2012 15th International IEEE Conference on Intelligent Transportation Systems
Anchorage, Alaska, USA, September 16-19, 2012

Urban Road User Classification Framework using Local Feature Descriptors and HMM

Toshimitsu Takahashi[1], HyungKwan Kim[1], and Shunsuke Kamijo[1]

Abstract— **Surveillance and safety systems for pedestrians and bicyclists are becoming much more important because there continue to be a large number of traffic accidents that involve vulnerable road users. In this paper, we propose an urban road user classification framework using local feature descriptors and hidden Markov models (HMM). Our framework achieved pedestrians, bicyclists, motorcyclist classification in high accuracy. The framework consists of two classification methods: pedestrian-bicyclist classification and bicyclist-motorcyclist classification. First, we discriminate between pedestrians and bicyclist-like objects using histograms of oriented gradients (HOG)-based classifiers. We implemented a cascade classifier using generic HOG and our original local feature descriptor called co-occurrence semantic HOG. Bicyclist-like objects mainly consist of bicyclists and motorcyclists. We focused on the objects' leg motions and classify them using the hidden Markov models (HMM)-based motion models. We conducted experiments with real traffic scenes to evaluate the performance of our framework. The experiments for pedestrian-bicyclist classification and bicyclist-motorcyclist classification are conducted independently and both methods achieve nearly 90% on classification.**

I. INTRODUCTION

Traffic accidents involving pedestrians and bicyclists are still a critical problem because they result in a higher rate of fatal accidents. Not only accidents between vehicles and pedestrians or bicyclists but accidents between bicyclists and accidents between bicyclists and pedestrians are dangers as well. Conflicts between people using different methods of transport are also an important problem. On sidewalks, conflicts between pedestrians and bicyclists do occur. In addition, conflicts between bicyclists and motorcyclists occur on roads. These conflicts cause not only inconvenience but also yield traffic inefficiency. To solve these problems, we need detailed information on the traffic conditions such as the trajectory and type of objects. Image processing and pattern recognition are widely used for this purpose.

To distinguish pedestrians from vehicles, Zhou and Aggarwal[1] classified moving objects into three categories using motion, spatial position, shape, and color information. Pai *et al.*[2] focused on human's walking rhythm and distinguish pedestrians from vehicles on infra-taken images. They proposed a non-rigid body model where the entropy on lower part of object is employed as a pedestrian indicator. Zhang *et al.*[3] adopted local feature descriptors and classified objects by boosting different descriptors.

[1]T. Takahashi, H.K. Kim, and S. Kamijo are with the Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan {takahashi, kim} at kmj.iis.u-tokyo.ac.jp, kamijo at iis.u-tokyo.ac.jp

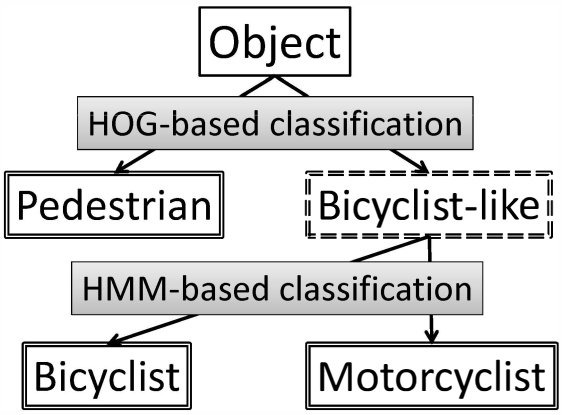


Fig. 1. Flowchart of traffic classification.

For a more detailed classification, Buch *et al.*[4] employed 3D models of objects and classified them into five categories. Three of them are for vehicles (bus/lorry, van, car/taxi) and the remaining two are motorbike/bicycle and pedestrian. Similarly, Chen and Zhang[5] proposed a vehicle classification framework based on independent component analysis and the support vector machine.

In this paper, we propose a novel object classification framework that aims to detect pedestrians and bicyclists. The system consists of two phases of classification: pedestrian-bicyclist classification and bicyclist-motorcyclist classification. Pedestrian-bicyclist classification is based on a HOG feature[6] and Fisher's linear descriminant. We also use our original local feature descriptor called co-occurrence semantic HOG (Co-Sem-HOG) that is based on a HOG feature. Bicyclist-motorcyclist classification is based on HMM. We focused on the objects' leg motions and use motion information for classification. Object detection and tracking are achieved using the spatio-temporal Markov random field (S-T MRF) model[7].

II. FRAMEWORK OVERVIEW

The framework consists of two phases of object classification (Fig. 1). The first phase distinguishes pedestrians and bicyclists/motorcyclists from moving objects based on a HOG feature , Co-Sem-HOG, and Fisher's linear discriminant. The second phase classifies bicyclists and motorcyclists based on HMM. It is difficult to distinguish bicyclists from motorcyclists precisely only by local feature descriptors because their appearances are similar. As such, we employed a two-phase classification and also used objects' motion

978-1-4673-3063-3/12/$31.00 ©2012 IEEE

information for further classification. The object detection and tracking is achieved by the spatio–temporal Malkov random field (S–T MRF) model, that appears in our previous work[7]. The S–T MRF model is the method to divide an image into blocks as a group of pixels, and to optimize the labeling of such blocks by referring to texture and labeling correlations among them, in combination with their motion vectors. Combined with a stochastic relaxation method, the S–T MRF optimizes object boundaries precisely, even when serious occlusions occur. Then we can get regions of interest (ROIs) of objects.

III. Pedestrian-Bicyclist Classification

For sidewalk surveillance, distinguishing pedestrians from bicyclists is important. Here, we show a classification system based on local feature descriptors and Fisher's linear discriminant.

A. HOG Feature

Among the various algorithms for pedestrian detection, HOG[6] is well-known for its performance in object classification. Combined with machine-learning algorithms such as the support vector machine (SVM) or Fisher's linear discriminant, it performs quite well for object classification. We employed a HOG feature to implement a pedestrian-bicyclist classifier. In what follows, we give an overview of the HOG extraction process implemented in our system.

The HOG feature represents the spatial distribution of edge on the scene. The pedestrian image set that consists of various moving poses and background images was used as training data for Fisher's linear discriminant. In this paper, training images are scaled into 64×128 pixels, and a block of 8×8 pixels is defined to extract a local HOG feature. An orientation of gradients is estimated by applying an edge operation to each pixel, and the orientation is quantized into nine measures. Sixty-four quantized orientations for 8×8 pixels are plotted into a histogram with respect to nine measures. This histogram is translated into a vector of nine dimensions, with each value representing one of the nine magnitudes in the histograms. Gradient strengths vary among images and locations within an image owing to illumination and foreground-background contrast. In order to cancel such effects, the descriptor vectors are normalized. In this paper, a descriptor block of 16×16 pixels consisting of four 8×8 pixel cells is defined, and a descriptor vector of 36 dimensions is obtained by connecting four 9 dimension vectors. A sequence of descriptor blocks is obtained by shifting the region by 8 pixels into the direction of raster scan, and thus a sequence should consist of 7×15 descriptor blocks. Each descriptor vector is normalized with the square norm to be a normalized descriptor vector. Thus, a vector of 3780 dimensions is obtained by connecting 105 normalized descriptor vectors with respect to a training image.

B. Co-Occurrence Semantic HOG Feature

In addition to conventional HOG, we adapted our original HOG-based feature to describe a pedestrian. We refer to it as the co-occurrence semantic HOG (Co-Sem-HOG). The feature aims to make a meaningful non-linear space that is effective in decreasing false positive results of the linear classifier trained in the HOG space. The Co-Sem-HOG feature can be extracted by calculating the co-occurrence matrices based on Semantic HOG. The Semantic HOG is a non-linear feature that replaces edge orientation histograms of HOG with 6 simbols according to deviations of the histograms. If the histogram of a cell shows an explicitly strong deviation to the pre–decided four directions, the cell is symbolized with a corresponding direction. If the cell hardly contains edge factors, then the cell is symbolized with "few texture" and the cells with small deviations are symborized as "complex texture" (see Fig. 2.) An example of the making of the semantic HOG is illustrated in Fig. 2. First, as the semantic HOG contains six symbols, the matrix should be a 6–dimensional square matrix. By comparing with 8 adjacent cells, a co–occurrence matrix is gradually voted with a matched combination of two symbols. The process can be described as follows:

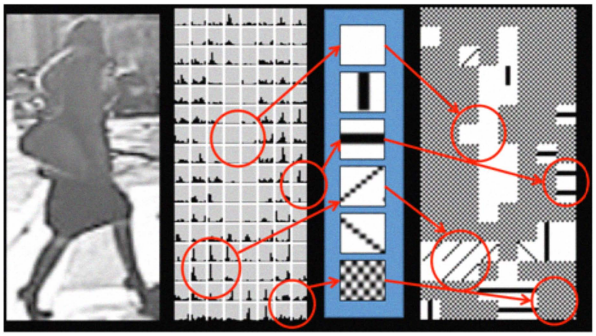


Fig. 2. Symbolization to Semantic HOG.

$$C_{m,n}(i,j) = \sum_{x=-1}^{1} \sum_{y=-1}^{1} \alpha,$$
$$\text{where } \alpha = \begin{cases} 1 & \text{if } S(m,n) = i, S(m+x, n+y) = j, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Here, $C_{m,n}$ is the co–occurrence matrix at the cell (m,n). i and j correspond to the symbols of the Semantic HOG. The process is conducted at each cell except for those in boundaries so that in the case of 16×8 size cells, the dimension of the feature vector would be $(16-2) \times (8-2) \times (6^2) = 3,094$. Further, in order to capture the characteristic "straight" distribution of non-pedestrian objects, if the target cell is one of the four directional symbols, the matrix is calculated based on investigation cells expanded according to a target symbol (see Fig. 3). The examples of cell–wise histograms of HOG features and Co–Sem–HOG features for pedestrians and bicyclists are shown in Fig. 4.

We constructed a 2-step cascade of the heterogeneous linear classifier. First classifier is HOG-Fisher classifier and second uses Co-Sem-HOG instead.

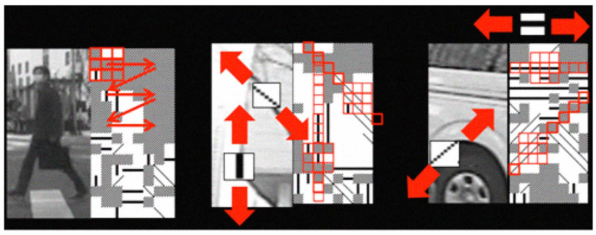
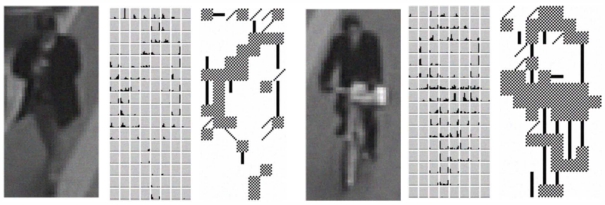


Fig. 3. Co-Sem-HOG feature extraction.



(a) (b)

Fig. 4. Comparison of the extracted HOG feature and Co-Sem-HOG between a pedestrian (left) and a bicyclist (right).

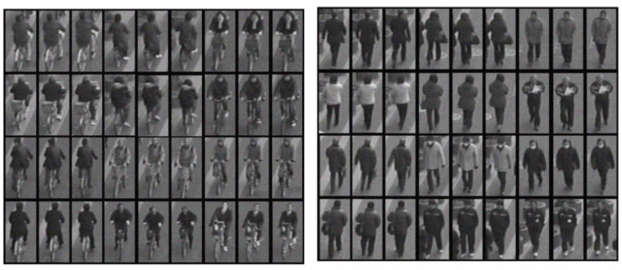
C. Fisher's Linear Discriminant and Training Data

The Fisher's linear discriminant is widely used in the field of object detection. Compared with other linear discriminants, Fisher's linear discriminant delivers the same level of detection rate with much lower computational cost. In this paper, we implemented a pedestrian-bicyclist classifier using a Fisher's linear discriminant trained in the space of the HOG feature. $10,000$ images each of a bicyclist and a pedestrian were prepared as training data. Fig. 5 shows the example of the training images. Because the pedestrians and bicyclists exhibit different patterns during the movement, three different poses was automatically cropped out from each bicyclist and pedestrian based on tracking results.

IV. Bicyclist-Motorcyclist Classification

On roads, bicyclists and motorcyclists run almost the same area. Discriminating between bicyclists and motorcyclists in the first phase is difficult because both have similar appearances.

As such, we propose a motion-based classification system to classify bicyclist-like objects into bicyclist and motorcyclist. We focused on the pedaling motion of bicyclists.



(a) (b)

Fig. 5. Examples of training images. Top : bicyclists, Bottom : pedestrians.

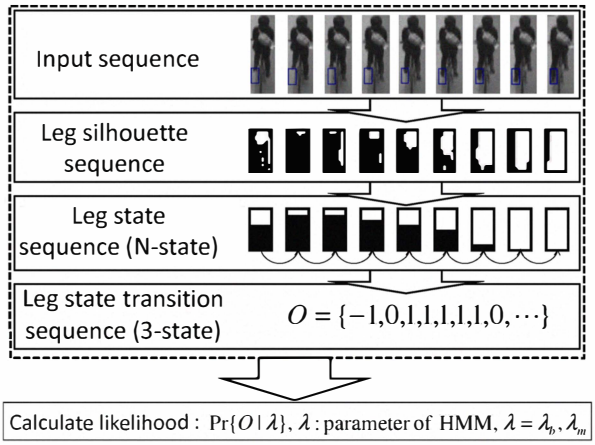


Fig. 6. Bicyclist-motorcyclist classification process overview.

The essential idea of the system is to model leg motions using HMM and classify objects with the leg motion state sequence.

A. Classification Overview

The processing we use to extract objects' leg motion state sequence from a cropped image sequence is shown in Fig. 6. First, we use simple rules to determine the ROIs that include the legs of the object. Then, we obtain the leg silhouette sequences using ROIs and background subtraction and convert them to N-state sequences. Based on the differences of the leg state sequence, we get the leg motion state sequence. Finally, we classify the object according to the HMM likelihood of the leg motion sequence. Because the parameters of leg motion HMM are unknown, we use the Baum-Welch algorithm[8] to estimate the parameters.

B. First Step: Determination of leg ROI

As for bicyclists and motorcyclists, the position of the leg (relative to head position) is quite restricted. As such, we can estimate the region of interest (ROI) where the leg of the objects appears, using a simple method.

For precise estimation of ROI, we need three–dimentional human body model and projective transformation of leg moving area. Projective transformation from a three-dimensional (real world) point Q to a two-dimensional (image) point q can be achieved as follows:

$$q = MWQ, \tag{2}$$

where M is the camera parameter matrix and $W = [R\ \mathbf{t}]$ (R is the rotation matrix and $\mathbf{t}$ is the parallel translation vector). M and W can be estimated from several sample points and camera calibration.

To determine a ROI including the object's leg, we should project the three dimentional cubic which shows the space that the leg moves through to the two dimentional rectangle as shown in Fig. 7. Here we set Q_0 as the head coordinates and Q'_i as the relative coordinates of vertex i. As such, the absolute coordinates of vertex i are $(Q_0 + Q'_i)$. Q'_i can be calculated using the direction of object θ.

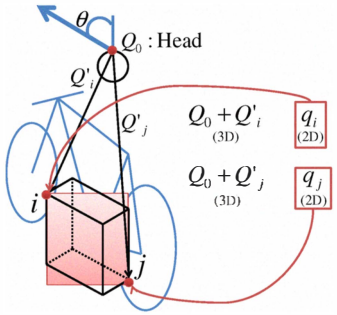


Fig. 7. ROI setting process.

TABLE I
LEG STATES (N=9).

State	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
Silhouette									

To aquire q_i, we should project $(Q_0 + Q'_i)$ using 2. However, we can roughly estimate q_i and q_j directly from θ and the object's height when the moving direction of the objects is limited. We employ this heuristic estimation.

C. Second Step: Acquiring the Sequence of Leg States

After the first step, we have a sequence of clipped silhouette images of the object's leg. In second step, we convert it to a N-state sequence using simple method. One of the leg states is selected for one leg silhouette as follows:

$$J = \arg\max_{s \in S} r(I, s), \tag{3}$$

where I is the observed silhouette, J is the corresponding leg state, S is the set of leg states, and $r(I, s)$ is the matching rate between the silhouette I and the corresponding silhouette of s. Each leg state s corresponds to one silhouette as shown in Table I.

After this conversion process, we get the leg state sequences of the objects.

D. Third step: Conversion of the sequence

After the second step, we have N-state Markov sequences. Since we are interested in leg motions, we use the differences of the leg state sequences. After converting difference sequences with threshold, we get the leg motion sequences as three-state discrete sequences (Fig. 8).

The state of the leg motion sequences can be up (state "1"), level (state "0"), or down (state "-1").

E. Fourth (final) step: Classification of the motion sequences using HMM

To classify the leg motion sequences, we use HMM. HMM here has three unobserved (discrete) states and three discrete observations (Fig. 9). Because the parameters are unknown, we use the Baum-Welch algorithm to estimate the HMM

$$\begin{aligned} x_{state,t=1\ldots T} &= \{6,7,5,5,3,\cdots\} \\ &\downarrow \\ x_{diff,t=2\ldots T} &= \{1,-2,0,-2,\cdots\} \\ &\downarrow \\ x_{motion,t=2\ldots T} &= \{1,-1,0,-1,\cdots\} \end{aligned}$$

Fig. 8. Example of conversion from state sequence to motion sequence.

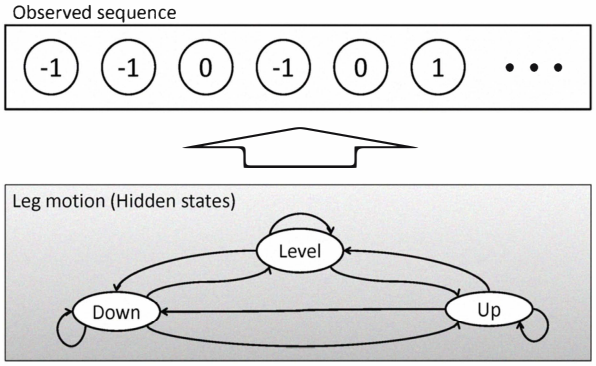


Fig. 9. HMM of leg motion.

parameters. The Baum-Welch algorithm is a well-known algorithm based on the expectation-maximization approach and uses sample sequences to estimate the HMM parameters. Here, we have to estimate two sets of parameters, one for bicyclists and the other for motorcyclists.

After estimating the HMM parameters for bicyclists and motorcyclists with sample sequences, the likelihoods of test sequences are calculated for bicyclist-HMM and motorcyclist-HMM and each sequence is classified based on the likelihoods.

V. EXPERIMENTS

We conducted experiments in a real traffic scene to demonstrate the performance of each classification system.

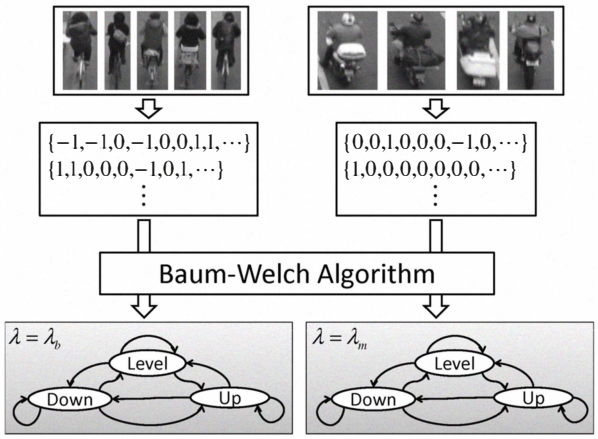


Fig. 10. Estimating the parameters of HMM using the Baum-Welch algorithm.

A. Pedestrian-Bicyclist Classification

In order to evaluate our classifier, we prepared 1,575 images of bicyclists and 4223 images of pedestrians. This video sequence contains various scenes taken from different angles and different spots. Therefore, we can evaluate how the system can be robust to such a deviation. Further, the images used for training are not included in this evaluation set. As an evaluation criteria, we used the detection rate. As the detection rate of each object is the trade-off relationship, we adjusted both values to have the same level. The result is shown in Table II.

We employed the S–T MRF model[7] for object tracking. By applying our classifier to the tracking results, we confirmed that the detection rates can be improved by more than 90%. However, evaluation on the system was still conducted as a feasibility study. Therefore, in this paper, we concentrate on the result images obtained from the system. Two result images on different scenes are shown in Fig. 11 and Fig. 12.

Fig. 11 shows the scene taken up-straight from the pedestrian road. If the system classified the object as a bicyclist, the box will be colored yellow. If not, the box will be colored blue. The bicyclists appear on all still cuts from Fig. 11(a) to Fig. 11(c). In Fig. 11(a) and Fig. 11(c), the back face and front face of the bicyclist is being tracked and the image is successfully classified as a bicyclist. Further, in Fig. 11(b) and Fig. 11(c), whether or not the bicyclist is carring a cargo is correctly detected. In Fig. 11(a), pedestrians in various poses and attires appear and are successfully detected and classified. From those results, we get that the system has a certain level of robustness to such deviations.

Scene B in Fig. 12 is much more challenging than scene A because the video sequence was taken from a skewed angle. In addition, as the shooting range is wide, objects that are under tracking appear much smaller in size in the approaching direction. In Fig. 12(a) and Fig. 12(b), bicyclists in a skewed direction are correctly detected and classified. Our system can thus be said to be robust to this kind of direction deviations. Further, in Fig. 12(c), the front-face and back-face pedestrians can be detected without any problems. From those results, our bicyclist-pedestrian classifier is robust to such deviations that easily occur in the real world.

B. Bicyclist-Motorcyclist Classification

We used several real traffic scenes that include 98 bicyclists and 96 motorcyclists. To estimate the parameters of HMMs, we used 2000 sequences of varied length each for bicyclists and motorcyclists as samples. Sample sequences are not included in test sequences. The likelihood of each object is calculated each time after the length of the motion sequence exceeds the threshold. The threshold here is 10. The classification process uses the sum of the likelihood and is applied each time. To precisely evaluate the performance of the bicyclist-motorcyclist classification, object tracking here is done manually. Table III shows the results of the experiments and Fig. 13 shows the example images.

VI. Conclusion

In this paper, we proposed an urban road user classification framework that consists of pedestrian-bicyclist classifier and bicyclist-motorcyclist classifier. Pedestrian-bicyclist classification was constructed by a cascade of Fisher's linear discriminants trained in the space of HOG feature and our original feature Co-Sem-HOG. Bicyclist-motorcyclist classifier uses HMM. Our detection rate was 90% on pedestrian-bicyclist classification and 88.9% on bicyclist-motorcyclist classification with real traffic scene.

As a future work, we plan to adapt another HOG-based feature for raising detection rate of pedestrian-bicyclist classification. Also, we are currently improving the performance of pedestrian-bicyclist classification by employing precise estimation on leg part of ROI and an optical flow.

VII. ACKNOWLEDGEMNT

This work was partially supported by PREDICT from Ministry of Internal Affairs and Communications.

References

[1] Q. Zhou and J. Aggarwal, "Tracking and classifying moving objects from video," in *Proceedings of IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, vol. 0012. Hawaii, USA, 2001. [Online]. Available: http://nguyendangbinh.org/Proceedings/CVPR/2001/CD1/PETS/zhou.pdf

[2] C.-J. Pai, H.-R. Tyan, Y.-M. Liang, H.-Y. M. Liao, and S.-W. Chen, "Pedestrian detection and tracking at crossroads," *Pattern Recognition*, vol. 37, no. 5, pp. 1025–1034, May 2004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0031320303003959

[3] Z. Zhang, M. Li, K. Huang, and T. Tan, "Boosting local feature descriptors for automatic objects classification in traffic scene surveillance," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4761317

[4] N. Buch, J. Orwell, and S. Velastin, "Urban road user detection and classification using 3D wire frame models," *IET Computer Vision*, vol. 4, no. 2, p. 105, 2010. [Online]. Available: http://link.aip.org/link/ICVEBI/v4/i2/p105/s1&Agg=doi

[5] X. Chen and C. Zhang, "Vehicle classification from traffic surveillance videos at a finer granularity," *Advances in Multimedia Modeling*, pp. 772–781, 2006. [Online]. Available: http://www.springerlink.com/index/gq6u1773r8040440.pdf

[6] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1467360

[7] S. Kamijo, Y. Matsushita, and K. Ikeuchi, "Occlusion robust tracking utilizing spatio-temporal markov random field model," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 1. IEEE, 2000, pp. 140–144. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=905292

[8] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966. [Online]. Available: http://www.jstor.org/stable/2238772

TABLE II

EXPERIMENTAL RESULTS ON PEDESTRIAN-BICYCLIST CLASSIFICATION.

	Object type	Object number	Classified as a bicyclist	Classified as a pedestrian	Detection rate (%)
HOG	Bicyclist	1,575	1,331	244	84.5
	Pedestrian	4,223	658	3,565	84.4
HOG + Co-Sem-HOG (2-step cascade)	Bicyclist	1,575	1,402	173	89.0
	Pedestrian	4,223	468	3,755	88.9

(a) (b) (c)

Fig. 11. Examples of experiments on pedestrian-bicyclist classification (Scene A). Blue rectangle : pedestrian. Yellow rectangle : bicyclist.

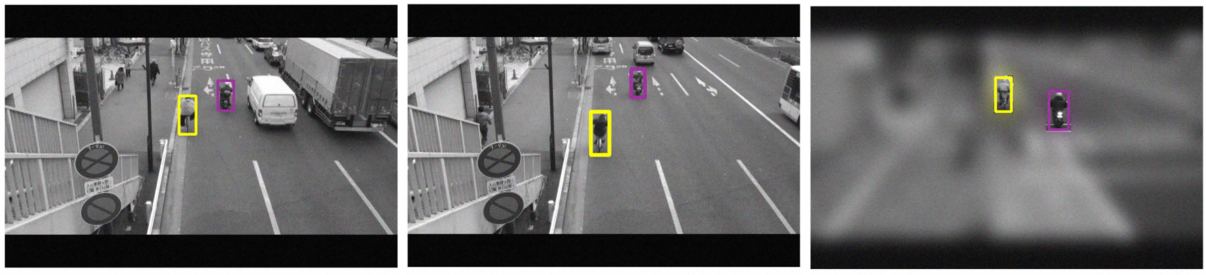
(a) (b) (c)

Fig. 12. Examples of experiments on pedestrian-bicyclist classification (Scene B). Blue rectangle: pedestrian. Yellow rectangle: bicyclist.

TABLE III

EXPERIMENTAL RESULTS ON BICYCLIST-MOTORCYCLIST CLASSIFICATION.

Object type	Sequence number	Classified as a bicyclist	Classified as a motorcyclist	Detection rate (%)
Bicyclist	2,300	2,102	198	91.4
Motorcyclist	909	158	751	82.6
Total	3,209	-	-	88.9



(a) (b) (c)

Fig. 13. Examples of experiments on bicyclist-motorcyclist classification. Yellow rectangle: bicyclist. Purple rectangle: motorcyclist. (a), (b): scene 1. (c): scene 2. The image of scene 2 is partially masked because of privacy issue.)