Vehicle Tracking System Robust to Changes in Environmental Conditions

Yasuo OGIUCHI*, Masakatsu HIGASHIKUBO, Kenji NISHIDA and Takio KURITA

Driving Safety Support Systems (DSSS) provide drivers with visual information on their surrounding traffic environment to alert them of possible driving-related dangers, preventing traffic accidents. The image processing sensors installed on the road for the systems need high reliability in tracking vehicles regardless of the environmental conditions. The authors have developed a tracking algorithm with high accuracy and stability even in adverse lighting or weather conditions. This paper outlines the developed algorithm along with the results of public-road testing.

Keywords: DSSS, image sensor, tracking, pixel-pair feature

1. Introduction

In cooperative Driving Safety Support Systems (DSSS) aiming at the reduction of traffic fatalities and serious injuries, it is essential to track a detected vehicle precisely and stably for estimation of the driving speed and collision alarm on it.

We developed a method that can track the target vehicle precisely without being affected by environmental conditions. In this paper, we report the outline of the developed tracking method and the result of the vehicle tracking experiment.

2. Background and Purpose of the Development

Traffic fatalities have been decreasing for 9 years, and became 4,914 in 2009[1]. This largely depends on improvements in medical technology and the spread of various kinds of in-vehicle safety equipment. Furthermore, the number of the injured has decreased to less than 1,000,000. The numeric target “Less than 5,500 traffic fatalities and less than 1,000,000 fatalities or injured people in total in 2010” has been achieved two years earlier than planned. Given this situation, National Public Safety Commission is working on traffic safety to achieve the target of “Reducing traffic fatalities to half, or less than 2,500, in 2018 and to become the most safety country in the world.”

Analysis of the accident situation in detail shows that the reason of more than 60 percent of the accidents is rear-end collision, head-on collision, or turning right collision. The law violation in more than three quarters of the accidents is the violation of safe driving practice (such as lack of safety check, looking aside when driving, or lack of confirmation of traffic movement).

In order to achieve the target of National Public Safety Commission, “Less than 2,500 fatalities in 2018,” it is necessary to realize the cooperative DSSS, in which infrastructure equipment and in-vehicle equipment cooperate to prevent an accident that cannot be prevented with each equipment only.

In order to realize the cooperative DSSS, it is essential to develop various technologies such as sensors, communication systems and traffic signal control. In particular, in sensing technology, image processing with monocular cameras is promising in the total balance of the measuring range, the product life, cost, and performance.

We have been developing vehicle detection systems which can detect the position of vehicle and motorcycles with high accuracy under various environmental conditions[2]. Cooperation of such a vehicle detection system with a high performance vehicle tracking system can realize more advanced processing such as accurate measurement of movement and speed of each vehicle, correction of vehicle detection failure (miss of detection or false alarm).

However, the previous tracking algorithms show poor performance when environmental conditions, such as illumination, change largely or when the tracking target is occluded by other vehicle, and they cannot satisfy the demand to track a road vehicle stably under various environmental conditions. A new tracking method based on the appearance of the vehicle is needed to satisfy such demands.

In this paper, we propose a new tracking method based on discriminative pixel-pair features selected every frame[3][4]. Pixel-pair feature is defined by difference of intensity of two pixels in an image. This feature is expected to assure robustness to changes in illumination conditions. In our experiments, the proposed method showed good performance under various conditions including low-contrast vehicles.

3. Outline of the Vehicle Tracking System

We define a tracking problem as a classification problem of obtaining an image patch that contains the object in the correct position from a new image frame. Under this definition, we need feature value and a detector that can detect the image patch with the track target in its center and other patches.

For the t-th frame, the image frame V_t and a vehicle position (and scale) L_t are obtained from the $(t-1)$ th frame (the initial position is given by the vehicle detector). Our tracking system can be used to crop a positive (correct) image patch I_t using V_t and L_t; then, F false (incorrect) image patches J_1, ..., J_F surrounding L_t are cropped. Next, the features for discriminating between I_t and J_s are extracted, instead of training a classifier. Finally, a search for an image patch I_{t+1}, which is the most similar to the positive (correct) image patch I_t, is carried out from the next frame V_{t+1}.

3-1 Pixel-pair feature

We adopt a pixel-pair feature, which is determined by a relative difference in the intensities of two pixels, as a feature for tracking. The pixel-pair feature is an extension of the statistical reach feature (SRF)[5], in which the restriction on the distance between pixel pair is removed.

The definition of the pixel-pair feature and the similarity index $c\,(I, J)$ of a given pair of images I and J of the same size are described as follows. Suppose the size of the input images is $W\times H$. Let grid Γ represents a set of pixel coordinates in the images I and J.

$$\Gamma := \{\ (i, j) \mid i=1,\ \cdots,\ W,\ j=1,\ \cdots,\ H\}$$

We regard the image of size $W\times H$ as an intensity function defined on Γ. For an arbitrary pair (p, q) of grid points in Γ, we define the value $ppf\,(p>q, T_p)$ as follows:

$$ppf(p>q\ ;\ T_p) := \begin{cases} 1 & I(p)-I(q)\geq T_p \\ -1 & I(p)-I(q)\leq -T_p \\ \phi & otherwise \end{cases}$$

Where T_p (>0) is the threshold of the intensity difference.

We refer a pixel pair (p, q) on the patch I as "a valid pair," when $ppf\,(p>q;T_p)\neq\Phi$. Hereafter, we write $ppf\,(p>q)$ rather than $ppf\,(p>q;T_p)$, unless there is any ambiguity.

If we do not restrict the selection of p and q from the image I, it is possible to extract a huge number of pixel-pair features. Therefore, we limit the number of pixel-pair features to N. By selecting a set of pairs (p, q) with selection policy s, we denote a random pixel-pair-feature set RP_s as follows:

$$RP_s\,(p,q,I,T_p,N) := \{(p,q) \mid ppf\,(p>q)\neq\Phi\}$$

where $\{p,q\in\Gamma\times\Gamma\}$, $p=\{p_1,\cdots\cdots,p_N\}$ and $q=\{q_1,\cdots\cdots,q_N\}$.

Next, we calculate the similarity of J to I based on pixel-pair features.

For a valid pixel-pair $(p,q)\in RP_s$ on the patch I, we define the incremental sign $b(p>q)$ on the patch J as follows:

$$b(p>q) := \begin{cases} 1 & J(p)\geq J(q) \\ -1 & otherwise \end{cases}$$

For a valid pixel-pair (p,q), a single pair similarity $r\,(p, q, J)$ on the patch J as follows:

$$r(p,\ q,\ J) = \begin{cases} 1 & ppf(p>q)=b(p>q) \\ -1 & ppf(p>q)\neq b(p>q) \end{cases}$$

We use the intensity difference threshold to select a valid pixel-pair on patch I, while we do not use the intensity difference threshold to validate the pixel-pair on patch J as shown in **Fig. 1**. This assures the robustness to appearance change.

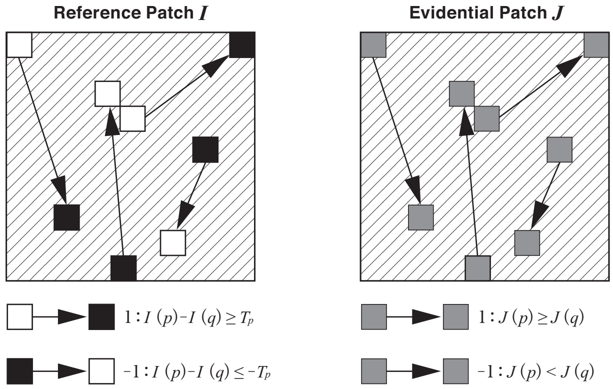

Fig. 1. Explanation of pixel-pair feature

The similarity index $c_s\,(I, J, RP_s)$ measured by using a pixel-pair feature set RP_s is defined as follows:

$$c_s(I,\ J,\ RP_s) = \frac{\sum_{(p,\ q)\in RP_s} r(p,\ q,\ J)}{|RP_s|}$$

3-2 Discriminative pixel-pair selection

Pixel-pair features are selected to maximize the discriminant criterion used for discriminating between a correct image patch I and incorrect image patches Js.

According to our problem definition, the discriminant criterion is defined under following conditions:

- ◆ the feature takes binary values $+v$, $-v$
- ◆ only ONE positive sample exists
- ◆ a large number F of false samples exist

The feature value for the positive sample p and false samples n^i are defined as

$$p=v,\ \{n^i\}_1^F=v,\ -v$$

Assuming $F\gg 1$, the total mean $\bar{\mu}_T$ is nearly equal to the mean for false samples $\bar{\mu}_n$. Defining m as the number of false samples values of which are $n^i = -v$ (i.e. not same value as the positive sample),

$$\begin{aligned} \bar{\mu}_T \approx \bar{\mu}_n &= \frac{1}{F}\sum_1^F n^i \\ &= \frac{1}{F}\left(\sum_1^m(-v)+\sum_1^{F-m}v\right) \\ &= \frac{1}{F}\,(F-2m)\,v \end{aligned}$$

The total variance σ_T^2 and inter-class variance σ_B^2 are defined as follows:

$$\begin{aligned} \sigma_T^2 &= \frac{1}{F+1}\left\{(v-\bar{\mu}_T)^2+\sum_1^F\,(n^i-\bar{\mu}_T)^2\right\} \\ &\approx \frac{1}{F+1}\left\{(v-\bar{\mu}_n)^2+\sum_1^F\,(n^i-\bar{\mu}_n)^2\right\} \end{aligned}$$

$$\sigma_B^2 = \frac{1}{F+1}(v-\bar{\mu}_T)^2 + \frac{F}{F+1}(\bar{\mu}_n - \bar{\mu}_T)^2$$
$$\approx \frac{1}{F+1}(v-\bar{\mu}_n)^2 + \frac{F}{F+1}(\bar{\mu}_n - \bar{\mu}_n)^2$$
$$= \frac{1}{F+1}(v-\bar{\mu}_T)^2$$

This equation indicates that minimizing the variance of false samples is equivalent to maximizing the discriminant criterion. The variance of false samples is redefined by substituting $\bar{\mu}_n = \frac{1}{N}(F-2m)v$ as

$$\sum_1^F (n^i - \bar{\mu}_n)^2 = \sum_1^m (-v-\bar{\mu}_n)^2 + \sum_1^{F-m} (v-\bar{\mu}_n)^2$$
$$= \frac{4v^2}{F} m(F-m)$$

The variance of false samples attains the minimum value zero at $m = 0$ and $m = F$. Since $m = 0$ implies that all the false samples have the same value as the positive sample, discrimination is impossible. At $m = F$, the discriminant criterion is maximized to one, where the similarity between a positive sample and a false sample is minimized to zero.

Therefore, minimizing the single-pair similarity index for a pixel-pair feature $c_s(I, J, RP_s)$ is equivalent to maximizing the discriminant criterion. It is also equivalent to minimizing the sum of similarity indices for a feature set C_{min} as follows:

$$C_{min} = \sum_{i=1}^{F} \{c_{min}(I, J^i, RP_{min})\}$$

where C_{min}, c_{min}, RP_{min} represent the selection policy for minimizing the single-pair similarity for each pixel-pair in the feature set.

There can be some implementations of discriminative pixel-pair features, we considered following implementation:

1. Randomly generate more pixel-pair features than needed
2. Select the number of required features from among those with similarity indices c_s.

4. Experiment

In this section, we present the performance result of our system for real conditions. In the experiment, the initial position of the tracked vehicle is defined manually, and tracker ends tracking when the car goes out of the image, or is entirely lost. In order to validate our algorithm based on the discriminative pixel-pair feature (DPF tracker), the results are compared with the results of tracking algorithm based on the least sum of the squared difference (SSD tracker).

4-1 Illumination change for a vehicle

Figure 2 shows the result in the case of illumination changes for the tracking vehicle. The appearance of the car changes as the illumination on the car changes when it moves from the shade into the sun.

The patch size for the object is 100×100 in the leftmost frame. In the SSD tracker, all the pixels in the image patch are used, and hence, 10,000 pixels are compared as features. The results for the DPF tracker and the SSD tracker are almost the same, even though a small number of features are used in the former.

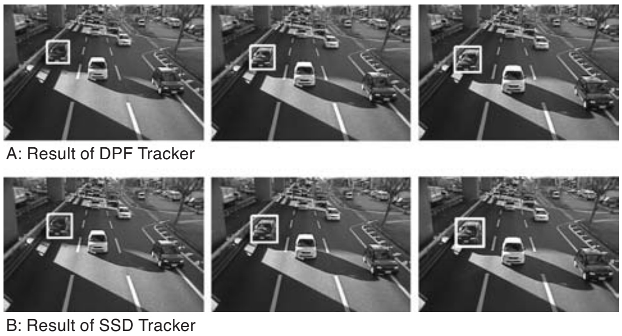

Fig. 2. Tracking result: illumination change

4-2 Partial occlusion

Figure 3 shows the result for a partially occluded vehicle. The truck on the right lane is tracked, and is partially occluded by the truck on the center lane. The precision of the tracking (especially for the scale) is not sufficient when the SSD tracker is used, while the object is successfully tracked when the DPF is used.

Fig. 3. Tracking result: partial occlusion

4-3 Nighttime

Figure 4 shows the result for the nighttime conditions. The truck that changes lanes from center to right is tracked. Error to manually defined ground-truth was evaluated.

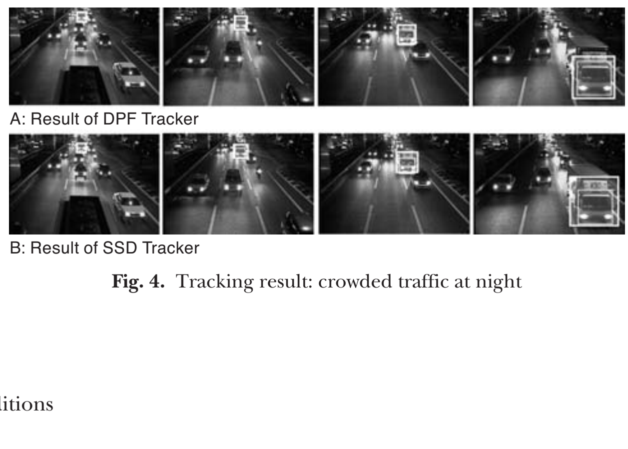

Fig. 4. Tracking result: crowded traffic at night

Since the initial patch includes some parts of the other vehicles as background, the SSD estimations are affected by the background (parts of other vehicles), and thus, the scale of estimations is larger than ground-truth. While the scale of estimation by the DPF tracker is larger than ground-truth, the estimations are more sufficient than that by the SSD tracker.

4-4 Low-contrast vehicle

Figure 5 shows the result for a low-contrast vehicle. The black car in the right lane is tracked. The appearance of the car is affected by low contrast of the car to road plane and high-contrast shadow in the background.

The DPF tracker is successfully used to track the vehicle, while the SSD tracker misses the car in an early frame.

Figure 6 shows a magnified version of the second video frame of **fig. 5**. The result indicates that the SSD tracker was affected by the high-contrast shadow in the background (road-plane), while the DPF tracker showed robustness to the background texture.

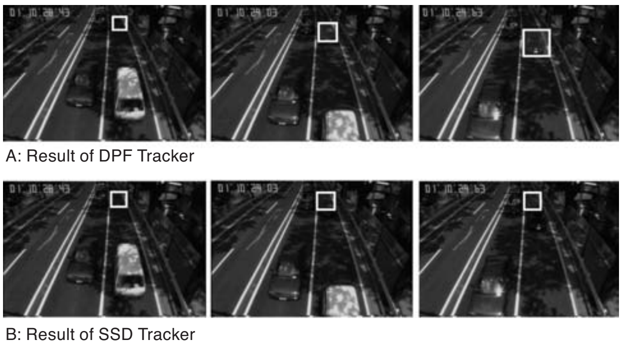

Fig. 5. Tracking result: low contrast vehicle

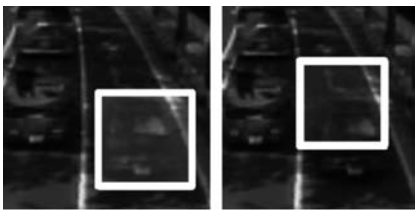

Fig. 6. Magnified figure around tracked vehicle (Left) DPF tracker (Right) SSD tracker

4-5 Reuse of pixel-pair features

In the experiments, pixel-pair feature set is fully renewed every frame to ensure the robustness to changes of the appearance of the vehicle. We refer this method as "full renewal of pixel-pair features."

However, position error was sometimes large because of the distribution pattern of pixel-pair features on the samples. We supposed that the position error could be reduced with continuous use of the pixel-pair features that meet the following conditions:

◆ The pixel-pair features are used in the previous frame.

◆ Discriminant criterions of the pixel-pair features are sufficiently high in the present frame.

We refer this method as "reuse of the pixel-pair features."

Figure 7 shows the results of the tracker with full renewal of pixel-pair features and that with reuse of pixel-pair features. The black vehicle in the second rightmost lane was tracked.

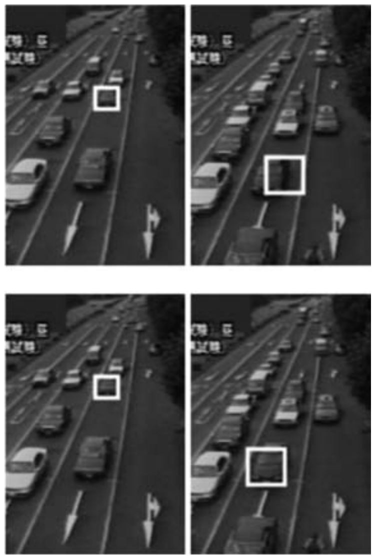

Fig. 7. Effect of the reuse of pixel-pair features (Upper) Renewal of the features every frame (Lower) Reuse of the discriminative features

When the tracker with full renewal of pixel-pair features is used, the error of position is relatively large, while the error is small with the tracker with reuse of pixel-pair features.

The pixel-pair features that are reused multiple times can be thought as a model describing the target. Thus, stability and robustness can be improved by using information on number of times of reuse of pixel-pair features.

4-6 Cooperation of vehicle detection and vehicle tracking

In the experiments in this paper, initial positions of the tracked cars are defined manually. However, the results of our other experiments show that the tracking precision was almost same even if initial position was set by vehicle detector that we had developed, and cooperation of the vehicle detection system and the vehicle tracking system is possible.

By comparing results of vehicle detection and tracking, we could correct detection failure such as miss of detection, false alarm, or error of detection positions/size. **Figure 8** shows a result of combining of vehicle detection and tracking.

The information of the correction of detection failure can be used as sample data for additive training of the vehicle detector. In addition, it is expected to use these data in automatic training of the vehicle detector with a proper training method as incremental support vector machine (SVM)[(2)].

Fig. 8. Result of cooperation of the vehicle detection and vehicle tracking Figures at the lower left show the identification number specific to each tracked vehicles or motorcycles

5. Conclusion

We have developed a tracking method that can track a target stably under adverse conditions, such as illumination change partial occlusion, or low-contrast, where previous methods are of poor performance. Additionally, cooperation of this method and the detection method we had developed can correct detection miss or false alarm of the detection system.

This method is promising for traffic accident reduction in the feature.

References

(1) National Police Agency, "The traffic accidents situation in 2009," 2010.

(2) Masakatsu Higashikubo, Epifanio Bagarinao, and Takio Kurita, "Development of Image Processing Sensor for Cooperative Driving Safety Support Systems," SEI Technical Review, No. 70, pp. 53-60, 2010.

(3) Kenji Nishida, Takio Kurita, Masakatsu Higashikubo, "Online Selection of Discriminative Pixel-Pair Feature for Tracking," in Proc. SPPRA2010, 2010.

(4) Kenji Nishida, Takio Kurita, Yasuo Ogiuchi, Masakatsu Higashikubo, "Visual Tracking Algorithm Using Pixel-Pair Feature," in Proc. ICPR2010, 2010.

(5) R. Ozaki, Y. Satoh, K. Iwata, K. Sakane, "Statistical Reach Feature Method and Its Application to Template Matching," in Proc MVA 2009, pp.174-177, 2009.

Contributors (The lead author is indicated by an asterisk (*)).

Y. OGIUCHI*

- Dr. Engineering
  Assistant Manager, Applied ICT Systems Department, Information & Communications Laboratories
  He is engaged in the research and development of image recognition systems, especially image processing sensors for road traffic measurements.

M. HIGASHIKUBO

- Manager, Applied ICT Systems Department, Information & Communications Laboratories

K. NISHIDA

- Dr. Engineering
  National Institute of Advanced Industrial Science and Technology

T. KURITA

- Dr. Engineering
  Professor, Hiroshima University