Nighttime Vehicle Detection and Tracking Base on Spatiotemporal Analysis using RCCC Sensor

Han-Wen Huang
Applied Sensor Technology Group
Automotive Research and Testing Center
Changhua, Taiwan (R.O.C.)
stanny@artc.org.tw

Chuan-Ren Lee
Applied Sensor Technology Group
Automotive Research and Testing Center
Changhua, Taiwan (R.O.C.)
harris@artc.org.tw

Hung-Pang Lin
Applied Sensor Technology Group
Automotive Research and Testing Center
Changhua, Taiwan (R.O.C.)
hplin@artc.org.tw

Abstract—With the consciousness of driving safety growing increasingly, traffic and transportation industry have been devoted to developing Advanced Driver Assistance System (ADAS) recently to help drivers observing the variation of the environmental conditions around the vehicle. Monocular camera is the essential sensor in ADAS to capture the information around the vehicle. Therefore, vision based technology is the main methods to detect obstacles. ADB systems typically exploit a camera that detect the front road view, if the system detect preceding vehicles or oncoming vehicles, the distribution of beam will changed to avoid glare other driver’s eyes. Consequently, vehicle detection at nighttime condition is the critical part of an ADB system. This study implements a nighttime vehicle detection method exploiting a RCCC sensor to replace Bayer sensor and aim to obtain a better quality of images at night. Candidate region selection is based on spatiotemporal analysis and three global verifications are adopted to reduce the false detecting results. A new adaptive tracking by-detection framework based on structured output prediction is applied with the detection process.

Keywords—ADAS, RCCC sensor, Vehicle detection, Spatiotemporal analysis, Tracking

I. INTRODUCTION

With the consciousness of driving safety grow increasingly, traffic and transportation industry have been devoted to developing Advanced Driver Assistance System (ADAS) recently to help drivers to observe the variation of the environmental condition around the vehicle. Such as Lane Departure Warning (LDW), Adaptive Cruise Control (ACC), and Pedestrian Detection System (PDS) etc., the goal of ADAS is to alert drivers under dangerous situations to prevent traffic accidents.

Besides radar and Lidar, monocular camera is the essential sensor in ADAS to capture the information around the vehicle. Therefore, vision based technology is the main methods to detect obstacles like vehicles, pedestrian, cyclists and motorists on road. According to National Safety Council, people are three times more likely to be killed in vehicle accidents at night than during the daytime. The lack of natural light results in most car accidents. Although the street lights can help to lighten the forward view, drivers may still experience decreased vision when no natural light presents at nighttime and have difficulties to identify the speed and distance of oncoming vehicles. Thus, the issue of obstacle detections in night condition is still an important area of focus.

To obtain greater field of view at night, some drivers would switch the headlight from low beam to high beam. But high beam will dazzle other drivers if it is not switched back correctly. This situation might cause some accidents. Therefore, ADAS systems also enable adaptive light distribution functions that headlights can adapt accurately not to dazzle other drivers. Adaptive Driving Beam (ADB) is a novel system that can provide drivers with the benefits of forward illumination similar to that from a high beam without the glare that beam would cause for other drivers. In 2016, a new SAE Recommended Practice, SAE J3069 was created to stipulate the test protocol and requirements of an ADB system. ADB systems typically exploit a camera that detect the front road view, if the system detect preceding vehicles or oncoming vehicles, the distribution of beam will changed to avoid glare other driver’s eyes. Consequently, vehicle detection at nighttime condition is the critical part of an ADB system. The challenges of nighttime object detection can be resulted from low luminance. Mostly, the targets of nighttime vehicle detection are headlight and taillight. Many approaches utilize monocular camera and transfer the image to binary image by a fixed threshold. In addition to typical image processing techniques, machine learning based classifiers like support vector machine are one way to implement object detection. Tracking is another important process and the common approach is Kalman filter. The advantage of using tracking is to reduce computational complexity of detection. This study exploit RCCC sensor to replace Bayer sensor and aim to obtain a better quality of images at night. Candidate region selection is based on spatiotemporal analysis and we implement three global
verifications to reduce the false detecting results. A new adaptive tracking by-detection framework based on structured output prediction [6] is applied with the detection process.

II. VEHICLE DETECTION

In the detection process, the image will be divided into non-overlapped blocks, and each block has the same size in every image. Because we consider both the spatial and temporal domain, we regard each block as a three dimension grid. We calculate all local features based on every single block and that means every block contains both the spatial and temporal information.

The width and height for a block is 8 pixels, and the dimension for the timeline is 8 frames as TABLE I and Fig. 1. shows. The input image is 720×480 pixels, so the image will be divided into 90×60 blocks.

<table>
<thead>
<tr>
<th>Dimension</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>8 (pixels)</td>
<td></td>
</tr>
<tr>
<td>Width</td>
<td>8 (pixels)</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>8 (frames)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Diagram of a single block

TABLE I. Block dimension

A. System Architecture

B. RCCC Sensor

In most vision based systems, Bayer sensors is the typical sensor type which contains a filter with three color channel: Red, Green, and Blue. The color images are usually transferred to gray level as the pre-processing. This study exploits a unique RCCC (Red/Clear) sensor. It contains three clear channel and one red channel that can filter G and B color and enhance the sensibility of red signals. The images of RCCC sensors contain less noise and have better imaging quality than Bayer sensors in the low luminance condition. This kind sensor is also applied to traffic sign recognition. In this study, to avoid imaging other non-vehicles light source, like the reflection of headlight on road, we adjust the EV value to shorten the exposure time. Figure 4 shows the obtained image here we called Red-channel Map which keeps the pixels that contains more information of red channel.

C. Local Feature Extraction

The targets of this approach are the headlight and taillight of the vehicles. The detection part is composed of local feature
extraction and global verifications. For local feature extraction, temporal spatial and spatio-temporal analysis is adopted to extract probably useful features. We exploit a high-pass filter and a low-pass filter on both domains to get the high and low frequency information. When the spatial or temporal analyses do not require pixel by pixel information, the mean and variance calculation can efficiently use for low-frequency and high-frequency analysis respectively. For example, calculating the mean of each block of a grid image is a low-frequency spatial analysis and calculating the variance of an area of local spatial features can analyze the spatial frequency of the area. Thus, we use mean and variance as the low-pass filter and high-pass filter respectively.

1) Temporal Spatial Analysis

In temporal spatial analysis, we use temporal high-pass and low-pass process on image sequences, and following by spatial high-pass and low-pass. We calculate the mean image and variance image on time domain. For the current image, we will calculate the Mean and Variance of the previous $p$ images to generate temporal mean image and temporal variance image. Temporal mean image keep the low frequency information of image sequences, like static background, and smooth the obvious moving objects, like cars and pedestrians. As shows in Figure 6(a), the background is retained in this image because it is almost static.

\[
\text{Mean}(x) = \frac{1}{N} \sum_{i=1}^{N} x_i \quad (1)
\]

\[
\text{Variance}(x) = \frac{1}{N} \sum_{i=1}^{N} [x_i - \text{Mean}(x)]^2 \quad (2)
\]

![Figure 5. Current image](image)

(a) Temporal mean image  (b) Temporal variance image

Conversely, temporal variance image keep the obvious moving objects and filter the static. As shows in Figure 6(b), only vehicles are retained in temporal variance image. This image shows a foreground-like result and we will mention it at next section. At every moment we will calculate these two temporal analyzed image pixel by pixel, and spatial high-pass and low-pass filter will be used on both images. Finally, we will get four local features based on every single block. If there is an ordinary moving object going through the scene then there will be a sudden signal change because of the transition from the background to the foreground object.

2) Spatio-Temporal Analysis

In spatio-temporal analysis, spatial high-pass and low-pass are used on a single image, and following by temporal analysis. The analyzed order is reverse from last section. The input is the current image, and we calculate the spatial mean and spatial variance of each block. Thus, we get spatial low frequency and high frequency of the image and the following processing are temporal low-pass and high-pass.

For the temporal analysis, each block in current image has its previous $p-1$ value on time domain and we calculate the temporal mean and variance from these eight values, then we will get another four local features.

3) Candidate Selection

Temporal variance image shows a foreground-like result. Only moving objects would be retained in the foreground image so the candidate blocks selection will base on temporal variance image. We accumulate gray levels within each block of temporal variance image.

If the block $B_k$ exceeds our predefined threshold $T_1$ and the red channel value at the same location on Red-channel map satisfies threshold $T_2$ simultaneously, the block will be regarded as a candidate block. The sum of foreground image for each pixel as shown in Equation (3) where $S_k$ is the $k$th block and $x$, $y$ is the coordinates of the image. $T_1$ is the predefined threshold. Only selected candidate blocks will be processed in global verifications.

\[
S_k = \begin{cases} 
1, & \text{if } \sum_{x,y} \left[ \text{foreground}(x,y,t) \right] > T_1 \\
0, & \text{otherwise}
\end{cases} \quad (3)
\]

![Figure 6](image)
the parameters of the tracker are not estimated correctly, non-target objects might be tracked, resulting in a significant increase in false positives. Thus, how to accurately predict the direction of the target will be the major issues to overcome.

This study uses Struck tracking algorithm [6]. It is not only using object’s historical trajectory to predict, while using Haar-like features and support vector machine (SVM) classifier to verify the detected objects. It is mentionable that the SVM in struck tracking is a structural support vector machine and its output is not a single label but a discriminant function $Y_r$. It is more flexible to classify different types of data, such as images or a number series which can usually be expressed as a matrix. Equation (4) shows he discriminant function $Y_r$ of struct SVM.

$$y_r = \arg \max_{d \in \text{decay}} F(x_p^{r-1}, d)$$ (4)

Struck tracking also has the ability to update training samples instantly. When the system enable the tracking algorithm, everytime a new target is detected, the new target information is returned to the background to be re-trained online to get the most stable tracking result.

IV. EXPERIMENTAL RESULTS

To demonstrate the performance of this study, we collected two scenarios including oncoming car and preceding car. The target cars were at the distance from 30 meters to 100 meters in front of the host car which the RCCC camera sensor were mounted on the windshield at nighttime. The resolution of the Images is 720x480. The two scenarios contained 1680 frames and 1763 frames respectively. We divided the distance to three ranges: 30–50(m), 50–80(m), and 80–100(m). The results are expressed in frame based detection rate and false detection rate.

<table>
<thead>
<tr>
<th>TABLE II. EXPERIMENTAL RESULTS OF PRECEDING CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (m)</td>
</tr>
<tr>
<td>Number of frames</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>30–50</td>
</tr>
<tr>
<td>50–80</td>
</tr>
<tr>
<td>80–100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III. EXPERIMENTAL RESULTS OF ONCOMING CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (m)</td>
</tr>
<tr>
<td>Number of frames</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>30–50</td>
</tr>
<tr>
<td>50–80</td>
</tr>
<tr>
<td>80–100</td>
</tr>
</tbody>
</table>
V. CONCLUSION

In this study, we present a nighttime vehicle detection method to detect and track the headlight and taillight of vehicle. The image sensor we used is RCCC sensor. This kind of sensor enhances the sensibility of red light and decrease more noise signals than typical Bayer sensor under low illumination conditions. The red-channel map is generated to decide the location of red light source in one image. In the vehicle detection process, the spatiotemporal analysis is utilized to extract local features and select the candidate region of the image. Foreground objects can be segmented from background without any background models. Following three global verifications are utilized to eliminate the false detected objects like the reflection of headlight of the road surface. Struck tracking is applied to the detected results to predict and decrease the range of detection for candidate objects, thereby reducing the computational complexity of detection. Finally, we collected two testing scenarios including the oncoming car and preceding car at 30~100 meters and the experimental results have shown high accuracy of the proposed method.

REFERENCES


