A Vehicle Image Tracking Method Robust to Size Reduction and Partial Color Change

Soutaro KANEKO, Osafumi NAKAYAMA
FUJITSU LABORATORIES LTD.
4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki-shi, Kanagawa 211-8588 JAPAN
{kaneko.soutaro, osafumi}@jp.fujitsu.com

***Abstract*—In recent years, vehicles have a driving support device which recognizes a forward vehicle for auto cruise or collision avoidance. Since these devices are related to human safety, they are required sufficient evaluation in their development process, but evaluation data preparing cost is too high because it needs to prepare manually on human recognition. Therefore, we propose an image tracking method useful to reduce human's annotation work. This method has robustness to size reduction and partial color change, which are apparent changes often occurring in a forward vehicle. Experiments showed that this method enables to generate accurate position annotations of many frames from the position annotation of 1 frame.**

Keywords—Forward vehicle image tracking; Template Matching; Video Processing

I. INTRODUCTION

In recent years, vehicles have a driving support device for auto cruise or collision avoidance, which is contributing to safe driving. These devices alert the driver or control the behavior of the vehicle according to the recognition result of another forward vehicle that becomes an obstacle in the traveling direction of the vehicle [1]. Since these devices are related to human safety, they are required sufficient evaluation assuming various scenes in the development process. This evaluation is performed by comparing the detection result of the driving support device with the correct detection result. This correct detection result information indicates the image region of the target and is called an annotation. Normally, as typified by the well-known dataset ImageNet [2], annotation is done manually based on human recognition. More recently, annotation data is often used not only for evaluation but also for improving recognition accuracy of a device using machine learning. That is, in the process of developing a safety support system for vehicles, since a huge amount of manual work such as annotation has occurred, cost reduction for preparing this data has become a big problem.

In order to solve the problem about preparing annotation data, using automatic detection technology can be considered as one approach. This approach is realized by the identification technology of the forward vehicle region at the pixel-wise level which is comparable to that of human annotation. Mainly, there are two image recognition methods for specifying the region of the forward vehicle. One method is machine learning which detects the target for each image, but it has two problems that cannot be solved. First, this approach is inconsistent because these annotation data are necessary to create those classifiers by machine learning. Second, machine learning acquires recognition robustness by absorbing differences in various shapes, so it loses boundary strictness. The other method is image tracking that expresses some features of the target vehicle image and specifies the position and size in the sequential images. Since the forward vehicle has a rigid body whose shape does not change, if the target vehicle is well expressed by image features, the method can track it. However, actually, it is necessary to solve two problems about forward vehicle image tracking shown as Figure 1. The first problem is the tracking failure caused by the size change of the forward vehicle in image. Because, since the image size of the forward vehicle always changes according to the relative distance between the host vehicle and the forward vehicle, it causes pattern matching failures. The second problem is that tracking may failure for the sake of the partial color change of the forward vehicle. If a part of the tracking target has some color changes, that target pattern does not match the pattern before the color changes, so general pattern matching method fails. These partial color changes are often occurred by lighting of rear lamps of the forward vehicle.

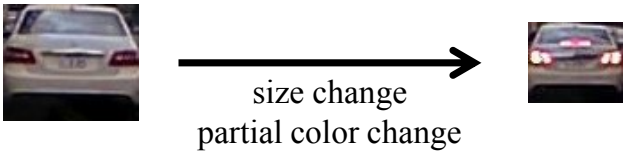


Figure 1 Change of forward vehicle image

In conventional method, SIFT [3], which is a typical pattern matching method invariant to size change, does not consider partial color change so it cannot continue accurate tracking. Meanwhile, there is a match method that using only a small number of stable color pixels selected from multiple color variations of the same object [4]. However, this also has two problems. First, it needs to have an extra method to know the color variation of the target forward vehicle before its tracking process is performed. Second, the partial color change occurrence regions that characterize an important shape for a vehicle such as rear lamps cannot be used as features at tracking. As mentioned above, conventional methods cannot track forward vehicle correctly. In addition to the problems with these conventional methods, the image

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

tracking method also has a problem of how to obtain the features of the target vehicle at the start of tracking.

In summary, image tracking can be considered as an approach of annotation method to support the development of forward vehicle recognition devices, but the method has two difficulties to solve. One is how to track targets that change their size, and the other is how to track targets that change their partial color.

II. METHOD

This proposed method solves the problems of preparing forward vehicle annotation data by image tracking with semi automatic annotation. Semi automatic means that only the position at the start of the tracking of the target forward vehicle which has the largest region in the time series images is given manually. This method solves two problems about forward vehicle tracking; one is the size change problem of tracking targets in time series images and the other is the partial color change problem of them.

First, for the problem of tracking targets with size changes, this method determines the tracking result of frames in which the size of the target is obvious from the entire time series image, and then determines the tracking result of the remaining frames using the determined result. That is, this method performs tracking in two stages. The first stage is that determines intermittent frames called key frames which enable to obtain accurate target's position. The second stage is that tracks remaining frames under the constraint of the accurate tracking results detected in the first stage. In order to obtain intermittently accurate position, this method generates a group of templates that can specify obvious size change. In addition, since this method is given the largest size of the target forward vehicle manually at the start of tracking, it can limit to track smaller size than the size of starting.

Second, for the problem of tracking targets with partial color changes, this proposal two-stage tracking method plays an important role. At the first tracking, since this method performs selective multi-color-channel matching that uses only stable color channels, this method can continue to track with avoiding an adverse effect caused by a large changes in a specific color channel. At the second tracking, this method tracks the target in consideration of a variety of partial color changes based on the target's color analysis using the first tracking result.

The proposal method consists of 4 steps as shown in Figure 2. In the remainder of this section describes this proposal method in the view of acquiring robustness to size change and partial color change.

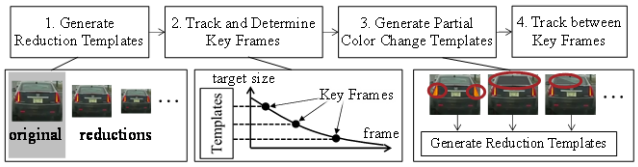


Figure 2 Overview of proposal method

1. Tracking robust to size change

In order to track the target whose sizes change, this method generates a group of templates of different sizes from the image region of the target vehicle specified manually. This template group consists of a template obtained by pixel-wise reducing the initial image region based on the width of the target vehicle. Resized templates are generated up to a predetermined minimum pixel width $w_{\min}$. At the first tracking, this method selects multiple templates with intermittent size from the templates $t_w ... t_{w\min}$ which are already generated. The subscript of t represents the width of the template. Selection of template group k is according to the following equation (1). W means the width of the template, and M means the maximum value of RGB-ZNCC.

$$\begin{aligned} &k_0 = t_w, \\ &k_i = t_r \ \ where \ \ r = \arg\min_r (W(k_{i-1}) - W(t_r)) \\ &\qquad and \ \ M(k_{i-1}, t_r) < h_m \end{aligned} \tag{1}$$

The first stage tracking uses up to 3 templates for 1 frame. When current tracking frame number is f_c, input image of f_c is I_{fc}, and the template is k_n in the last key frame number f_l, this method evaluates whether there is an obvious size as shown in equation (2). Here, f is $f_l \le f \le f_c$.

$$C(f_c - 1) = \begin{cases} -1 & if\ n-1 = \arg\max_v M(I_{fc}, k_v) \\ & and\ f_c - 1 = \arg\max_f M(I_f, k_v) \\ 1 & if\ n+1 = \arg\max_v M(I_{fc}, k_v) \\ & and\ f_c - 1 = \arg\max_f M(I_f, k_v) \\ 0 & otherwise \end{cases} \tag{2}$$

The matching result of k_{n-1} if $C(f_c - 1) = -1$ or the result of k_{n+1} if $C(f_c - 1) = 1$ at the frame $f_c - 1$ is determined as key frame result. Then, the template used for the key frame result is set as a new k_n, and tracking is continued from $f_c - 1$. It enables to determine the key frame as the best result of template matching at the time of obvious size change. Even when not determined as a key frame, tracking is continued for all three templates until the best match score falls below the end threshold. As a result, accurate matching positions are obtained over frames intermittently.

At the second tracking, this method determines the position and size of the target in each frame between key frames under constraints of both neighbor key frames which have accurate tracking results. Since the target vehicle certainly exists in a frame between the key frames, the tracking result is determined with the position and size of the most matched template between the candidates. After the last key frame, this method only uses the constraint of the size, and tracks until it falls below the match threshold.

2. Tracking robust to partial color change

At the first tracking, in this matching function M in equation (2), in order to reduce errors caused unknown partial color change, ZNCC scores are calculated using only channels with sufficient matching peak out of R, G, B, horizontal and vertical edge images.

Before the second tracking, this method generates a new template group reflecting partial color changes based on the result of the key frames as shown in Figure 3. First, this method performs pixel-wise color clustering with the result of the key frame regions and the initial template. Each result region of the key frames is normalized to the size of the initial template, and the feature vector of each pixel is represented by RGB channels of these normalized images . That is, when the normalized image size is $w \times h$, it is clustering of $w \times h$ $3n$ dimensional vectors. Since the relationship between the same pixel position at different times is set as one feature vector, this clustering result is including the co-occurrence of color changes appeared in the tracking process. These vectors are classified by k-means. Next, this method performs partial color change extraction. It compares the templates already generated and the key frame tracking results for each cluster, and it extracts the cluster region which have large color change. Finally, this method generates new templates which is reflected the partial color changes. It adds the change amount of the previously extracted color change region to one of the templates already generated which most matches the tracking result of the key frames when this color change occurs. Since the color change region is based on the normalized template size, it enables to reflect the color change occurred by the smaller size in the template of the initial size.

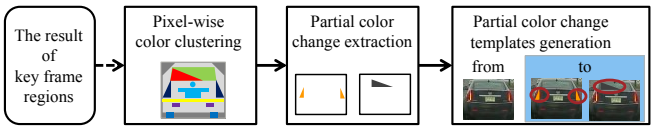


Figure 3 Generation of partial color change templates

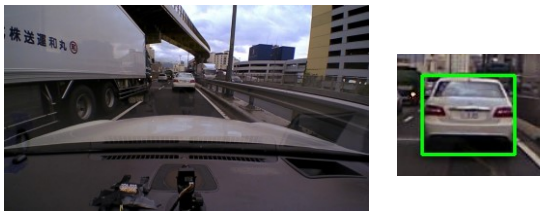
At the second tracking, the tracking is performed with these newly generated templates which are resized under constraints of the result of both neighbor key frames.

III. EXPERIMENTS AND RESULTS

This section shows two experiments and results. First, it is focusing on one tracking process to verify how this method works. Second, it is to verify the performance as an automatic annotation method from the result of tracking for various scenes. In these experiments, input images are RGB images with a resolution of 1280 x 720 and a frame rate of 30 fps on Japanese roads.

1. Experiment focusing on tracking process

Tracking was started from the inside of the rectangle shown in Figure 4 (b) of the input image Figure 4 (a) where the target vehicle appears the maximum size. Its size is 76 x 65. Key frames detection result is shown in Figure 5 partly. 28 key frames were detected from the entire 923 frames. This result can be confirmed that this method detects the vehicle correctly even if partial color change is occurred.



(a) (b)

Figure 4 Tracking start frame
(a) input image (b) initial template

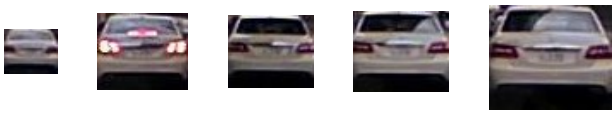


Figure 5 Example of key frames' images

The result of clustering with color in the template including the co-occurrent color change shown in Figure 6. The number of clusters is 10. For example, the rear glass is classified into different clusters because of its different color changes. Figure 6 (b)-(d) shows some examples of new templates generated reflecting partial color change. It is confirmed that new templates reflect partial color change even if these change occur in smaller size.

Figure 6 Result of color analysis
(a) color clusters including co-occurrence changes
(b)-(d) newly generated templates

The result of overall tracking is shown in Figure 7. It plots the comparison of vehicle pixel width. Actually, position errors are within 2 pixels, size errors are within 4 pixels, this is equivalent to the accuracy of annotation by hand. In Addition, the conventional RGB-ZNCC matching is failure at #537.

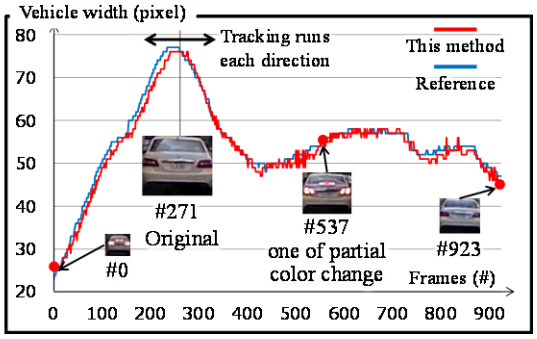


Figure 7 Vehicle tracking result

2. Experiment in various scenes

This experiment compared the annotation by proposal method and by hand in various scenes. In this experiment, up to 20 vehicle width pixel sizes are targeted for annotation.

Two indicators were used for this evaluation, one is the tracking rate R_t. This is based on how much matching was continued, R_t is expressed by the following equation (3) from the amount of tracking frames F_t and the amount of correct reference frames F_r.

$$R_t = \frac{F_t}{F_r} \tag{3}$$

The other indicator is the accurate rate R_a that determines how many frames of F_t are sufficiently close to the person's annotation is defined by equation (4). F_a is the amount of frames within the target accuracy in F_t. The target accuracy was determined from variation of human annotations, it is within 4 pixels when the correct width is 80 pixels or less, within 5% of vehicle width when the correct width is 80 pixels or more, it needs all x, y, w and h are within the range.

$$R_a = \frac{F_a}{F_t} \tag{4}$$

The results are shown in Table 1. Since there are multiple target vehicles in one video scene, 63 cases were tracked from 26 scenes. As a result, the tracking rate was 81.1% (25,314 frames) for all annotations of 31,200 frames. In addition, the target accuracy rate was 81.9% (20,741 frames).

Table 1 Experiment result of various scenes (overall)

input	Scenes	26
	Cases	63
	Frames	31,200
result	tracking rate (frames)	81.1% (25,314)
	accurate rate (frames)	81.9% (20,741)

For detailed analysis, Table 2 shows the results classified by position at the start of tracking of Table 1. The distance classification at the lower end position of the image is 5 meters ahead and the vehicle width is 80 pixels approximately. Besides, it is classified by lane position of the target.

The best result was the far-self-lane. This is because there are few shape changes other than simple size reductions and partial color changes. The main reason for tracking failure is extreme decrease in contrast or direction change as shown in Figure 8 (a). In the far-side-lane, tracking failures early appeared when the target vehicles have different depths on their back as shown in Figure 8 (b). In the near-self-lane, due to the change in the aspect ratio, there were several cases where the target accuracy range was out of the lower end position of the vehicle shown in Figure 8 (c), however, since the actual distance errors caused by them were less than 20 cm, the influence of their errors is small. The near-side-lane had the lowest accuracy. This is because the appearance of the back side changes greatly, such as the invisible rear lamp becomes visible as shown in Figure 8 (d). However, as a whole, it was confirmed that many annotations can be automatically performed. Additionally, we confirmed that the entire work can be saved by conducting tracking again from the changed appearance.

Table 2 Experiment result of various scenes (classified by target's start position)

input	distance	far		near	
	lane	self	side	self	side
	cases	15	31	5	12
	frames	12,606	11,571	1,788	3,220
result	tracking rate	96.3%	70.6%	98.8%	61.8%
	accuracy rate	90.9%	83.4%	51.3%	61.5%

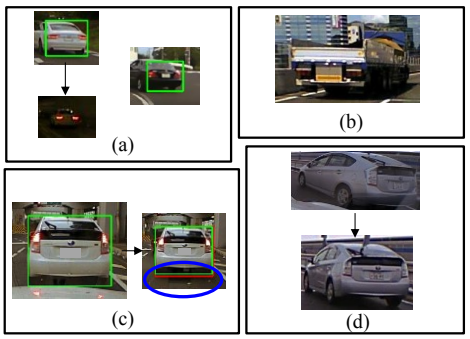


Figure 8 Examples of tracking failure

IV. CONCLUSION

In this paper, we proposed a template matching method that has robustness to size reduction and partial color change, which can be utilized for annotating positions of forward vehicles. Experimental results suggested that it enables to generate accurate position annotations of many frames from the position annotation of 1 frame, and it needs to have robustness to change of direction as a further study.

REFERENCES

[1] Choi, H-C., et al. "Vision-based fusion of robust lane tracking and forward vehicle detection in a real driving environment." International Journal of Automotive Technology 13.4 (2012): 653-669.

[2] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[3] Ng, Pauline C., and Steven Henikoff. "SIFT: Predicting amino acid changes that affect protein function." Nucleic acids research 31.13 (2003): 3812-3814.

[4] Saito, M., and M. Hashimoto. "A Fast and Robust Image Matching for Illumination Variation using Stable Pixel Template based on Co-occurrence Analysis." IEEJ TRANSACTIONS ON ELECTRONICS INFORMATION AND SYSTEMS C 133.5 (2013): 1010-1016.