Development of the system which recognize of mark of “STOP” and judge stop status by in-vehicle camera

Ryota Kaneto\textsuperscript{a,*}, Lifeng Zhang\textsuperscript{a}

\textsuperscript{a}Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-8550, JAPAN

\textsuperscript{*}kaneto.ryota397@mail.kyutech.jp

Abstract

In recent years, in order to prevent traffic accidents caused by automobiles, automobiles equipped with a driving support system have become widespread. Although traffic accidents themselves are relatively decreasing, the absolute number is still significant. Driving assistance systems are attracting attention as ones that enhance driving safety. In the driving support system, there are many functions to enhance driving safety, such as a false start suppression system, a lane departure suppression system, and a tracking function. These functions are loaded on relatively new cars, but when they are a little older, they may only be partially loaded or may not have a driving assistance system at all. There is a need for a driving support system that can be installed later and easily incorporated into those vehicles. In this research, paying attention to the sign recognition function among many driving support systems, the camera is used to recognize the sign of “STOP”, extract the feature points from the input image, and analyze the movement vector of the point indicate a system that determines the status of compliance with the “STOP” sign, warns, and prompts for a stop.


1. Introduction

1.1 Backgrounds

In recent years, the number of accidents caused by automobiles has reduced, but it is not at the very least. Possible causes of the accident include ignoring and missing signs and signals, faulty driving operation, and insufficient confirmation of the presence of other vehicles and pedestrians. The automotive industry is working on various things to get rid of such causes. For example, driver assistance systems such as a false start prevention system and a collision prevention system, which have been installed widely in recent automobiles.

Also, looking at the number of traffic accidents by age, the younger generation in their teens and 20s and the elderly aged 80 and older are more likely to have a traffic accident, and the proportion of teenagers, in particular, tends to be higher. As described above, the driving support system is installed only on relatively new models, and it is difficult for young people and older people with high traffic accident rates to purchase the latest models from an economic point of view. In this research, a warning system for oversight the traffic sign with a mobile-phone camera is proposed for such kind of people. The purpose of this study was to develop a system to recognize “STOP” signs that can be easily installed later and to determine the stop status.

2. Principle

2.1 AdaBoost

In AdaBoost learning, given an example of a pattern belonging to multiple classes to be identified, the class to which you belong can obtain a function for identifying an unknown input pattern. AdaBoost creates different classifiers while changing the weight of learning samples sequentially, performs weighted majority voting on these classifiers, and gives the final classification function. The Fig. 1 shows the

Fig. 1. Discriminator obtained by AdaBoost
classifier obtained by AdaBoost. For an input pattern \( x \), output the label \( y \) of the class to which the pattern belongs. A strong classifier is configured by combining \( T \) weak classifiers \( h(t) \) with a reliability \( \alpha_t \). When learning, first select \( h_1(x) \), and the next \( h_2(x) \) is chosen sequentially to identify the samples for which \( h_1(x) \) could not be identified.

**Learning algorithm**

1. Preparation of learning sample
   Prepare \( N \) learning samples \( (x_1, y_1), \ldots, (x_N, y_N) \). \( x_i \) is a sample, and means an image pattern of the target to be detected and another target to be distinguished from it. \( y_i \) is the label of the class to which the sample \( x_i \) belongs, and target of detection is generally labeled as \( y_i = +1 \), otherwise \( y_i = -1 \).

2. Initialize weight of each sample
   Initialize the weight of each sample with \( D_1(i) = \frac{1}{N} \).

3. Iterative learning
   
   (i) Train weak classifier \( h_t(x) \) in sample distribution \( D_t \)
   Choose \( h_t(x) \) to minimize error rate for learning samples
   \[
   \epsilon_t = \sum_{i : y_i \neq h_t(x_i)} D_t(i) \tag{1}
   \]
   The error rate is calculated based on the weight, not the number of samples.

   (ii) Calculation of reliability \( \alpha_t \) from error rate
   Reliability calculated from error rate
   \[
   \alpha_t = \frac{1}{2} \log \left( \frac{1 - \epsilon_t}{\epsilon_t} \right) \tag{2}
   \]
   The smaller the error rate, the higher the reliability.

   (iii) Update sample weights
   
   \[
   D_{t+1}(i) = D_t(i) \exp[-\alpha_t y_i h_t(x_i)] \tag{3}
   \]
   About the sample that was correctly identified by the weak classifier, since \( y_i h_t(x_i) = 1 \), it becomes \( D_{t+1}(i) = D_t(i) \exp(-\alpha_t) \). That is, the weight decreases. Since \( y_i h_t(x_i) = -1 \) for the wrong sample, \( D_{t+1}(i) = D_t(i) \exp(\alpha_t) \). In other words, the weight increases. If \( \epsilon_t = 0.5 \), then \( \alpha_t = 0 \), and the weight is not updated, so there is no point in continuing learning.

   (iv) Normalize so that the sum of sample weights is 1
   
   \[
   D_{t+1}(i) = \frac{D_{t+1}(i)}{Z_{t+1}} \tag{4}
   \]
   here,
   \[
   Z_{t+1} = \sum_{i=1}^{N} D_{t+1}(i) = \sum_{i=1}^{N} D_t(i) \exp[-\alpha_t y_i h_t(x_i)] \tag{5}
   \]

4. Completion of final classifier
   The final classifier performs weighted majority voting with confidence on all weak classifiers, and is expressed as follows.
   
   \[
   H(x) = \text{sign} \left[ \sum_{t=1}^{T} \alpha_t h_t(x) \right] \tag{6}
   \]

2.2 OpticalFlow

**Lucas-Kanade method**

In this experiment, optical flow is performed using the Lucas-Kanade method. Since this method uses only local information derived from a small window surrounding each of
Table 1. Overview of verification video

<table>
<thead>
<tr>
<th>Video</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Daytime, Pedestrian presence</td>
</tr>
<tr>
<td>B</td>
<td>Daytime, Downhill, Two consecutive signs, Pedestrian presence</td>
</tr>
<tr>
<td>C</td>
<td>Daytime</td>
</tr>
<tr>
<td>D</td>
<td>Daytime, The sign is reflected diagonally and is slightly hidden in the tree</td>
</tr>
<tr>
<td>E</td>
<td>Daytime, The crossing of the track is visible</td>
</tr>
<tr>
<td>A’</td>
<td>Nighttime, Same place as A</td>
</tr>
<tr>
<td>B’</td>
<td>Nighttime, Same place as B</td>
</tr>
<tr>
<td>C’</td>
<td>Nighttime, Same place as C</td>
</tr>
<tr>
<td>D’</td>
<td>Nighttime, Same place as D</td>
</tr>
<tr>
<td>E’</td>
<td>Nighttime, Same place as E</td>
</tr>
</tbody>
</table>

3. Proposed system

The processing flow of the proposed system is shown in Fig. 2.

First, the “STOP” sign is detected from the input image sequence first time, the “STOP” sign tracking started. When there was a false detection happened during the tracking period, in fact, such irregularly false detection was confirmed frequently. However, such case of false detection is considered that can not be captured continuously, and it is checked whether detection is continuous during a fixed period of detection. After determining that it is not a false positive, warn of the presence of the sign. When the car is running close to sign, a warning for compliance is given. Moreover, when the car approaches further, it is judged whether it has stopped.

A machine learning algorithm called AdaBoost is used to detect “STOP” signs.

Also, in the determination of the stop state, a method called “OpticalFlow” is used. The feature point is extracted from the image. And the motion is analyzed from the movement vector of the feature point. In order to stabilize the judgment, the stop state is achieved by suppressing the extraction of feature points from the upper part and the lower part of the moving image, where the same color spreads such as roads and sky, and the feature points are stable. The central part is not used for moving detection because the movement is small and have more extra noise from other moving objects.

4. Simulations

In this research, both the “STOP” sign detection and the judgment of stop state of the car are aimed to be examined only from the video sequence. The experiment is designed as following.

4.1 Outline of experiment

First, examine whether “STOP” sign can be correctly recognized from image sequence from a video camera. In addition, it is evaluated whether the judgment of the stop state can be performed correctly. The video used in this experiment is described as A to E and A’ to E’ shown in Table 1, and the features of each video are explained below.

4.2 Evaluation method

Aside from the correct image used for learning, an image to be used for evaluation is prepared, a detection program.
Table 2. Inspection result

<table>
<thead>
<tr>
<th>Video</th>
<th>Number of frames recognized</th>
<th>Number of misrecognized frames</th>
<th>Stop judgment</th>
<th>Number of feature points used for stop judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>91</td>
<td>18</td>
<td>○</td>
<td>75</td>
</tr>
<tr>
<td>B</td>
<td>(1) 98; (2) 76</td>
<td>7</td>
<td>(1) ○; (2) ○</td>
<td>(1) 59; (2) 61</td>
</tr>
<tr>
<td>C</td>
<td>54</td>
<td>31</td>
<td>○</td>
<td>44</td>
</tr>
<tr>
<td>D</td>
<td>119</td>
<td>4</td>
<td>○</td>
<td>36</td>
</tr>
<tr>
<td>E</td>
<td>98</td>
<td>7</td>
<td>○</td>
<td>54</td>
</tr>
<tr>
<td>A’</td>
<td>33</td>
<td>3</td>
<td>○</td>
<td>36</td>
</tr>
<tr>
<td>B’</td>
<td>(1) 10; (2) 66</td>
<td>1</td>
<td>(1) ○; (2) ○</td>
<td>(1) 無し; (2) 10</td>
</tr>
<tr>
<td>C’</td>
<td>43</td>
<td>2</td>
<td>○</td>
<td>10</td>
</tr>
<tr>
<td>D’</td>
<td>27</td>
<td>6</td>
<td>○</td>
<td>32</td>
</tr>
<tr>
<td>E’</td>
<td>54</td>
<td>0</td>
<td>○</td>
<td>8</td>
</tr>
</tbody>
</table>

is executed, and classification is performed according to the detection result.

As shown in Fig. 3, the “STOP” sign was correctly recognized from the captured image. In Fig. 4, the “STOP” sign was correctly recognized, but another object was also detected as a “STOP” sign which is the “false recognition”. And in Fig. 5, only something other than the “STOP” sign was detected as “false recognition.”

5. Results

The number of recognized frames obtained by executing the verification program, the number of misrecognized frames, and the evaluation of stop judgment are shown in Table 2.

There are two consecutive “STOP” signs appeared in video B and B’, the one near to the camera was called sign(1) and the other one called sign(2) which can be found in Table 2. Comparing the five daytime videos, the number of frames that can be recognized by the movie C is less, and the number of misrecognized frames is relatively more than other videos. Besides, because video C continued to be caught targets other than “STOP” sign, it moved to the procedure of stop judgment even there has not the “STOP” sign. When comparing nighttime videos, the number of frames that could be recognized showed variation, and also the number of feature points extraction in the stop judgment also showed variation. Comparing the daytime and nighttime result, the number of frames that could be recognized and the number of feature points used for stop judgment were trends to smaller at nighttime, and there were cases where stop judgment could not be made at nighttime.

6. Discussion

From the experimental results, comparing the videos of different road conditions, the number of frames that were detected the “STOP” sign showed variation. This is thought the sign itself is inclined or hidden by trees, which makes it challenging to recognize. Besides, since the learning data of the cascade classifier used in this experiment was extracted from videos taken at multiple locations, the number of data is large but also with a considerable redundancy. Furthermore, in the breakdown of learning data, the daytime samples are more abundant than that of the nighttime. Also, when it comes to weather, there are many data when it is sunny or cloudy, and data under other weather conditions such as rain and snow are not used. Therefore, the detection performance was thought depending on the location due to the bias of the learning data. Although the number of erroneously recognized frames also showed variation, the cause may be that a small number of incorrect images in the training data.

Looking at the result of the stop judgment, several scenes where it was not possible to judge well, and the number of feature points used for the stop judgment decreases notably at night. In this experiment, after recognizing the “STOP” sign, feature points are extracted from the whole screen when approaching a certain distance, and the stop is judged from the movement of the feature points. When a fast-moving object such as a car or a bicycle crosses the front, it disappears from the screen during judgment, so there is no problem in stopping the judgment. However, it is difficult to catch the same position from the wall that is uniform in color, and the object that is seen through the judgment of stopping pedestrians, etc. Therefore, the judgment is made when the fixed number of feature points is less than the defined movement distance. At night, the number of feature points that can be obtained from a dark screen as a whole decrease, and it is
considered that the stop judgment could not be made due to the factors described above.

7. Conclusions

In the present study, it was possible to recognize the sign and judge the stop state in the bright time zone where it is easy to recognize the “STOP” sign. However, in conditions where it is difficult to see signs such as nighttime, there have been cases where accurate operation cannot be performed. As future issues, it is considered that it is necessary to review learning data in order to improve the recognition rate, and to preprocess the input image so that features can be extracted from the nighttime image.

References