ROBUST AND PRECISE MEASUREMENT METHOD OF VEHICLES AND MOTORCYCLES FOR COOPERATIVE DRIVING SAFETY SUPPORT SYSTEM WITH COMBINATION OF HOG-SVM DETECTION AND DISCREMINATIVE PIXEL-PAIR FEATURE TRACKING

Yasuo Ogiuchi, Masakatsu Higashikubo, Yuuki Ono
Sumitomo Electric Industries Ltd.
1-1-3 Shimaya, Konohana-Ku, Osaka, 554-0024, Japan
+81-6-6466-5615, {ogiuchi-yasuo, higashikubo, ono-yuuki}@sei.co.jp

Takio Kurita
Hiroshima University
1-7-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8521, Japan
+81-82-424-6468, tkurita@hiroshima-u.ac.jp

Kenji Nishida, Hiroaki Inayoshi, Reiji Arata
National Institute of Advanced Industrial Science and Technology (AIST)
Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568, Japan
+81-29-861-5879, {kenji.nishida, h.inayoshi, arata.reizi}@aist.go.jp

ABSTRACT

In this paper, a new vehicle measurement system with image processing for cooperative driving safety support system (DSSS) is proposed. Image sensors for cooperative DSSS are required to have extremely high performance under any condition changes or environmental noise. However, there have been no known image sensor systems to meet the requirements. Therefore, we developed our own image sensor system. Inputs to the system are gray-scale images from a roadside video camera. The system consists of a detector and a tracker. The detector is based on histogram of oriented gradient feature and support vector machine trained with a newly developed efficient training method. The tracker is based on discriminative pixel-pair feature. Since both a detector and a tracker are of high-performance, they enables to correct miss of detection, error of size, position, or both, and false positives. Moreover, since the inputs to the system are gray-scale images, existing roadside cameras can be easily converted to image sensors for cooperative DSSS. The results of our experiments shows that the proposed system detects each vehicle with less than 1.0% misses and one false positive per frame, and can be processed in 30ms per frame in average. In addition, the system is robust to condition changes or environmental noise.

Keywords: Visual Tracking, Support Vector Machine, Histogram of Oriented Gradient, Incremental SVM, Discriminative Pixel-pair Feature Tracker, Cooperative Driving Safety Support System

INTRODUCTION

The numbers of traffic fatalities and the injured in Japan was 9,066 and 1,155,697 in 2000, but dropped to 4,863 and 911,108 in 2010 (1). This largely depends on improvements in medical technology and the spread of various kinds of in-vehicle safety equipment. 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 by National Police Agency 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, it is necessary to realize the cooperative DSSS (Driving Safety Support Systems), in which infrastructure equipment and in-vehicle equipment cooperate to prevent an accident that cannot be prevented with each equipment only.

Figure 1 shows an example of cooperative DSSS for a crossing. The image sensors should spot automobiles, motorcycles or pedestrians in a driver's blind corners, and measure the object's position, speed and direction of movement. Then, communications between infrastructures and vehicles by radio provide the obtained traffic data and images to drivers, or offer drivers the information, and drivers can prevent collision or other traffic accidents. As shown in this example, various technologies such as sensors, communication systems and traffic signal control are essential for the cooperative DSSS.

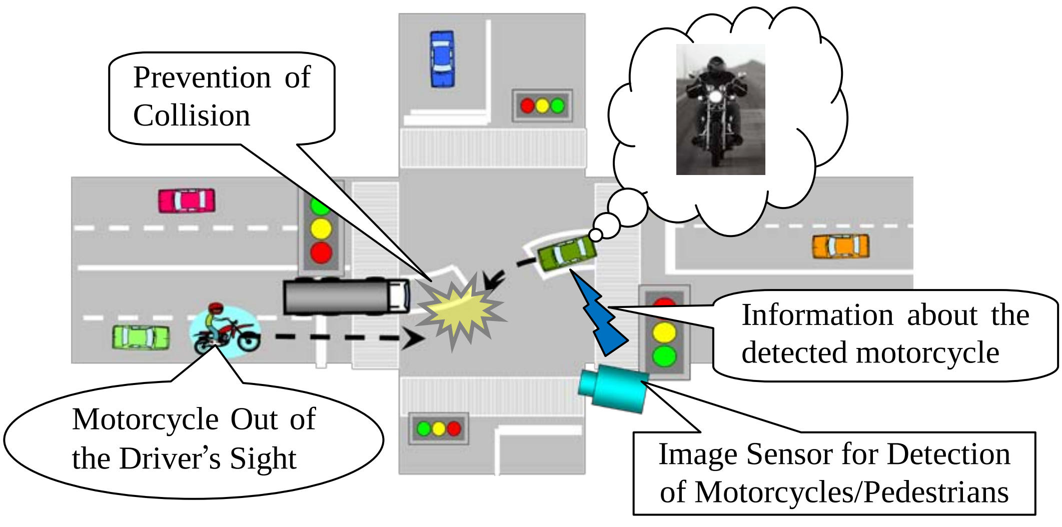


Figure 1. Example of Driving Safety Support System for a Crossing

In sensing technology, image processing with a monocular camera is promising in the total balance of the measuring range, the product life, cost, and performance. However, the image sensors for cooperative DSSS are required to have extremely higher performance compared with the conventional image sensors such as traffic flow counters.

For example, cooperative DSSS require almost 100% detection for vehicles, and the performance has to be reserved under any condition changes or environmental noise, while traffic flow counters usually allows about 10% error under some conditions. In addition, the detection of motorcycles have to be ensured in cooperative DSSS since they are often involved in serious or fatal accidents, while traffic flow counters does not attach great importance to motor cycles since they have little influence on traffic flow.

There have been no known image sensor systems to meet such requirements essential for cooperative DSSS. The previous algorithms for detection and tracking show poor performance under some environmental condition, such as illumination change, shadows or reflectance on the road surface, or occlusion of the tracking target by other vehicle.

Therefore, we developed our own image sensor system that is based on the HOG and SVM detector and the discriminative pixel-pair feature tracker. The features of the method proposed in this paper are that it is robust to condition changes or environmental noise, meets the requirements to detect each vehicle or motorcycle with high accuracy, and is a real-time processing. Cooperation of a vehicle detection system with a high performance vehicle tracking system can realize more advanced processing such as accurate measurement of movement or speed of each vehicle, or even correction of vehicle detection failure (miss of detection or false alarm). Moreover, since inputs to our system are supposed to gray-scale images, existing cameras for traffic flow counters can be easily converted to image sensors for vehicle detection.

Common specifications of image sensor systems for cooperative DSSS in Japan is now under consideration, so we set original target specification shown in table 1, which is based on the experience of our previous projects.

Table 1. Target specifications of image sensor systems for DSSS

Articles	Specification	Regulation
Measuring Range	up to 4 lanes and 150m length	
Processing time	100ms / frame	
Vehicles detection accuracy	less than 3% of misses under one false positive per frame	Rate of misses = 100 – (number of detected vehicles) / (number of ground truth) * 100

METHOD

OUTLINE OF OUR METHOD

Our purpose is to develop a vehicle measurement system with image processing. The inputs to the system are gray-scale images from a roadside video camera, and are obtained under various and varying illumination conditions, such as daytime or nighttime, conditions with a high-contrast shadow in the road plane, and with low-contrast vehicles passing through.

Figure 2 shows outline of the proposed method.

The system consists of a detector based on histogram of oriented gradient feature and support vector machine, and a tracker based on discriminative pixel-pair feature. The detector based on histogram of oriented gradient and support vector machine realized detection of more robustness to noise such as shadow or reflection, and more precision than the previous methods, and the discriminative pixel-pair feature tracker (DPF tracker) (2)(3) realized tracking robust to illumination changes, low-contrast vehicles.

Since DPF tracker attains a high precision for vehicle positions, we could correct position or size errors and false detections by comparing with the result of tracking stage.

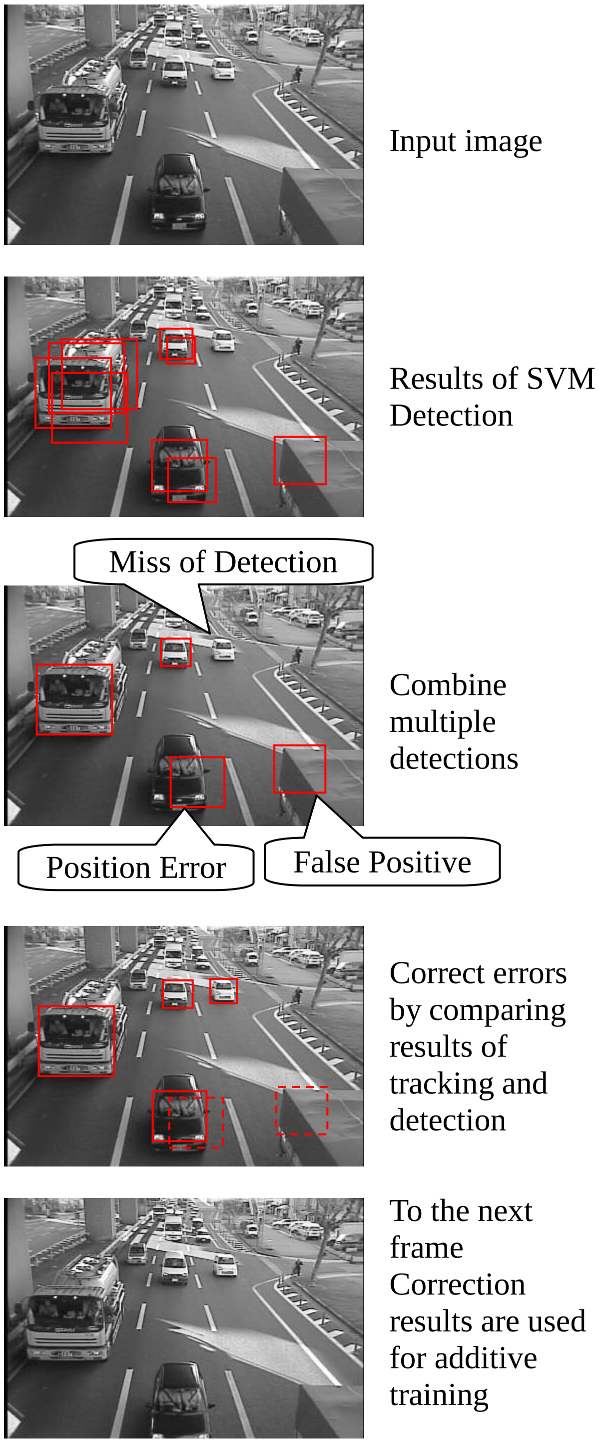


Figure 2. Outline of detection and tracking

FEATURE AND DETECTOR

Our vehicle or motorcycle detector is based on histogram of oriented gradient (HOG) feature and support vector machine.

Robustness against illumination changes and geometric change (parallel transition, rotation) can be achieved with HOG and SVM. We developed a parameter set optimum for vehicle detection, such as the size of local area, the number of bins of gradient directions.

TRAINING OF A DETECTOR

Figure 3 shows the outline of training of a detector.

To train a vehicle detector, hand labeled square training images of vehicles and paved road surface are used as positive data and negative data respectively.

First, an initial detector is trained. Procedure of training of the initial detector is as follow:

1. Select subset of the samples randomly.
2. Train a detector for the subset.
3. Check the performance for all the samples
4. Repeat 1 to 3, and pick up the best detector.

Each training time and memory size required in this method is relatively small since detectors are trained for only small subset of the samples. In addition, generalization is ensured by verifying the performance for all the samples.

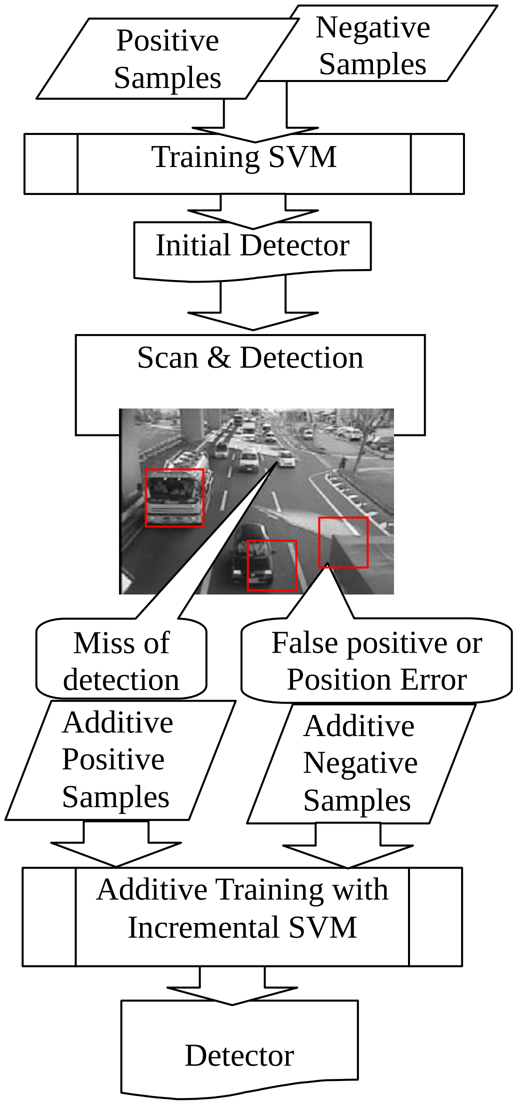


Figure 3. Training of the Detector

The initial detector scans across images, and misses of detection and false positives are collected by hand. Then they are added to the training image data set. Additive training is performed with the Incremental SVM technique (2). With the Incremental SVM technique, additive training data set can be included in the existing detector without re-training from scratch.

This training method, along with some other improvements, achieves a high-performance detector within a realistic training time. Since our method can handle a large set of samples, by using mixture of samples collected from various sequences, one common detector, i.e. detectors that does not depend on target places, can be trained with the balance between generalization and performance.

A detector for motorcycles is trained in the same way as a detector for vehicles, but hands labeled square training images of vehicles are included in negative data in addition to images of paved road surface.

DETECTION

The final detector scans across target area of the frame. We set measuring range, and change the range of detection window according to the vertical position in the frame. With these restraints, we can suppress the number of scans per frame.

Multiple detections will usually occur around each vehicle or motorcycle in a scanned image, and the same is often true of false positives. However, in cooperative DSSS, it is essential to return one final detection result per object. Therefore, overlapping detections are combined into a single detection based on the output value of the SVM for each. With some assumptions specific to vehicles and motorcycles, we can pinpoint the vehicles or motorcycles with high precision, and suppress false positives to one or less per frame.

TRACKING

Since the inputs are gray-scale images, and may contain partially occluded or low-contrast vehicles, or illumination changes because of traffic conditions on the road, tracking algorithms based on color such as mean-shift algorithm, or template matching are difficult to be applied. Therefore, a new tracking algorithm is necessary.

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.

We adopt the pixel-pair feature, which is defined by difference of intensity of two pixels in an image, as a feature for tracking (3) (4). Figure 4 shows the definition of similarity of images based on pixel-pair feature. This feature is expected to assure robustness to changes in illumination conditions, sensitiveness to position error, and low computation cost, since it based on only difference of intensity of two pixels.

Procedures of tracking based

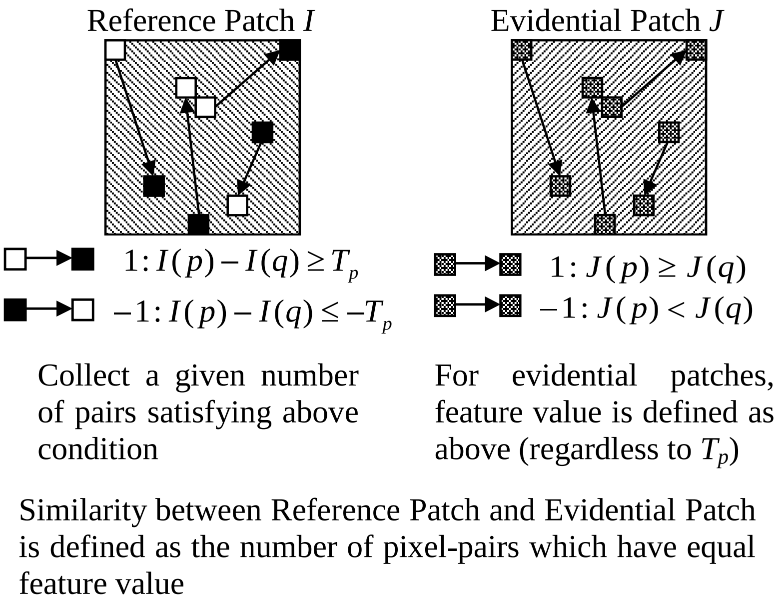


Figure 5. Similarity of images based on pixel-pair features.

on pixel-pair features are as follow:

1. At frame *T*, select pixel-pair features to minimize the similarity between the correct image patch *I* (i.e. the target of tracking) and incorrect image patches *J*s, which are near the correct image patch (figure 5).
2. At frame *T*+1, search the image patch that is the most similar to *I* on the pixel-pair features selected above. The image patch is the result of tracking.

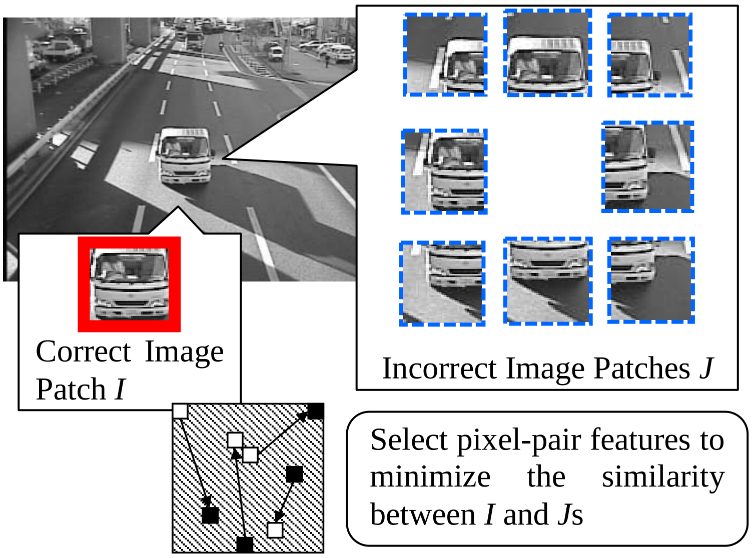


Figure 5. Selection of the discriminative pixel-pair features.

We refer such pixel-pair features as selected in above procedure as discriminative pixel-pair features. Accuracy of discrimination between correct and incorrect image patches is secured by selecting adequate discriminative pixel-pair features.

POST PROCESS

Since we use a place-independent detector, there should be place-dependent errors, such as false detection of pavement marking on the road surface. This post process handles such errors.

By comparing the results of both detection and tracking, we can correct errors such as miss of detection, false positive, and error of position and/or size. For example, if an object is once detected as a vehicle, but its movement (result of tracking) is not typical for vehicle, the object and the first detection of the object are judged as “not a vehicle” and “false positive” respectively.

The information of miss of detection or false positive can be used as samples for other additive training of the detector.

EXPERIMENT

TEST SEQUENCES

We selected 5 test sequences for the experiment, in which it was difficult for the conventional vehicle detection methods due to environmental change or noise, or tracking of the tails of the vehicles. Descriptions of the selected sequences are shown in table 2. Figure 5 shows a typical frame of each sequence.

The images in the sequences were captured in 10 frames per second from the video files taken with fixed cameras above the road. The resolution of each frame was 360x240.

Table 2. Description of the Selected Sequences

Sequence	Time	Features	Number of Frames	Max Distance (approx.)
DAY1	Daytime	3 lanes / Shadows of roadside buildings and columns of a highway above the road / Heavy traffic	12,947	80m
DAY2	Daytime	2 lanes / Shadow of roadside trees (swayed by wind) / Change of the daylight	26,204	50m
NIGHT1	Nighttime	3 lanes (Same place as DAY1) / Reflection of headlights on the road surface / Heavy traffic	11,119	80m
NIGHT2	Nighttime	2 lanes (Same place as DAY2) / Reflection of headlights on roof of vehicles	3,000	50m
TAIL	Daytime to Nighttime	2 lanes / Highway / Tracking of tails of vehicles / Medium to Heavy Traffic	9,999	100m


(a) DAY1


(b) DAY2


(c) NIGHT1


(d) NIGHT2


(e) TAIL (the 2 lanes from the left)

Figure 5. Typical frame of each sequence.

DETECTION AND TRACKING

Four detectors were trained in this experiment. Descriptions of the detectors are shown in table 3.

Table 3. Description of the detectors

	Target	Orientation	Sequences Applied
1	Vehicles	Front	DAY1, DAY2
2	Vehicles	Front	NIGHT1, NIGHT2
3	Vehicles	Tail	TAIL
4	Motorcycles	Front	DAY1, DAY2, NIGHT1, NIGHT2

Training samples of each detector were collected from the sequences where it was applied and from other sequences that were similar to them. Therefore, the detectors were robust to environmental change or noise in various sequences.

The minimum size of detection was 16x16 pixels (vehicles), and 8x16 pixels (motorcycles).

In calculation of HOG feature, we extracted 8x8-pixel cells overlapping each other by half size. Therefore, 9 cells were set in a 16x16-pixel image. We set the number of bins of the edge direction to 8. Thus, the number of the feature vector dimensions was 72.

RESULTS

Vehicles or motorcycles that were not detected properly (i.e. the detected window did not contain the center of the ground truth, or size error is larger than 50% of the size of ground truth) during passing through the image were counted as misses. Detections that did not correspond to any vehicles or motorcycles in ground truth in each frame were counted as false

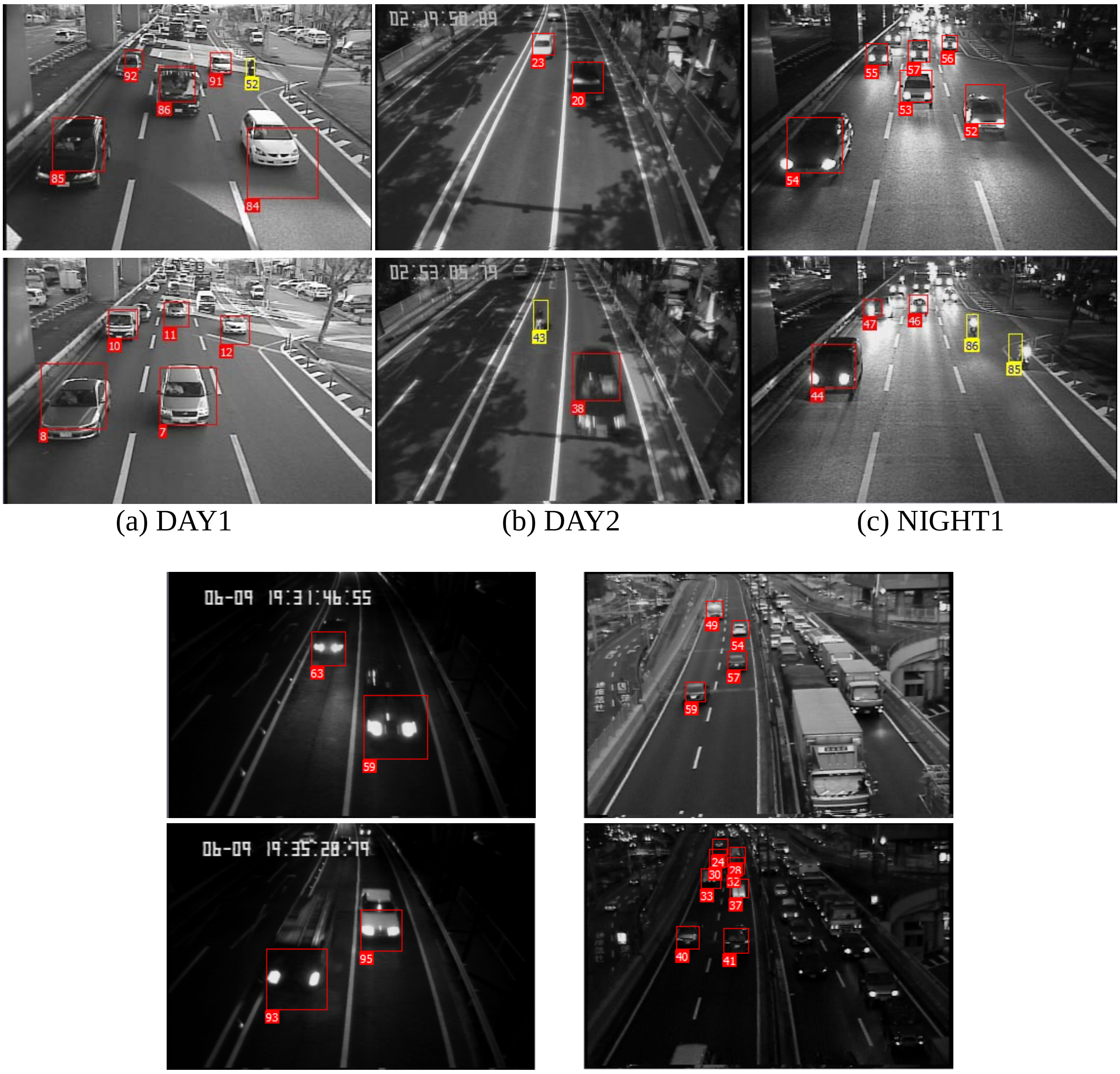
Table 4. Performance of our tracking system in the test sequences. Upper line and lower line show performance for vehicle and motorcycles respectively (motorcycles are not targets in "TAIL"). Percentages shown in the blankets are ratio of number of misses of detection to number of ground truth.

Sequence Name	Number of Frames	Number of Vehicles / Motorcycles	Number of Misses of Detection	Number of False Positives	Number of False Positives per Frame
DAY1	12947	885	2 (0.226%)	2371	0.183
		46	5 (10.9%)	1038	0.0802
DAY2	26203	989	1 (0.101%)	6898	0.263
		57	7 (12.3%)	295	0.0113
NIGHT1	11118	741	5 (0.675%)	931	0.0837
		49	5 (10.2%)	1296	0.117
NIGHT2	3000	98	2 (2.04%)	186	0.062
		11	0 (0.0%)	98	0.0327
TAIL	9999	682	5 (0.733%)	2452	0.245

detections.

Table 4 shows performance of our tracking system in the test sequences. The ratio of misses to the total vehicles was less than 1.0 %, and numbers of false positive were less than 0.3 per frame in all the sequences. Figure 6 shows examples of experimental result.

The processing time per frame was less than 30ms on the average (implemented on Windows XP / Intel Core 2 Quad Q9550 / C language) in all the sequences.



(a) DAY1 (b) DAY2 (c) NIGHT1

(d) NIGHT2 (e) TAIL

Figure 6. Examples of experimental results. Red and yellow rectangles indicate tracked vehicles and motorcycles respectively. Numbers at lower left of the rectangles are ID numbers for each tracked vehicle or motorcycle.

DISCUSSION

Ratio of misses to the total vehicles, number of false positives, and processing time per frame satisfied the target specifications shown in table 1.

The causes of the misses in the experiments were classified as follow:

- Position error (Detected and tracked part of a vehicle such as a bumper, roof, and so on)
- Size error (Detected and tracked part of a large vehicle)
- Type error (Mistook a motorcycle as a vehicle)
- Hidden behind other vehicle
- Blur of Input Images
- Lights of the target vehicle (head or tail) were off or dim

Table 5 shows causes of misses of detection in the experiment.

Table 5. Misses of detections in the experiment classified by cause. Upper line and lower line show number of miss of vehicles and motorcycles respectively.

Sequence	Position Error	Size Error	Type Error	Hidden behind other vehicle	Blur of Input Images	Headlights were off / dim	Total
DAY1	1	1	0	0	0	0	2
	0	0	3	2	0	0	5
DAY2	0	1	0	0	0	0	1
	1	0	3	0	3	0	7
NIGHT1	3	0	0	2	0	0	5
	0	0	2	3	0	0	5
NIGHT2	1	0	0	0	0	1	2
	0	0	0	0	0	0	0
TAIL	0	0	0	0	0	5	5

In practical cooperative DSSS, drivers can prevent a traffic accident if the system issue a timely warnings whatever the trigger of the warning is.

From this point of view, misses caused by the former 4 reasons are not critical. On the other hand, misses by the latter 2 reasons can be critical because there is no detection for a missed vehicle. However, improvement of imaging device can solve the problems.

Therefore, no misses are critical for cooperative DSSS. Improvement of accuracy of position and size of detections are our future work.

CONCLUSIONS

We have proposed a vehicle measurement method for the cooperative DSSS. The method mainly consists of 2 steps, detection based on HOG and SVM, and tracking based on discriminative pixel-pair feature tracker. Misses of detection and false positives can be removed by comparing the results of detection and tracking.

Secondly, we have presented a set of experiments in 5 sequences, in which it was difficult for the conventional vehicle detection methods due to environmental change or noise, or tracking of the tails of the vehicles. The results of the experiments shows that our system is of high accuracy, robust to environmental noise or changes, fast enough for real-time processing, and sufficient for cooperative DSSS.

REFERENCES

(1) National Police Agency, *The traffic accidents situation in 2010* (in Japanese), February, 2011.

(2) Epifanio Bagarinao, Takio Kurita, Masakatsu Higashikubo, and Hiroaki Inayoshi, "Adapting SVM Image Classifiers to Changes in Imaging Conditions Using Incremental SVM: An Application to Car Detection", *Proc. ACCV 2009 (The Ninth Asian Conference on Computer Vision)*, September, 2009, pp. 363-372

(3) Kenji Nishida, T. Kurita, M. Higashikubo, "Online Selection of Discriminative Pixel-Pair Feature for Tracking", *Proc. SPPRA2010 (The Seventh IASTED International Conference on Signal Processing, Pattern Recognition and Applications)*, February, 2010, pp. 333-340

(4) K. Nishida, T. Kurita, Yasuo Ogiuchi, and M. Higashikubo, "Visual Tracking Algorithm Using Pixel-Pair Feature", *Proc. ICPR2010 (The Twentieth International Conference on Pattern Recognition)*, August, 2010, pp. 1808-1811