1 & d_Y(x, y) > \tau_L \quad M \leq x \leq X - M - 1 \\
0 & \text{otherwise} \quad N \leq y \leq Y - N - 1.
\end{cases}
$$

(11)

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

(12)

Finally, the united likelihood map $L_U$ is computed by a basic logical OR operation of the texture likelihood map $L_T$, luminance likelihood map $L_Y$ and chrominance likelihood map $L_C$,

$$
L_U(x, y) = L_T(x, y) + L_Y(x, y) + L_C(x, y)
$$

\[ M \leq x \leq X - M - 1 \]

\[ N \leq y \leq Y - N - 1. \]

(13)

There are many algorithms for selecting optimal threshold such as isodata algorithm, background-symmetry algorithm and triangle algorithm. However, there is no universal approach for threshold selection that is guaranteed to work for all images. In this paper, the optimal setting of parameters $\tau_T$, $\tau_Y$ and $\tau_C$ are determined by the isodata algorithm as isodata algorithm is simple, automatic and the error rate is low when compared with human justification [14]. According to equations (7), (9) and (10), the values of texture difference map $d_T$, luminance difference map $d_Y$ and chrominance difference map $d_C$ 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 texture difference map $d_T$, luminance difference map $d_Y$ and chrominance difference map $d_C$ are the same.

Isodata algorithm is based on an iterative technique. The values of texture difference map $d_T$, luminance difference map $d_Y$ and chrominance difference map $d_C$ are firstly quantized into the number of levels $2^B$, where $B$ is any positive integer. The histogram of each difference map is then constructed. Figure 6 depicts the histogram of luminance difference map $d_Y$ of motorcycle test case in section 4. 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,

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

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

(14)

where $h[i]$ is number of pixels at position $i$. A new threshold value $\tau_1$ is computed as the average of these two sample means. The process is repeated, based upon the new threshold,
τ_k = \frac{m_{r,k-1} + m_{b,k-1}}{2}, 
\text{ until the threshold value does not change any more, } \tau_k = \tau_{k-1}. 

Figure 6: Histogram of luminance difference map \( d_Y \) with \( B = 8 \) for the motorcycle test case in section 4.

4. SIMULATION RESULTS AND DISCUSSION

Some typical outdoor traffic image sequences on different roads in Hong Kong have been captured in order to test the effectiveness and robustness of the proposed method. The image sequences were captured under different lighting conditions; including sunny, cloudy, and different time of the 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. Some of the images were selected to illustrate the working of the proposed method.

In the first case, an input image frame \( f_i \) of a motorcycle is shown in Figure 7(a). In Figure 7(b), a background reference frame \( f_b \) was generated by the background estimation algorithm [11]. In Figures 7(c), 7(d), 7(e) and 7(f), the results of texture likelihood map \( L_T \), luminance likelihood map \( L_Y \), chrominance likelihood map \( L_C \) and united likelihood map \( L_U \) are shown respectively. The vehicle mask can be created by performing the morphological operations: The background noise can be initially detached by performing morphological erosion as shown in Figure 7(g), and then the inner boundaries can be removed by performing morphological dilation on opening as shown in Figure 7(h). Finally the vehicle mask can be restored by performing morphological closing of opening as shown in Figures 7(i). The contour of the vehicle can be created by subtracting the morphological erosion from the vehicle mask as shown in Figure 7(j) and the extracted vehicle can be bounded by the vehicle mask as shown in Figure 7(k). The moving motorcycle can be well extracted with our method while preserving the concavity of the vehicle as the convex hull is not applied to the vehicle mask. Also the region corresponding to the black tires can also be extracted without being classified as the region of cast shadow. However, a narrow background region near to the tires and a small part of background region next to the left side mirror are also extracted as part of the vehicle.

Our method also works well with dark colored vehicle as shown in Figure 8(a) even the luminance values of the vehicle are lower than the cast shadow. The final extracted vehicle is depicted in Figure 8(h). All parts of the dark colored vehicle can be extracted and the cast shadow near to the end of the vehicle was successfully removed. However, the area of the umbra region near to the tires and a small part of background region next to the left side mirror are also extracted as part of the vehicle.
Figure 7: Motorcycle case: (a) Input image frame $f_i$; (b) Background reference frame $f_b$; (c) Texture likelihood map $L_T$; (d) Luminance likelihood map $L_Y$; (e) Chrominance likelihood map $L_C$; (f) United likelihood $L_U$; (g) Morphological erosion; (h) Morphological dilation on opening; (i) Vehicle mask; (j) Contour; (k) Extracted vehicle.
In the third case, the results of extracting a taxi under different lighting conditions are shown in Figure 9. The light source was only partially projected on the road as there were some buildings along the road where their cast shadows can be seen in Figure 9(b). The final extracted vehicle is depicted in Figure 9(h). All parts of the taxi are extracted successfully and the cast shadow is automatically removed. However, the area of the umbra region at the back of the taxi is also extracted as foreground region.

In the forth case, the results of extracting a white vehicle under a cloudy weather condition are shown in Figure 10. The final extracted vehicle is shown in Figure 10(f). All parts of the vehicle are extracted, but the umbra regions beside and in front of the white vehicle, and a small part of background region next to the left side mirror are also extracted as foreground region. Further classification of segmented regions of the white vehicle is shown in Figure 11.
Figure 9: Taxi case: (a) Input image frame $f_i$; (b) Background reference frame $f_b$; (c) United Likelihood map $L_U$; (d) Morphological erosion; (e) Morphological dilation on opening; (f) Vehicle mask; (g) Contour; (h) Extracted vehicle.

Figure 11(a) shows different segmented regions by lowering the number of quantization levels. However different regions may belong to the same component of the vehicle. For example, Regions A, B and C are segmented into three regions, but regions A and B should be considered as the one region, and therefore the evaluation of three regions are A, B and C calculated by the autocorrelation difference as shown in Figures 11(a), 11(b) and 11(c) respectively. It is obvious that the profiles of autocorrelation function $R$ of region A and region B are similar, but significantly different than the profile of autocorrelation function $R$ of region C.
Figure 10: White vehicle case: (a) Input image frame $f_i$; (b) Background reference frame $f_b$; (c) United Likelihood map $L_U$; (d) Vehicle mask; (e) Contour; (f) Extracted vehicle.

Figure 11: Region classification. (a) Segmented regions of the white vehicle; Profiles of autocorrelation function $R$ of region: (b) A, (c) B, (d) C.
5. CONCLUSIONS

In this paper we have presented an efficient method for extracting moving objects, which can effectively separate the cast shadow from the vehicle under different environment and vehicle color. In this method, a united likelihood map is constructed based on three different domains: texture, luminance and chrominance. A foreground mask can be subsequently computed by performing the morphological operations of the united likelihood map. Also, segmented regions can be categorized into different components according to the autocorrelation index. We have tested our proposed method on different vehicle samples under typical outdoor scenes. Our proposed method is tested to be successful for various outdoor daylight environments and vehicles. Our extensive simulations demonstrated that moving objects can be well extracted while preserving object concavity. It also has an advantage that cast shadow regions are automatically removed as they have the same textural feature as the background reference frame.

6. REFERENCES