2017 International Conference on Security, Pattern Analysis,
and Cybernetics (SPAC)

G2P: A New Descriptor for Pedestrian Detection

1st Ming Yang
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
MingYANG@sjtu.edu.cn

2nd Yeqiang Qian
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
qianyeqiang@sjtu.edu.cn

3rd Linji Xue
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
xuelinji9568@163.com

4th Hao Li
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
haoliarts@gmail.com

5th Liuyuan Deng
Department of Automation
Shanghai Jiao Tong University
Shanghai, China
lydeng@sjtu.edu.cn

6th Chunxiang Wang*
Robotics Institute
Shanghai Jiao Tong University
Shanghai, China
wangcx@sjtu.edu.cn

***Abstract*—Pedestrian detection plays an important role in intelligent vehicle applications. Since its birth 12 years ago, the Histogram-Of-Gradient (HOG) descriptor has become a popular descriptor for pedestrian detection, thanks to its effectiveness in capturing implicit human characteristics. Besides its original instantiation, the HOG also reflects a general methodology of constructing descriptors based on histograms of gradients of certain image sub-blocks. Following this general methodology, a number of HOG-style descriptors have been reported in literature. Three contributions are made in this work. First, a general model called Descriptor Generation Model (DGM) is proposed, which can be used to systematically construct a wide range of HOG-style descriptors for pedestrian detection. Second, based on the DGM, a pedestrian detection experimental framework (PDEF) is introduced to find the optimal HOG-style descriptor. In the PDEF, the performance of each descriptor can be evaluated. At last, the genetic algorithm is employed to search the optimal (or semi-optimal) HOG-style descriptor in the descriptor space. And a new descriptor named Second-order Gradient for Pedestrian detection (G2P) is presented. Experimental results demonstrate the advantage of the G2P descriptor over the standard HOG descriptor with ETH, CVC-02-system, NITCA and KITTI dataset, which also reflects the effectiveness of the DGM-based PDEF in finding better descriptors for pedestrian detection.**

***Index Terms*—computer vision, pedestrian detection, descriptor, G2P, HOG, genetic algorithm.**

I. Introduction

Pedestrian detection plays an important role in intelligent vehicle applications, as it is a crucial functionality for guaranteeing vehicle navigation safety in human-existing traffic environments. Pedestrian detection is a challenging task, due to a huge number of possibilities of pedestrian appearances and poses. In the intelligent vehicle field, pedestrian detection has already been incorporated into various aspects of applications, such as environmental perception [12], collision avoidance [6] and motion planning [1].

During past decades, numerous pedestrian detection algorithms have been proposed in literature. Since its birth 12 years ago, the Histogram-Of-Gradient (HOG) descriptor [2] has become a popular descriptor for pedestrian detection, thanks to its effectiveness in capturing implicit human characteristics.

As the HOG methodology is commonly followed, a question arises naturally: how can we find a HOG-style descriptor (i.e. an instantiation of the HOG methodology) that is optimal or at least semi-optimal (in certain sense)?

Before answering above question, we need to know how to systematically generate HOG-style descriptors that may serve as candidates for further evaluation. Inspired by works presented in [7][8][5][14][9][14], we propose a general model, coined as Descriptor Generation Model (DGM), which can be used to systematically generate HOG-style descriptors for pedestrian detection. The DGM can generate a wide range of HOG-style descriptors; for examples, existing HOG-style descriptors such as the original HOG [2] and the LBP [10] are among the descriptors that the DGM can construct.

Based on the DGM, we introduce a pedestrian detection experimental framework (PDEF) to find the optimal HOG-style descriptor. In the PDEF, the performance of each descriptor can be evaluated with appropriate criteria; the genetic algorithm [13] is employed to search the optimal (or semi-optimal) HOG-style descriptor in the descriptor space.

Based on the NICTA pedestrian dataset [11], the best descriptor that is found via our implemented PDEF is presented. This new descriptor, which is named Second-order Gradient for Pedestrian detection (G2P), is also a contribution of the works presented in this paper.

This paper is organized as follows: The Descriptor Generation Model (DGM) is detailed in Section II; the pedestrian detection experimental framework (PDEF) and the new descriptor G2P are presented in Section III; experimental results on the G2P are demonstrated in Section IV, followed by a conclusion in Section V.

II. Descriptor Generation Model

The DGM consists of the Filter module, the Repetition-Kernel module and the Histogram module, each of which

*Resrach supported by The National Natural Science Foundation of China (91420101).

contains various components. A component may further have various parameters. Fig.1 illustrates the block diagram of the DGM, its modules and their relationship. The green circles in the modules represent the components and parameters the module possesses.

The Filter module contains a group of image filters. The Repetition-Kernel module consists of two sub-modules: the Repetition sub-module and the Kernel sub-module. The Repetition sub-module provides some patterns to describe the positional relationship among image pixels. The Kernel sub-module provides a (large) number of ways to transform image blocks. The Histogram module is used to calculate the histograms of the image block and gather relevant statistics.

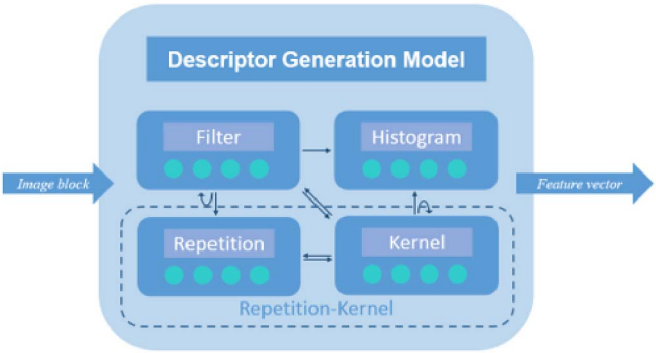

Fig. 1. The block diagram of the Descriptor Generation Model. The two arrows on either side of the fig denote the input and the output of the DGM. The rectangles denote different modules. Moreover, the circles in the rectangles mean different components in different module. The small arrows between the rectangles mean the connection between different modules.

A. Filter

Image filtering is a sort of standard image processing method; it has various functions, such as image smoothing, noise suppression, gradient extraction and so on.

The output of the Filter module is usually a single image, but in some cases can be two images. For example, the Sobel filter (a gradient extraction filter) takes the operator direction as its parameters; the operator direction may be set to vertical, horizontal and vertical-horizontal. If the operator direction is set to vertical-horizontal, then the Sobel filter computes gradients in both vertical and horizontal direction and output two images (a magnitude image and an angle image).

When the magnitude image and the angle image are both available as filter outputs, the magnitude image will be sent to the Repetition-Kernel module, whereas the angle image will be sent directly to the Histogram module (the Repetition-Kernel module and the Histogram module are detailed in following sub-sections).

Each of above gradient extraction filters except the Laplace filter has three filter parameters which are the directions of the filter kernels; i.e. vertical, horizontal and vertical-horizontal (both vertical and horizontal). The Laplace filter is a second-derivative-based edge detector, does not have any direction-related parameter.

B. Repetition-Kernel

The Repetition-Kernel module consists of two sub-modules i.e. the Repetition sub-module and the Kernel sub-module which always work together in cascade. The Repetition sub-module provides the position relationship and the Kernel sub-module provides the calculation method.

1) Repetition: The input of the Repetition sub-module can be an image block as well as a pixel. Given an input image (denote as I) whose height and width are H and W respectively, the preliminary step of the Repetition sub-module is to reshape the input image into a vector of $H * W$ elements (denote as $V_{reshaped}$) by concatenating the input image column vectors from left to right; as formalized in equation (1).

$$V_{reshaped} = [I(0,0)\ I(0,1)...I(0,W-1)\ I(1,0)\ I(1,1)\ ...\ I(1,W-1)\ .....\ I(H-1,0)\ I(H-1,1)\ ...\ I(H-1,W-1)] \quad (1)$$

In summary, the Repetition sub-module is a function that converts an image block to a repetition matrix. An example of the repetition process is illustrated in Fig.2.

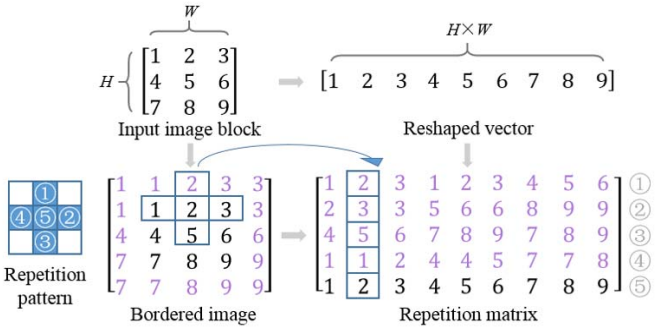

Fig. 2. The height and width of the input image block is H and W. The reshaped vector is to reshape the input image to a row vector whose length is H ∗ W. The bordered image can be got from adding border process. Then traverse all the pixels in input image. The relate pixels which are indicated in the repetition pattern are rearranged in column. Moreover, the order of them depends on the number in repetition pattern. Finally, lower-right corner of the fig demonstrates the output repetition matrix.

2) Kernel: As mentioned previously, the Repetition sub-module and the Kernel sub-module always work together in cascade. More specifically, the Kernel sub-module takes the output of the Repetition sub-module as its input and output a vector, as formalized compactly in equation (2).

$$R_{Kernel} = K(M_{Rep}) \quad (2)$$

Function $K()$ can be defined in various ways and can be linear or non-linear.

The examples in Fig.3 illustrate the relationship between the Repetition-Kernel module and the Filter module. Normally, the function of the Repetition-Kernel module is similar to that of the Filter module. Fig.3 shows how the Mean Filter and the Sobel Filter can be derived from repetition patterns and kernel functions.

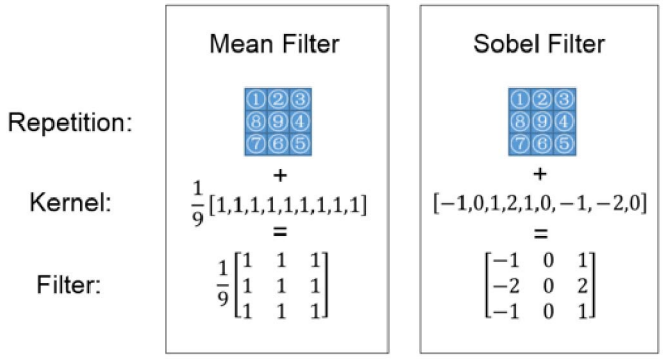

Fig. 3. Two examples to show the relationship between the Repetition and Kernel sub-module. The contents in two rectangles show how to get the mean filter and the Sobel filter through the combination of the Repetition and Kernel sub-module.

C. Histogram

The Histogram module is to get the statistics from the input image. This module will finally convert the input image block into a feature vector.

The histogram method presented in this paper is inspired by the HOG histogram method [2] which divides an image into windows, blocks and cells. Different block and cell size will have an influence on the process of the histogram extraction and the dimension of the output feature vector.

D. Examples

In this sub-section, we give two examples to demonstrate how the DGM constructs the HOG descriptor and the LBP descriptor.

To construct the HOG descriptor [2], the Filter and Histogram modules can be used. First, send the image block into HOG-style filter, which has vertical-horizontal parameter and output a magnitude image and an angle image. Then, send the two images to the Histogram module in which the parameters can be chosen by the user.

III. A NEW DESCRIPTOR FOR PEDESTRIAN

With different combination of modules, components and parameters, the DGM can construct a huge number of descriptors. In order to find the optimal (or semi-optimal) descriptor for pedestrian detection in DGM, we construct the pedestrian detection experimental framework (PDEF) in this section. In PDEF, the performance of each descriptor can be evaluated with appropriate criteria.

A. Genetic Algorithm

The DGM can potentially construct an infinite number of descriptors i.e. a descriptor space that cannot be exhaustively traversed in practical implementation. There are two important operations in the genetic algorithm, namely the crossover and mutation operations.

In the DGM, the crossover operation (if treated as a function) takes two descriptors as input and outputs a new descriptor. For example, the newly generated descriptor may be a combination of the Filter and Repetition-Kernel modules of an input descriptor and the Histogram module of the other descriptor.

Mutation is another important operation in the genetic algorithm. The implemented mutation operation can be carried out randomly in the following five ways:

- Add a new module to the input descriptor.
- Delete a module from the input descriptor.
- Change a module from the input descriptor.
- Change a component of a module of the input descriptor.
- Change a parameter of a module of the input descriptor.

A pedestrian detection experimental framework (PDEF) is proposed, based on the descriptor evaluation method introduced in previous sub-section and specially designed genetic algorithm introduced in this sub-section. The flow diagram of the genetic algorithm based PDEF in illustrated in Fig.4.

B. New Descriptor: G2P

The new descriptor, coined as Second-order Gradient for Pedestrian detection (G2P), can be extracted as follows.

First, performing the Gamma correction on the input image block. For the G2P, the Gamma parameter in the Gamma correction is set to 1/32.

Second, send the output of the Gamma correction to a Sobel filter with the horizontal.

Third, send the output of the Sobel filter to a Roberts filter with the parameter vertical-horizontal. This Roberts filter uses both filter kernels.

Fourth, compute the magnitude and angle images and then send these two images to the Histogram module to compute the gradient histogram. The parameters in the Histogram module are set as follows: the block size is set to 8*8; the cell size is set to 4*4; the bin number is set to 9; the range of orientation bins is set to 0-180 degrees (in other words, the orientation bins are unsigned), and the L2-Hys [9] is chosen as the normalization scheme.

The flow diagram in Fig.5 summarizes the procedures of extracting the G2P features.

IV. EXPERIMENTS

A. Datasets and Experimental Platform

Three datasets were used to evaluate the performance of the G2P descriptor. They are the ETH [3] pedestrian dataset, the CVC-02-system [4] pedestrian dataset, the NITCA [11] pedestrian dataset and KITTI dataset. The ETH and CVC-02-system datasets contain plenty of images captured by on-vehicle cameras in traffic environments.

We select the test videos recorded by on-vehicle cameras especially those recorded in intersection and zebra crossing scenarios. Two video sequences which have 1450 frames and contain 10395 pedestrians are chosen from the ETH dataset. In addition, five video sequences which have 1552 frames and contain 3262 pedestrians from the CVC-02-system are chosen, moreover, 6000 pedestrian samples and 24000 non-pedestrian samples are selected. The video sequences which are chosen from the ETH and CVC-02-system and the image samples

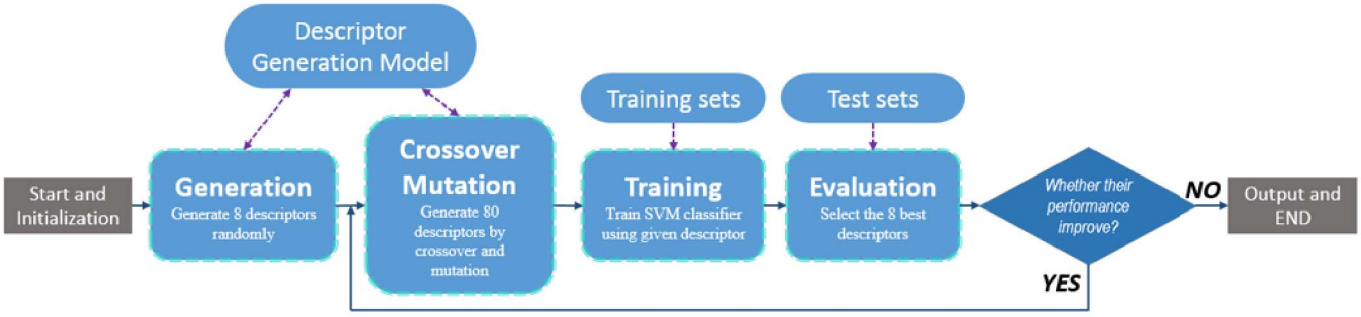

Fig. 4. The flow diagram of the genetic algorithm based PDEF. In the first step, the DGM, linear SVM classifier and all data are initialized. The second step is the Generation step, which the DGM generates 8 descriptors randomly. The third step is the Crossover and Mutation step which generates 80 descriptors by performing the crossover and mutation operations. The forth step is the Training step which trains the corresponding SVM detectors by the input descriptors. In the fifth step i.e. the Evaluation step, the performance of the trained SVM detectors are evaluated then the best 8 descriptors are selected. The whole algorithm will not stop until the best descriptors performance does not improve. At last, the algorithm will find and output the best 8 descriptors.

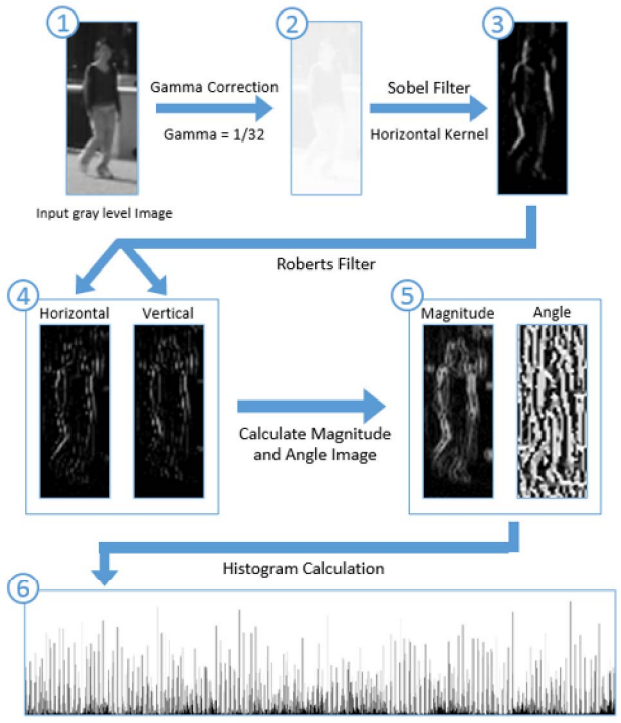

Fig. 5. The flow diagram of the calculation process of the G2P descriptor. 1. The input grey level image block. 2. The Gamma correction result. 3. The Sobel filter result with horizontal kernel. 4. The Roberts filter result with both horizontal and vertical kernel. 5. Calculate the magnitude and angle image. 6. Compute the histogram, then get the feature vector.

which are chosen from the NITCA and KITTI are used to evaluate the performance of the pedestrian detection system and compare the G2P and the HOG descriptors.

Processing of all the experimental data presented in this paper was performed by a computer with an i7-4702MQ CPU.

B. Experimental Results and Analysis

Based on the ETH, CVC-02-system, NITCA and KITTI pedestrian datasets, a pedestrian detector based on the linear SVM and the G2P descriptor was evaluated. The traditional linear SVM + HOG detector [2], which served as a baseline method for comparison, was also evaluated in the same conditions.

Fig.6 demonstrates the Receiver Operating Characteristic (ROC) and the Precision-Recall (PR) curves of the tests with the ETH, CVC-02-system and NITCA pedestrian datasets. Fig.7 demonstrates the Precision-Recall (PR) curves with KITTI dataset. Since the KITTI dataset is not used for training, the algorithm shows good robustness as it is shown in the Fig.7. Moreover, with single template, G2P has similar effect with DPM, which has multiple templates.

As demonstrated by Fig.6, the G2P descriptor outperformed the HOG in all the tests we carried out. The left two sub-figures demonstrate a considerable advantage of the G2P over the HOG in tests with the ETH dataset. On the other hand, the advantage of the G2P over the HOG was not obvious in tests with the CVC-02-system dataset. That is because the G2P descriptor is the descriptor which has optimal (or semi-optimal) pedestrian detection performance in PDEF, and the test set in PDEF is chosen from the ETH dataset. The tests with the NITCA dataset also demonstrate the advantage of the G2P descriptor.

Table I. lists the processing time per frame of the G2P-based method and the HOG-based method. The time spent by the G2P-based method to process one image frame is only few milliseconds more than that spent by the HOG-based method, yet the G2P-based method achieves apparently better performances than the HOG-based method.

Fig.8 demonstrates some experimental results of testing the linear SVM based G2P detector with the ETH and CVC-02-system datasets. The four sub-figures in the first and second rows show some detection results in tests with the ETH and CVC-02-system datasets respectively. On the contrary, the sub-figures in the last row in Fig.8 show some false positives.

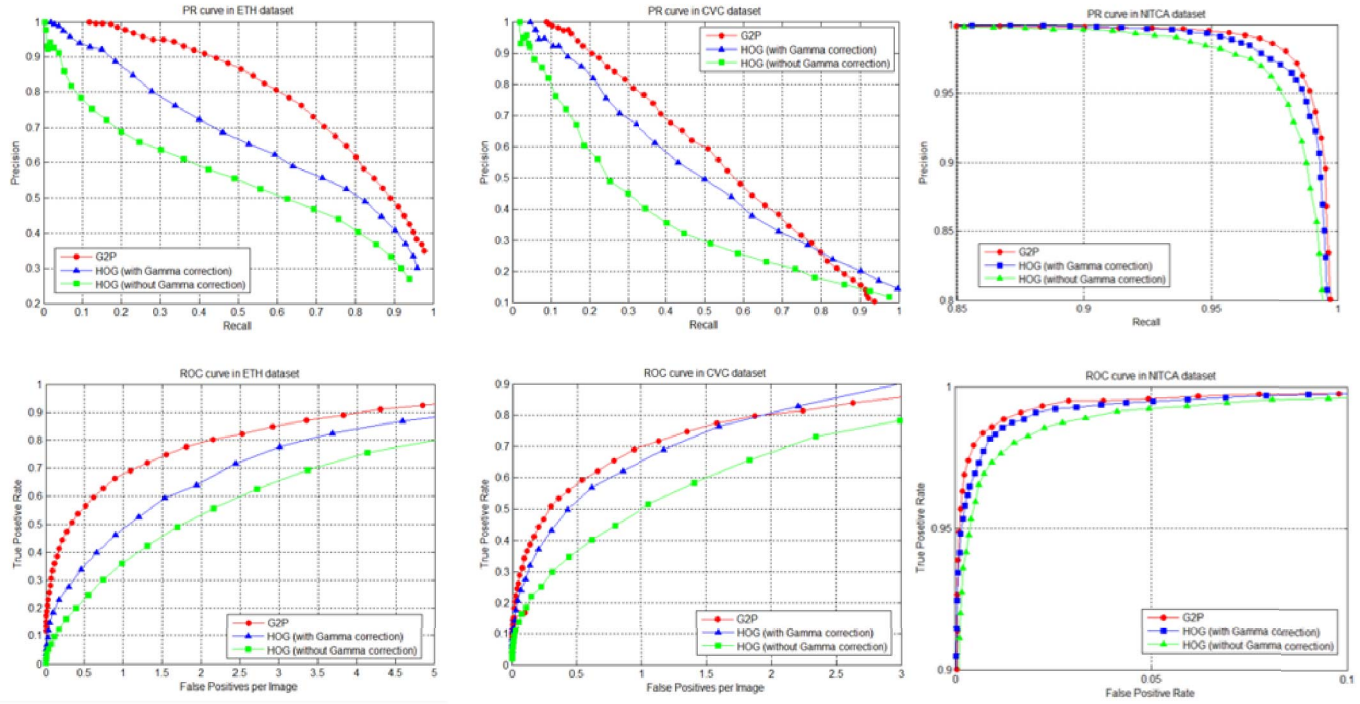

Fig. 6. The pedestrian detection experimental results of the G2P and the HOG. There are three kinds of descriptors involved in the experiments. The red line with circle dots represents the experimental results of the G2P descriptor, the blue line with triangular dots represents the experimental results of the HOG with Gamma correction and the green line with square dots represents the experimental results of the HOG without Gamma correction.

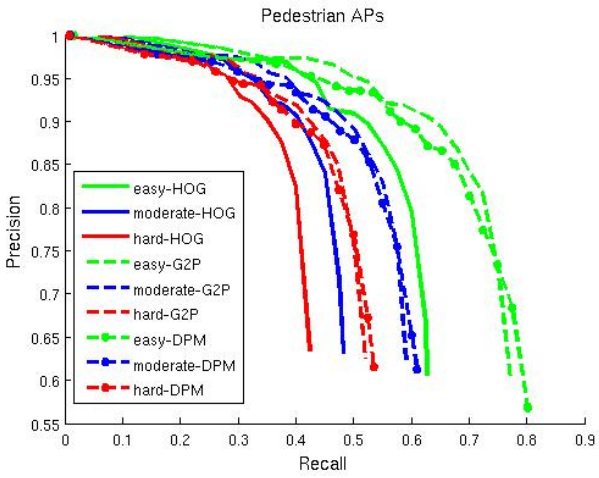

Fig. 7. The pedestrian detection experimental results of the G2P, HOG and DPM using KITTI dataset. Easy, moderate and hard categories are designed according to the KITTI standard.

Green bounding box represent true positives, whereas red boxes represent false positives.

V. CONCLUSION

In this paper, we have proposed a general model, coined as Descriptor Generation Model (DGM), which can be used to systematically generate a wide range of HOG-style descriptors for pedestrian detection. Based on the DGM, we have further introduced a pedestrian detection experimental framework (PDEF) to find the optimal (or semi-optimal) HOG-style descriptor. We have detailed our implementation of the PDEF and have also described a new descriptor that is found via our implemented PDEF; the newly found descriptor is named Second-order Gradient for Pedestrian detection (G2P).

TABLE I
REAL-TIME PERFORMANCE(DETECTION TIME PER FRAME)

ALGORITHM	ETH	CVC
G2P	262.772ms	261.497ms
HOG(with Gamma Correction)	255.916ms	257.627ms
HOG(without Gamma Correction)	257.952ms	257.382ms

REFERENCES

[1] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. Intention-aware motion planning. In *Algorithmic Foundations of Robotics X*, pages 475–491. Springer, 2013.

[2] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[3] Andreas Ess, Bastian Leibe, and Luc Van Gool. Depth and appearance for mobile scene analysis. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[4] David Gerónimo, Angel D Sappa, Daniel Ponsa, and Antonio M López. 2d–3d-based on-board pedestrian detection system. *Computer Vision and Image Understanding*, 114(5):583–595, 2010.

Fig. 8. Some pedestrian detection results in traffic environment from the ETH and CVC-02-system dataset. The sub-figures in the first row are the detection results in tests with the ETH dataset. The sub-figures in the second row are the detection results in tests with the CVC-02-system dataset. And the sub-figures in the last row show some false positives. The left two sub-figures in last row are from tests with the ETH dataset, and the right two are from tests with the CVC-02-system dataset. The green boxes denote true positives, the red boxes denotes false positives.

[5] Zhenyu He, Shuangyan Yi, Yiu-Ming Cheung, Xinge You, and Yuan Yan Tang. Robust object tracking via key patch sparse representation. *IEEE transactions on cybernetics*, 47(2):354–364, 2017.

[6] Christoph G Keller and Dariu M Gavrila. Will the pedestrian cross? a study on pedestrian path prediction. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):494–506, 2014.

[7] Henning Lategahn, Johannes Beck, Bernd Kitt, and Christoph Stiller. How to learn an illumination robust image feature for place recognition. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 285–291. IEEE, 2013.

[8] Henning Lategahn, Johannes Beck, and Christoph Stiller. Dird is an illumination robust descriptor. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 756–761. IEEE, 2014.

[9] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[10] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[11] Gary Overett, Lars Petersson, Nathan Brewer, Lars Andersson, and Niklas Pettersson. A new pedestrian dataset for supervised learning. In *Intelligent Vehicles Symposium, 2008 IEEE*, pages 373–378. IEEE, 2008.

[12] Antonio Prioletti, Andreas Møgelmose, Paolo Grisleri, Mohan Manubhai Trivedi, Alberto Broggi, and Thomas B Moeslund. Part-based pedestrian detection and feature-based tracking for driver assistance: real-time, robust algorithms, and evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):1346–1359, 2013.

[13] Darrell Whitley. A genetic algorithm tutorial. *Statistics and computing*, 4(2):65–85, 1994.

[14] Xinge You, Liang Du, Yiu-ming Cheung, and Qiuhui Chen. A blind watermarking scheme using new nontensor product wavelet filter banks. *IEEE Transactions on Image Processing*, 19(12):3271–3284, 2010.