Extended CoHOG and Particle Filter by Improved Motion Model for Pedestrian Active Safety

Hirokatsu Kataoka, Kimimasa Tamura, Yoshimitsu Aoki
Graduate School of Science and Technology
Keio University
Yokohama, Kanagawa, Japan
Email: kataoka@aoki-medialab.org
http://www.hirokatsukataoka.net/

Yasuhiro Matsui
National Traffic Safety and Environment Laboratory
Chofu, Tokyo, Japan
Email: ymatsui@ntsel.go.jp

***Abstract*—The percentage of pedestrian deaths in traffic accidents is on the rise. In recent years, there have been calls for measures to be introduced to protect such vulnerable road users as pedestrians and cyclists. In this study, a method to detect pedestrians using an in-vehicle camera is presented. We improved the technology in detecting pedestrians with highly accurate images using a monocular camera. We were able to predict pedestrians' activities by monitoring them, and we developed an algorithm to recognize pedestrians and their movements more accurately. The effectiveness of the algorithm was tested using images taken on real roads. For the feature descriptor, we found that an extended co-occurrence histogram of oriented gradients, accumulating the integration of gradient intensities. In tracking step, we applied effective motion model using optical flow for Particle Filter tracking. These techniques are valified by using images captured on the real road.**

I. Introduction

In Japan, the number of pedestrian deaths was 4,611 in 2011. Over the next several years, the number is likely to decrease still further. However, the percentage of pedestrian deaths among all deaths in traffic accidents is increasing [1]. In an effort to reduce pedestrian deaths, investigations are being made in the area of a pedestrian intelligent transport system (ITS).

Along these lines, we are currently studying a pedestrian active safety (collision avoidance) system, which is able to detect and track pedestrians by means of an in-vehicle sensor and employ automatic braking. And we can expect much of this system. The use of in-vehicle cameras is efficient in detecting obstacles, and many studies have been devoted to pedestrian detection by means of cameras. Researchers have been studying about active safety technologies for traffic safety [2][3]. Dalal *et al* present a human classification feature, called Histograms of Oriented Gradients (HOG), and a linear SVM as a learning classifier. Now many HOG-based methods are proposed, moreover, HOG is recently improved for high-accuracy detection. Joint HOG [5] and CPF (Co-occurrence Probability Feature) [6] are the one of improved HOG method. These features consider co-occurrence feature using returning value of classifiers. Haar-like, Haar-wavelet and Edge Orientation Histograms are also used to detect pedestrian in real-world roads. Co-occurrence of Histograms of Oriented Gradients (CoHOG) is known as state-of-the-art method in the field of pedestrian detection. CoHOG feature descriptor represents edge-pair as a histogram from image patch.

Recently, object detection using stereo cameras has been investigated. Stereo-based method can easily refine background area and separate from pedestrian area. Many methods are adopting stereo-based approach, e.g. Broggi *et al*[8], Krotosky *et al*[9], and so on.

However, because of the high costs of stereo technology, a monocular camera technique is preferable. On the other hand, the use of monocular cameras as driving recorders is on the rise and they are frequently used as such by taxis all over the world. Accordingly, we devoted our efforts to developing a pedestrian active safety technique using a monocular camera.

In this paper, we present high-accuracy pedestrian detection and tracking techniques using monocular camera. The improvement version of CoHOG [7] is used for detection, and optical-flow-based motion model is implemented as tracing method. An in-vehicle video is captured using a camera attached to the rearview mirror. The system detects any pedestrians that appear in the vehicle's path. After detection and tracking, the system sends a warning signal to the driver warning or it automatically engages the brake.

A signal is emitted by the vehicle to warn the pedestrian. The system is able to judge the situation and the human reaction, and if it determines that there is the possibility of collision it engages the brake after warning the driver. By avoiding collisions, the system is able to reduce the incidence of traffic accidents. Owing to the recent increase in the use of monocular in-vehicle cameras, it is necessary that such cameras be used to detect pedestrians.

II. Proposed Method

Fig.1 shows the sequence of events in our proposed method. A video is made of the area in front of the vehicle by means of the in-vehicle camera. Symmetry is judged from the left and right images of the pedestrian by means of high-speed processing, and the pedestrian candidate area is narrowed down. The edge detector and binarization is carried out on the source image. The current method for capturing a pedestrian's shape employs co-occurrence histograms of oriented gradients

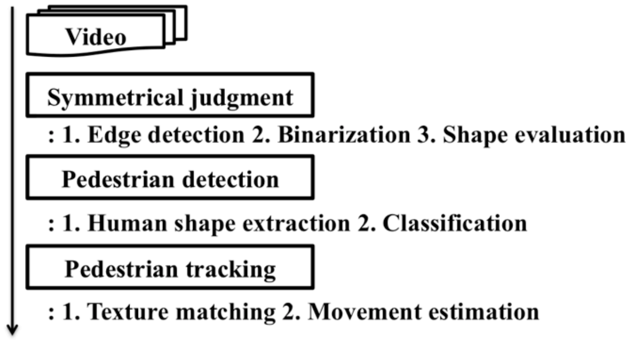

Fig. 1. The flow of pedestrian active safety system

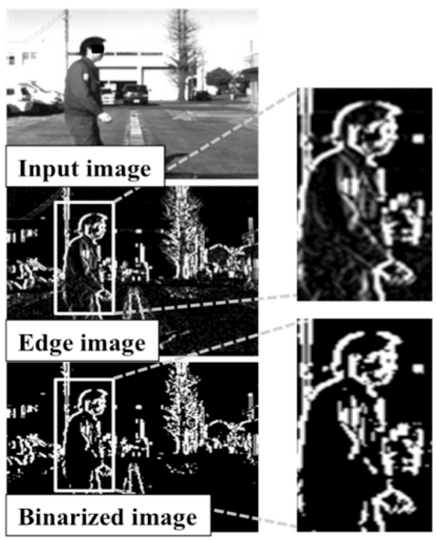

Fig. 2. Edge detection and binarization

(CoHOG) [7]. In this paper, however, we propose the use of extended CoHOG (ECoHOG) as an improvement of CoHOG. We implemented Particle Filter [11] for pedestrians tracking. Optical flow [12] is used for motion model that aquires the movement of vehicle. To test the applicability of ECoHOG, we carried out a comparison between it and CoHOG by means of commonly used datasets.

A. Symmetrical Judgment

Fig.2 shows the shape profile and binarization image of a pedestrian. From the pedestrian's symmetry it is possible to evaluate the person's shape. The edge image is used to capture the rapidly changing pixel value. Thus, it is possible to distinguish, for example, the boundary line of the human shape from the background. Binarization is performed after edge detection. As a result of the binarization, it is possible to exclude the less important edges within the overall edge of the human shape. It is thus easy to judge the outer contour.

The whole image is scanned so that the shape profile can be judged. Fig.3 shows the method of right and left symmetrical

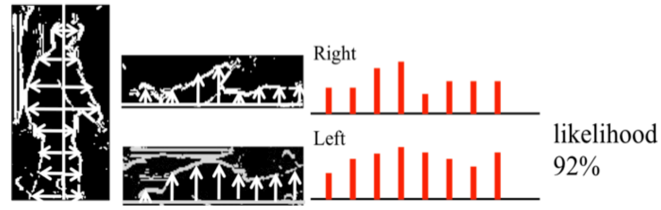

Fig. 3. Symmetrical judgment from left and right image

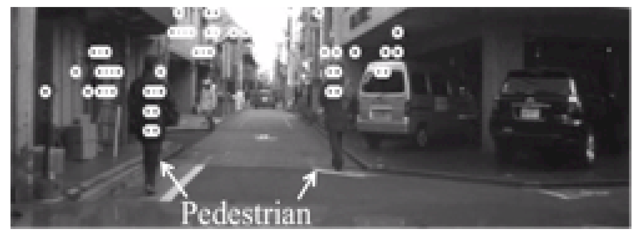

Fig. 4. Result of symmetrical judgments

judgment. The distance between the center of the window and the edge is presented as a distant histogram. The similarity between the two distant histogram is calculated using the Bhattacharyya coefficient [10]. The range of the Bhattacharyya coefficient is 0.0-1.0. We decided the threshold as 0.77 - 0.98 from 3,000 positive (pedestrian) images. The average was 0.88 and we can capture 95.44% (2 σ) samples with the threshold. Fig.4 shows the pedestrian candidate area detection by threshold value. The small circles in the image represent the pedestrian candidate area. Some circles indicate confusion of the background area with the pedestrian area. However, we can extract the pedestrian candidate area and exclude the background area.

B. Pedestrian Detection

We used ECoHOG as an improvement of CoHOG [7] for pedestrian detection. For pedestrian-shape extraction, CoHOG investigates the pedestrian's edge direction minutely and uses two different pixels in an extraction window as a direction pair. From the edge direction pair, CoHOG can describe such features as the head and shoulders, the back and legs (shown in Fig.5) and reduce misdetection by means of pair extraction. Unlike CoHOG, ECoHOG acquires not only the edge-direction pair but the edge-magnitude pair. ECoHOG can represent the edge-magnitude characteristics of the pedestrian shape.

ECoHOG : this approach is possible as edge-magnitude accumulation method to generate the pedestrian shape model in minute detail. The plus method can describe detail more comprehensively than CoHOG feature descriptor. In other words, ECoHOG considers not only the pedestrian shape features but also the degree of edge magnitude that can describe "shape strength". We believe the addition of edge-magnitude

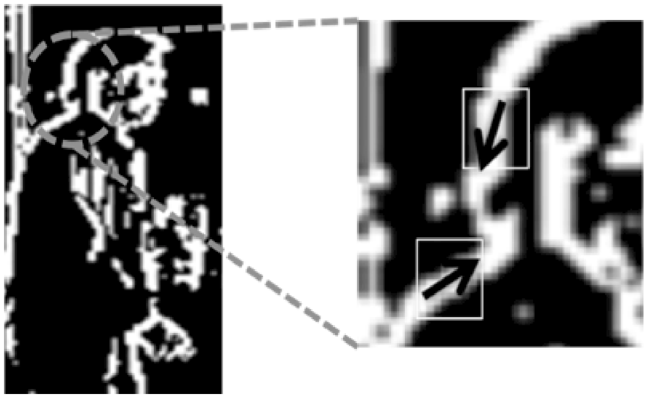

Fig. 5. How to acquire an edge pair from a pedestrian shape

pair is effective way to classify pedestrian and background high-accurately. The equation of ECoHOG is shown as below:

$$m_1(x_1, y_1) = \sqrt{f_{x1}(x_1, y_1)^2 + f_{y1}(x_1, y_1)^2} \quad (1)$$

$$m_2(x_2, y_2) = \sqrt{f_{x2}(x_2, y_2)^2 + f_{y2}(x_2, y_2)^2} \quad (2)$$

$$f_x(x, y) = I(x+1, y) - I(x-1, y) \quad (3)$$

$$f_y(x, y) = I(x, y+1) - I(x, y-1) \quad (4)$$

$$C_{x,y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases} m_1(x_1, y_1) + m_2(x_2, y_2) \\ (if\ d(p,q) = i\ and \\ \ d(p+x, q+y) = j) \\ \\ 0 \\ (otherwise) \end{cases} \quad (5)$$

where $m_1(x,y)$,$m_2(x,y)$ show the magnitude from attention pixel (center of window) and objective window, $f_x(x,y)$,$f_y(x,y)$ is the subtraction x and y direction. $C(i,j)$ is the co-occurrence histogram that accumulates pairs of attention pixel and objective pixel. Each $C(i,j)$ has 64 dimensions because the direction is quantized 8 and ECoHOG combine two directions. Coordinate (p,q) indicates the attention pixel (center of window) and coordinate $(p+x, p+y)$ is the objective pixels. m and n mean the width and height of feature extraction window. $d(p,q)$ is the function to quantize edge direction as integer value 0 - 7 on pixel (p,q).

ECoHOG normalizes the co-occurrence historam in order to set off the difference of edge-magnitude due to image brightness. Co-occrrence histogram is divided by the norm of co-occurrence histogram. Normalization is shown here:

$$C'_{x,y}(i,j) = \frac{C_{x,y}(i,j)}{\sum_{i'=0}^{7}\sum_{j'=0}^{7} C_{x,y}(i',j')} \quad (6)$$

C. Pedestrian Tracking

The searching and optimization are performed by Particle Filter [11]. Particle Filter is an iterative algorithm including prediction, update and resampling. This is a kind of Bayesian estimator that applies Monte Carlo framework. The current state is estimated as the expectation of likelihood distribution. Likelihood calculation that evaluates condition of object is texture comparison between model image and an image around particle.

Step1 Initialization

In this step, we execute initialization after specifying pedestrian's position by detector. We arrange particles aroud pedestrian and capture pedestrian's texture as tracking model.

Step2 State prediction

We apply the vehicle movement model as a motion model. Vehicle movement is calculated by optical flow [12] that tracks feature point between spatio-temporal frames Fig.6. The average of optical flow direction and value in a frame is movement model. And more, we add Gaussian noise to particle's velocity and location. We estimate by linear uniform motion if there's no vehicle movement as below:

$$X_{t+1} = FX_t + w_t \quad (7)$$

$$X_t = (x, y, v_x, v_y)^{\mathrm{T}} \quad (8)$$

$$F = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (9)$$

where x, y are the location of particle, v_x, v_y are the velocity between frames. w_t shows system noise adding paritcle location and velocity.

Step3 Likelihood calculation

The likelihood is calculated comparing two textures. The system captures model texture in detected area and compares model with a texture around particle Fig.7. The evaluation method is Normalized Cross Correlation (NCC). The calculated value is likelihood between two textures.

$$L = \frac{\sum_{j=0}^{N-1}\sum_{i=0}^{M-1}(I(i,j)T(i,j))}{\sqrt{\sum_{j=0}^{N-1}\sum_{i=0}^{M-1} I(i,j)^2 \sum_{j=0}^{N-1}\sum_{i=0}^{M-1} T(i,j)^2}} \quad (10)$$

where L is the returned likelihood in NCC, $I(i,j)$, $T(i,j)$ are the texture of model and image on coordinate (i,j). And N and M are the size of texture.

Step 4 Likelihood evaluation

The center of gravity is calculated in this step. We calculate the center of gravity shown as below:

$$(g_x, g_y) = (\sum_{i=1}^{n} L_i x_i, \sum_{i=1}^{n} L_i y_i) \quad (11)$$

g is the center of gravity of pedestrian. L is each particle's likelihood and x and y are position of particle.

Fig. 6. Flow vector calculation in driver's view

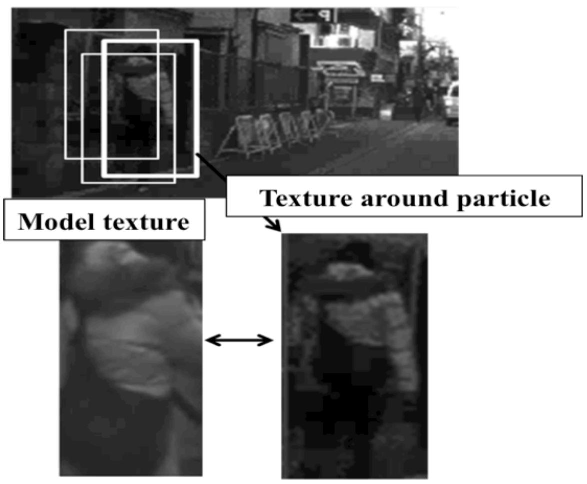

Fig. 7. Likelihood calculation by texture matching

III. EXPERIMENT

We carried out the experiments for validating detection and tracking approaches. Experiment 1 was executed to choose the feature descriptor, experiment 2 was a comparison between the previous method and the newly proposed method and experiment 3 was a velification about tracking methods. We applied the INRIA person dataset as a benchmark for experiment 1, and videos captured in real-road for experiment 2 and 3.

A. Experiment 1: Feature Descriptor Verification on INRIA Person Dataset

In this experiment, we applied the INRIA Person Dataset [13], including pedestrians and the detection-error tradeoff (DET) curve to evaluate the detection accuracy. The effectiveness of CoHOG has been established. We carried out a comparison of CoHOG with ECoHOG. The DET curve's vertical axis indicates the misdetection rate, when the system is unable to detect a pedestrian; the horizontal axis is the false-positive rate, which is incorrect detection. The DET curve shows high accuracy close to the lower left part of the graph.

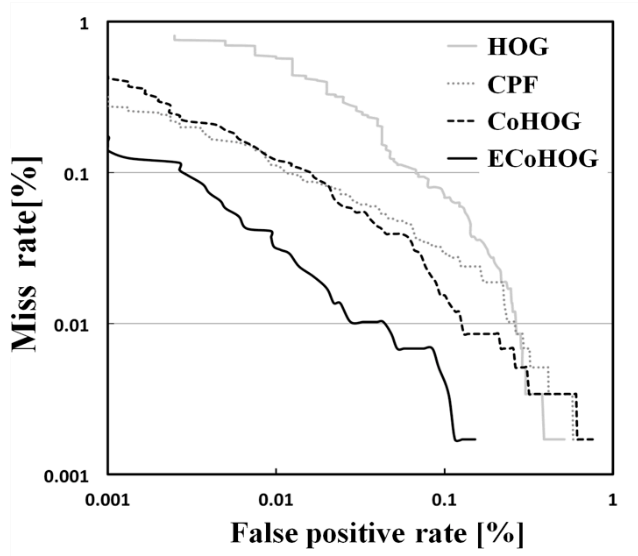

Fig. 8. Detection Error Tradeoff (DET) Curve

We evaluated the detection accuracy of the proposed method and previous methods, the comparison of the ECoHOG and CoHOG, CPF and HOG. The results of the evaluation experiment are shown in Fig.8 in accordance with the DET curve.

The evaluation experiment revealed that the ECoHOG was better than other methods. We added edge-magnitude and normalization step to ECoHOG from CoHOG. The ECoHOG method is improved using magnitude pair representation. And more, And normalization step sets off the intensity differences due to pixel values. From the result between CoHOG and HOG, the effectiveness of co-occurrence extraction is indicated. The CPF feature is better than HOG feature due to the CPF is also extracting co-occurrence values.

In the next section, we address the ECoHOG plus method for pedestrian detection using the in-vehicle video to reduce over-detection and false alarms.

B. Experiment 2: Effectiveness of Detection in Real-scenes

We verified the pedestrian-detection accuracy using videos with real-load capture. We used precision, recall, and the F measure as indicators of detection accuracy, whereby precision indicates the accuracy of the system, recall is the detection rate of system in an actual pedestrian area, and the F measure is calculated as the harmonic mean from precision and recall. Precision, recall, and the F measure are as shown here:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (12)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (13)$$

$$Fmeasure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (14)$$

TABLE I
THE ACCURACY OF DETECTION IN REAL-ROADS BY PRECISION, RECALL AND F MEASURE

	Precision	Recall	F measure
CoHOG	0.6811	0.5949	0.6321
ECoHOG	0.7922	0.5949	0.6795

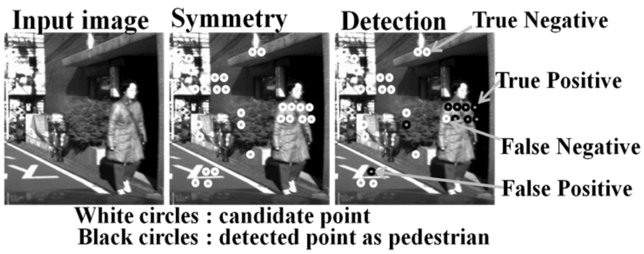

Fig. 9. The method of pedestrian detection on road

The system being able to correctly identify a pedestrian signifies a true-positive result, whereas identifying a pedestrian as the background is a false-negative result. Similarly, the system identifying the background as a pedestrian is a false-positive result, and correctly identifying the background is a true-negative result.

The settings were the same as those described in experiment 1 - the verification experiment on feature descriptor improvement. To help detect pedestrians, we prepared learning images for generating the pedestrian-shape model. There were 5,000 positive images and 20,000 negative images. Evaluation images including pedestrians were extracted from the in-vehicle video, which was taken in the Tokyo metropolitan area. Fig.9 shows the evaluation images. Table.1 shows the experimental results using the previously applied method (CoHOG) and the newly proposed method (ECoHOG) evaluated in terms of precision, recall, and the F measure.

In experiment 2, the newly proposed method-ECoHOG-gave a higher value in the F measure than CoHOG. This indicates that ECoHOG is superior in terms of accuracy (Fig.10). We believe this accuracy derives from the edge-magnitude accumulation and normalization of the ECoHOG method. ECoHOG can correctly identify pedestrians and the background in cases where the background includes object that have a similar edge-magnitude accumulation. We can configure a pedestrian model that correctly detects pedestrians by digitizing the characteristics of the human edge magnitude from a large number of learning images. Normalization is able to eliminate histogram changes due to image-strength variation by means of histogram shape setting. By means of normalization, we can decrease problematic cases that arise due to light-source variations and pedestrian shadows that appear in the images as a result of environmental changes.

When the vehicle is moving fast, features of a pedestrian become difficult to identify because the edges become less distinct. Difficulties in pedestrian detection emerge from discrepancies between the pedestrian-shape model and the extraction image. We need to add faded-edge pedestrian images to generate better pedestrian-shape models. We are currently preparing further pedestrian samples so as to improve the detection precision because the 5,000 positive images used in this study were insufficient.

Fig. 10. Sample of detection

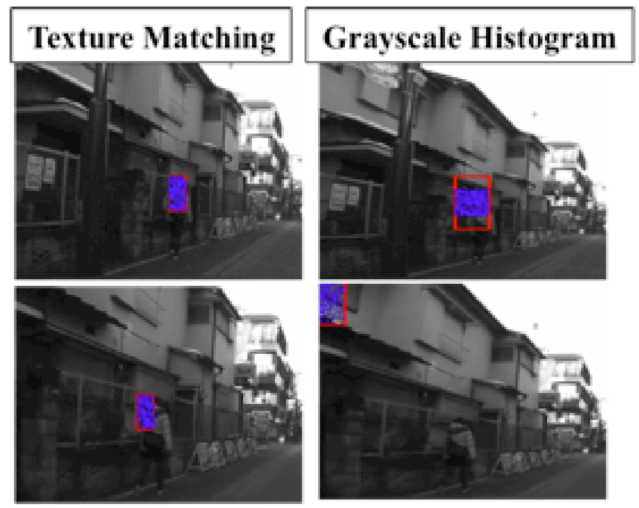

Fig. 11. Pedestrian tracking

C. Experiment 3: Effectiveness of Tracking in Real-scenes

We carried out tracking experiment using our real-road dataset. The objective of this method is to track one pedestrian, it doesn't include occlusion or crossing some pedestrians. We applied two methods (i) grayscale histogram matching + linear uniform motion and (ii) texture matching + linear uniform motion as related approaches. Our proposed method (iii) texture matching + optial flow of automotive is compared with related approaches. The initial position was set manually and histogram or texture is extracted from initial rectangle. Table.2 shows the result of tracking experiment and Fig.11 shows the tracking on real-road dataset.

The grayscale histogram doesn't evaluate human likelihood effectively. This result comes from ignoring the location of human's texture of pose and close. The histogram feature is confusing unless it is color depth. And more, grayscale value is changing dramatically in real time(Fig.12). We applied texture matching including location information from pedestrian. In this case, the time-series texture of pedestrian has few changing. Texture matching is better way to capture pedestrian from

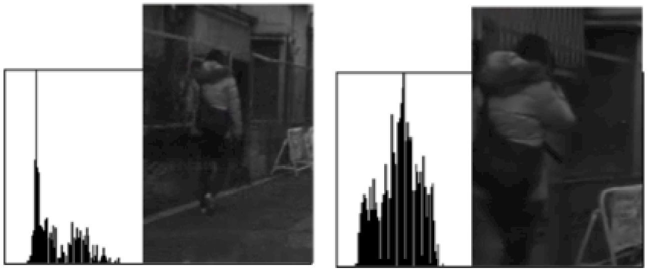

Fig. 12. Difference between two scenes' histograms

TABLE II
THE RESULT OF TRACKING EXPERIMENT

Tracking method	Accuracy(success/all)	Processing time (ms)
(i)	18.7% (280/1498)	16.02
(ii)	72.3% (1084/1498)	19.23
(iii)	97.2% (1456/1498)	37.05

in-vehicle video. However, texture accumulates gap due to different angle and scale if tracking continues long time. We need to fix the template using interest point tracker.

According to processing time calculation, our proposed approach can track in real-time. The processing time, 37msec (27fps) is high-speed, however, detection step is time-consuming. We try to complement detection method from tracking information for fast implementation.

IV. Conclusion

In this paper, we proposed the detection and tracking method for pedestrian active safety. The flow of approach is (1) narrowing down pedestrian candidate area from symmetry judgment (2) pedestrian detection using Extended CoHOG feature descriptor and (3) Particle Filter tracking by vehicle motion model.

The next work is to detect and track pedestrians in the context of occlusion. We can capture pedestrian position if slightly overlap each other, however, crowded situation is difficult to detect pedestrian. We try to complement information from detection and tracking for high-speed and highly accurate detection.

References

[1] "Institute for Traffic Accident Research and Data Analysis of Japan (ITARDA)", Annual Traffic Accident Report in 2010, Tokyo, 2011

[2] Tarak Gandhi, Mohan M. Trivedi, "Pedestrian Collision Avoidance Systems:A Survey of Computer Vision Based Recent Studies", IEEE Intelligent Transportation Systems Conference (ITSC2006), pp.976-981, 2006

[3] David Geronimo, Antonio M. Lopez, Angel D. Sappa, Thorsten Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.32, No.7, 2010

[4] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR2005), pp886-893, 2005

[5] T. Mitsui, Y. Yamauchi, and H. Fujiyoshi, "Object Detection by Two-Stage Boosting with Joint Features", The Institute of Electronics, Information and Communication Engineers Transactions on Information and Systems,Vol. J92-D, No. 9, pp. 1591-1601, 2009

[6] Y. Yamauchi, H. Fujiyoshi, Y. Iwahori, and T. Kanade, "People Detection based on Co-occurrence of Appearance and Spatio-Temporal Features", National Institute of Informatics Transactions on Progress in Informatics, No. 7, pp. 33-42, 2010

[7] Tomoki Watanabe, Satoshi Ito, Kentaro Yokoi, "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection", PSIVT2009, LNCS vol.5414, pp37-47, 2009

[8] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-Based Pedestrian Detection", IEEE Intelligent Vehicles Symp., pp.215-220, 2000

[9] S. Krotosky and M.M. Trivedi, "Multimodal Stereo Image Registration for Pedestrian Detection", IEEE Int Conf. Intelligent Transportation Systems, pp.109-114, 2007

[10] A. Bhattacharyya,"On a measure of divergence between two statistical populations defined by their probability distributions", Bulletin of the Calcutta Mathematical Society, 35, pp99-109, 1943

[11] Isard M, Blake A, "Condensation : conditional density propagation for visual tracking", International Journal of Computer Vision, vol.28, no.1, pp.5-28, 1998

[12] B.D.Lucas, Takeo Kanade, "An iterative image registration technique with an application to stereo vision", Proceedings of Imaging Understanding Workshop, pp.121-130, 1981

[13] INRIA Person Dataset, http://pascal.inrialpes.fr/ data/human