Distant Pedestrian Re-Detection from an in-Vehicle Camera based on Detections by Other Vehicles

Yasutomo Kawanishi*, Daisuke Deguchi†, Ichiro Ide*, Hiroshi Murase* and Hironobu Fujiyoshi‡
*Graduate School of Information Science, Nagoya University, Nagoya, Aichi, Japan
Email: kawanishi@is.nagoya-u.ac.jp
†Information & Communications Headquarters, Nagoya University, Nagoya, Aichi, Japan
‡Department of Robotics Science and Technology, Chubu University, Kasugai, Aichi, Japan

***Abstract*—In this paper, we propose the re-detection paradigm, which is a detection with prior knowledge of the detection targets, and we introduce an implementation of the re-detection for distant pedestrian detection from an in-vehicle camera. We focus on the fact that other vehicles including forward vehicles can observe and detect pedestrians before the own vehicle observes them. Since appearances of pedestrians do not significantly change even though their locations are different, sharing images of the detected pedestrians among the vehicles, the own vehicle can use them as prior knowledge for detecting them again. Results of applying the proposed method to a dataset obtained by an in-vehicle camera demonstrate that the accuracy of pedestrian detection results can be significantly increased if prior knowledge of the pedestrians could be obtained.**

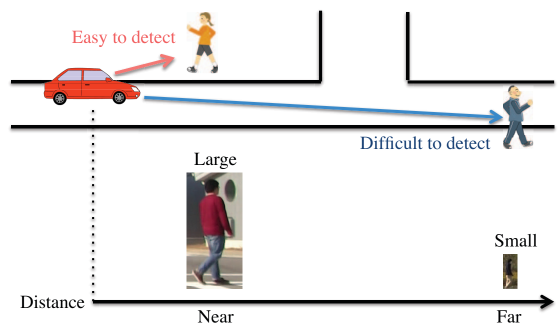


Fig. 1. Detecting distant pedestrians is difficult because they are too small in the observed image.

I. Introduction

In this paper, we address the problem of detecting distant pedestrians from an in-vehicle camera image.

Recently, technologies related to Intelligent Transportation Systems (ITS) have been actively developed. The main focus of the technologies is to reduce the risk of traffic accidents. Especially, since many pedestrians are killed in traffic accidents, their safety is one of the most important problems. To avoid traffic accidents involving pedestrians, the driver of a vehicle needs to find pedestrians as early as possible. To find pedestrians earlier, the driver needs to find distant pedestrians. However, finding distant pedestrians is a hard task for drivers while driving a vehicle.

To support the driver finding pedestrians earlier, various kinds of sensors, such as Radar, LIDAR and cameras are installed on the state-of-the-art vehicles. Many methods are proposed to detect pedestrians using one of these sensors or their fusions [1]–[4]. Radar / LIDAR are very effective to find obstacle on the road, including pedestrians. However, since their resolution is very low, they only work well within a very limited range of distance. Therefore, to detect distant pedestrians far from the vehicle, cameras are the best choice in terms of the resolution.

The sizes of pedestrians in an in-vehicle camera image depend on the distance to the pedestrians. Even though the resolution of a camera is relatively higher than other sensors, detecting distant pedestrians is very difficult because their sizes are very small (Fig. 1). This difficulty has been reported by Dollar *et al.* [2].

In general, it is difficult to detect distant pedestrians with no prior knowledge. However, if we have prior knowledge of the target pedestrians, can we detect them more easily? In this paper, we focus on the fact that other vehicles can observe and detect the pedestrians before the own vehicle observe them, and their appearance do not significantly change even though their locations are different. Here, other vehicles include forward vehicles of the own vehicle. We assume that vehicles can share the detection results of the others through a communication technology such as Vehicle-to-Vehicle communication. We introduce a pedestrian detection method utilizing the observation of other vehicles as prior knowledge to customize the pedestrian detector to the pedestrians who have been observed by other vehicles beforehand. The customized detector will detect the pedestrians even though they are in the distance. We named the detection paradigm which utilizes prior knowledge for detection targets to pull up the detection performance as "re-detection".

Our contributions are as follows:

- This paper proposes a new concept of the re-detection paradigm, which is a detection with prior knowledge.
- This paper presents an application of the re-detection for distant pedestrian detection from an in-vehicle camera with V2V communication.
- This paper proposes an implementation of a filtering based re-detection method.

II. Related Work

Various pedestrian detectors for a camera image have been proposed [2], [5]. Most of them use a pre-trained pedestrian

978-1-4673-6596-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ITSC.2015.200

classifier and they follow the exhaustive search with sliding windows approach; they scan each image with a sliding window and classify each sub-image cropped by a sliding window into positive or negative classes. Here, the positive class is a pedestrian and the negative class is a non-pedestrian. For the classification, to detect various pedestrians, it is preferred to use a feature which captures the common appearance of various pedestrians and ignore specific appearances of each pedestrian. Then, similar detection results are merged by clustering and output as the final detection result. To detect pedestrians in various sizes, many methods use the exhaustive search with multi-scale sliding windows strategy, which applies sliding window scans many times while chaining the window size from the smallest size to the largest size.

The most popular and practical pedestrian detection method is the Support Vector Machine (SVM) with the Histograms of Oriented Gradients (HOG) descriptor [6]. This method extracts an edge-based feature and classifies it with a discriminatively trained classifier. A HOG descriptor of a sub-image cropped by a sliding window is calculated by accumulating edge histograms in local regions of the sub-image. The descriptor can roughly describe the shape and the textures of the sub-image. Since the descriptor can capture the rough shape of a pedestrian, this method can detect pedestrians robust to slight differences of pedestrians. However, when the pose of pedestrians varies largely, it becomes difficult to detect them by this method.

To tackle this problem of large pose variation, Felzenszwalb *et al.* proposed a discriminative part based approach [7], [8]. It extracts HOG-based features for each body part and learn their relative positions by Latent SVM. Since this method extracts features of body parts, a relatively high resolution image is required. Therefore, it could not be applied to small pedestrians in the distance.

In recent years, Deep Learning based detection methods as represented by Joint Deep Learning have been proposed [9]–[12]. They simultaneously train feature extractor and classifier to extract more suitable features for pedestrian detection. These methods also require a relatively higher resolution image.

Recently, some detection methods not based on the exhaustive search with sliding windows approach has been proposed. One of the latest method of this approach is Regionlets [13], [14]. Instead of the sliding windows, these method extract region candidates using the selective search strategy [15]. Then they classify each region candidates into the pedestrian class or not. Since they avoid the sliding window search, the processing speed is relatively fast.

In nature, the methods based on the sliding window approach cannot detect pedestrians when they are smaller than the smallest sliding window. Selective search based methods also cannot detect them when they are smaller than the smallest region size. Therefore, to detect distant pedestrians who are very small in an image, the smallest sliding window size or the smallest region size should be small enough. When we scan an image with small sliding windows, since images cropped by the sliding window are very small, we could not extract information from them sufficient to accurately classify them. As a result, the pedestrian detector outputs many false positives.

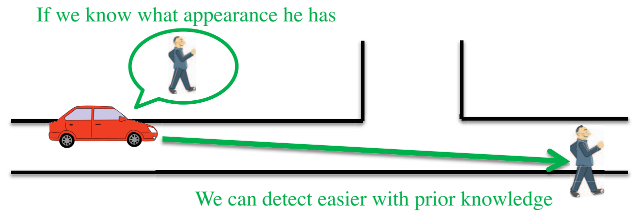


Fig. 2. Re-detection: Detecting pedestrians with prior knowledge.

III. Person Re-detection

A. Definition

We expect that pedestrian detection could become easier if we have prior knowledge on who there are, what clothes they wear, or what appearances they have. An example of an actual situation is shown in Fig. 2. If we can know beforehand that there is a person wearing blue in the distance, detecting the person should become easier even though they are very small in size. We define person "re-detection" as detecting pedestrians with prior knowledge obtained by observations from other cameras.

B. Re-detection from an in-Vehicle Camera

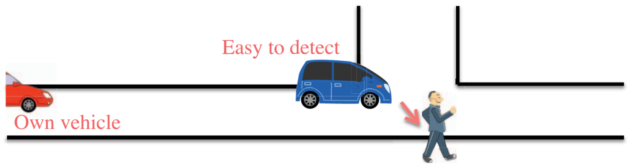
The problem which we tackle in this paper is pedestrian detection from an in-vehicle camera. Here, we adapt the re-detection paradigm to pedestrian detection from an in-vehicle camera.

In this paper, we focus on the fact that before the own vehicle observes pedestrians, other vehicles including forward vehicles and oncoming vehicles have a chance to observe them closely and detect them in high-resolution images. Several examples of such situations are shown in Fig. 3. In general, it is easy to detect pedestrians when they are close to the vehicle and observed largely, while it is very difficult to detect pedestrians when they are distant from the vehicle as shown in Fig. 1. We propose a method to utilize the high-resolution pedestrian images provided by other vehicles to pull up pedestrian detection performance of the own vehicle.

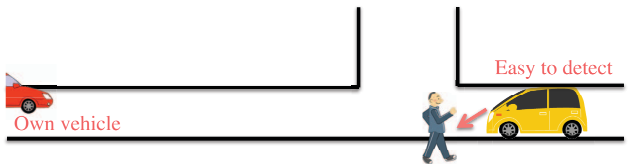
Recently, Vehicle-to-Vehicle (V2V) communication is actively researched in the ITS field. This technology is developed for various purposes, especially for cooperative driving such as collision avoidance of vehicles or auto-merging in a highway. In this paper, we assume that vehicles can take advantage of the technology, and transmit the detected pedestrian images to other vehicles. The situation that a forward vehicle transmits detected pedestrian images to its backward vehicle is illustrated in Fig. 4.

We assume that appearances of pedestrians do not significantly change even though their positions changed. Utilizing the high-resolution pedestrian images transmitted from the forward vehicles, we customize the pedestrian detector in the own vehicle to them. We expect that the customized pedestrian detector can find the pedestrians easier even though they are far from the vehicle.

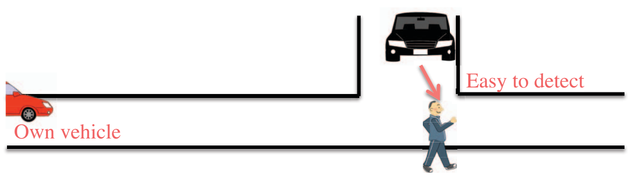
By combining the customized pedestrian detector with a general pedestrian detector, we can detect pedestrians close



(i) A situation that the pedestrian is captured by a forward vehicle.



(ii) A situation that the pedestrian is captured by an oncoming vehicle.



(iii) Another situation.

Fig. 3. Other vehicles have chances to observe and detect pedestrians before the own vehicle observe them. Since these vehicles can observe them closely, they can capture high-resolution images of the pedestrians.

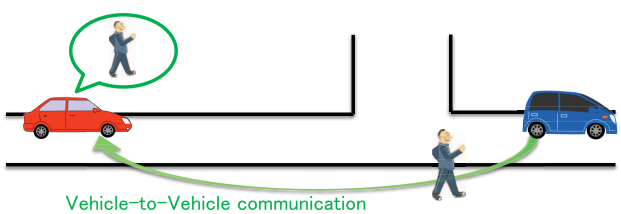


Fig. 4. Transmitting the detected pedestrian images from forward vehicles to a backward vehicle through V2V communication.

to the vehicle and additionally detect distant pedestrians if the pedestrians are captured by other vehicles. The overall scheme is illustrated in Fig. 5. In the following sections, we describe the implementation of the proposed method in detail.

IV. An Implementation of The Proposed Method

We propose an implementation of the pedestrian re-detection from an in-vehicle camera. Here, we describe the essential part of the system, which is emphasized with red color in Fig. 5. It is realized by the following procedure:

1) Receive high-resolution pedestrian images detected from other vehicles through V2V communication.
2) Detect pedestrian candidates with smaller sliding windows.
3) Filter the pedestrian candidates based on the prior knowledge obtained from the other vehicles.

The detailed process flow is illustrated in Fig. 6. In the following sections, D^p denotes the high resolution images

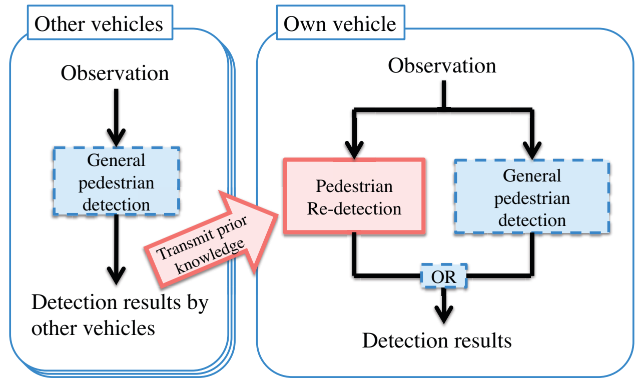


Fig. 5. Proposed pedestrian detection scheme.

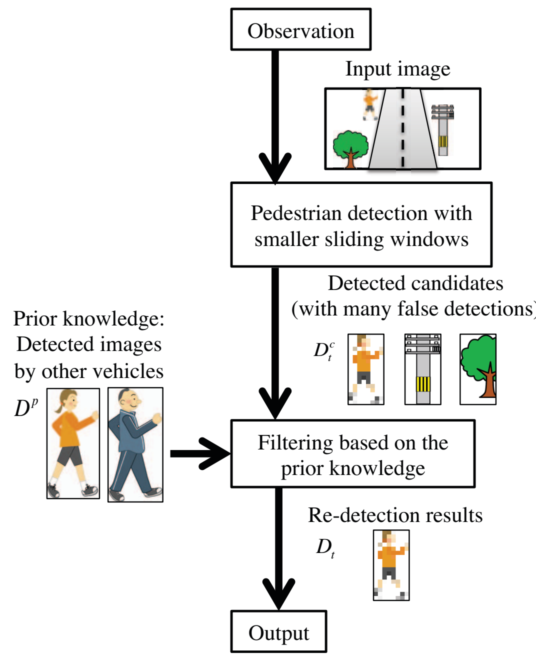


Fig. 6. Detailed process flow of the proposed method.

(prior knowledge shared with other vehicles), which consists of images of pedestrians P, D_t^c denotes the detection candidates for a frame t of the own vehicle, and D_t denotes the final detection results for the frame t.

A. Pedestrian Detection by Other Vehicles

We assume that when pedestrians close to the vehicle are observed, they could be easily detected and tracked for several frames. By tracking pedestrians P, we obtain a cropped image sequence s_i for each pedestrian $i \in P$. Since we consider that an image with higher resolution has much information on the pedestrian than an image with lower resolution, a pedestrian image with the highest resolution is selected for each image sequence. The set of high-resolution pedestrian images

$$D^p = \{r_i^p | r_i^p \text{ is the largest image in } s_i, \forall i \in P\}, \quad (1)$$

are shared within vehicles.

B. Detecting Pedestrian Candidates

Any kind of exhaustive search with sliding windows based or selective search based detector can be applied. To detect pedestrians distant from the vehicle, in this paper, we use an SVM based pedestrian detector and HOG descriptor [6] with small sliding windows. For an image I_t of the own vehicle at time t, detection candidates

$$D_t^c = \{r_1, \dots, r_{n_t}\}, \quad (2)$$

are obtained by the pedestrian detector. In this process, to reduce the number of miss-detections, the proposed method detects pedestrians with a lower threshold.

C. Filtering Pedestrian Candidates

Among the detected pedestrian candidates, there are many false detections. Therefore, the results should be filtered to suppress the false detections based on the prior knowledge.

Although most false positives have similar HOG features to that of pedestrians, their appearances can be different in terms of other features. Actually, in most cases, humans can distinguish them correctly. The similarities of the appearances of each detection candidate and each pedestrian image provided by the forward vehicles are calculated, and candidates whose similarities are smaller than a threshold τ are filtered out. Finally, the detection results are obtained;

$$D_t = \{r | f(r, D^p) \geq \tau, \forall r \in D_t^c\}, \quad (3)$$

where $f(r, D^p)$ is a function which returns the similarity between r and D^p, defined as

$$f(r, D^p) = \min_{r^p \in D^p} g(r, r^p). \quad (4)$$

Here, $g(r, r^p)$ returns the similarity of the image features of r and r^p.

In order to compare the provided high-resolution images and the detected low resolution images, color features can be a good choice since they are robust to the difference of image resolution. In person re-identification researches, many features including color features are used to compare pedestrians precisely [16]. These features can be also applicable for our purpose. Among them, Major Color Spectrum Histogram Representation (MCSHR) [17] is known as one of the most useful color representations to compare pedestrian images. Actually, the feature is also used in the field of object tracking within a video [18].

Here, we use MCSHR to filter out the false detections. To exclude background pixels, the left and the right 1/3 of the image are cropped. We define the function $g(r, r^p)$ as the similarity between MCSHR of the detected image r and the provided image r^p.

V. Evaluation and Discussion

A. Evaluation Settings

To show the effectiveness of the proposed method, especially the use of prior knowledge for filtering out false detections, we performed an evaluation on a dataset.

A Canon G20 video camera was prepared as an in-vehicle camera, and installed in a car to gather images for the dataset.

Fig. 7. Samples from the dataset. The ground truth pedestrians are bounded by blue rectangles.

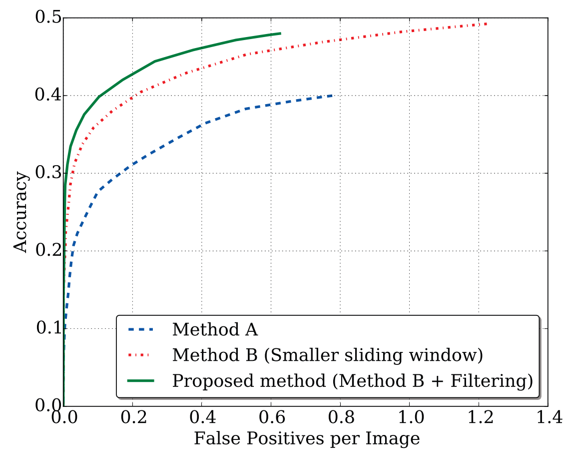


Fig. 8. FROC curves of pedestrian detection.

The resolution of the camera was Full-HD ($1,920 \times 1,080$ pixels) and the frame rate was 10 fps. The dataset consists of four image sequences captured by the in-vehicle camera. Each sequence was taken along the same route, contained about 500 images with two to four pedestrians. All of the pedestrians were manually annotated, which resulted in approximately 1,000 true positives in each sequence, and more than 4,000 true positives in total. A sample of the dataset is shown in Fig. 7.

To simplify the evaluation, here we assumed that all pedestrians had been detected accurately by other vehicles. We selected the largest image for each pedestrian in the sequences and used them as the detected high-resolution images by other vehicles.

For comparison, we prepared the following three pedestrian detection methods:

- Method A: HOG feature with normal window size (minimum 64×128 pixels).
- Method B: HOG feature with smaller window size (minimum 48×96 pixels).

- Proposed method: HOG feature with smaller window size (minimum 48×96 pixels) for detection and MCSHR feature for filtering.

All of them used an SVM-based classifier with the sliding window approach. The pedestrian detectors were trained using training samples provided in the Daimler Pedestrian Detection Benchmark Dataset [19]. Additionally, for each sequence of the dataset, we collected false detections of the detectors from the rest of the sequences and added them into negative samples. The detectors were re-trained several times until they converged for each sequence.

A detected bounding box b^d was considered as positive when it overlapped with the corresponding ground truth b^g sufficiently. As to follow in the popular evaluation method [2], we evaluated each detected bounding box using the following equation:

$$\frac{a(b^d \cap b^g)}{a(b^d \cup b^g)} > 0.5, \tag{5}$$

where $a(b)$ is a function which returns the number of pixels in area b.

B. Results

The results are visualized by Free-response Receiver Operating Characteristic (FROC) curves in Fig. 8. FROC is usually used to compare object detection results, where its horizontal axis shows false positives per image and its vertical axis shows the detection accuracy.

We can see that method B showed higher accuracy than method A. It is because method B could detect smaller pedestrians. The proposed method showed higher accuracy than method A, and at the same time, the proposed method suppressed false positives by almost half of method B.

Examples of the detection result are also shown in Fig. 9, where the proposed method detected a small pedestrian that method A could not detect, and suppressed false positives. The results of method A, method B and the proposed method are shown in Fig. 9 (i), (ii), and (iii) respectively. In Fig. 9, detection results of each method are indicated in green rectangles. The images in Fig. 9 are cropped and magnified for visualization.

(i) Detection results by method A.

(ii) Detection results by method B.

(iii) Detection results by the proposed method.

Fig. 9. Examples of the detection results. Detection results of each method are indicated in green rectangles.

C. Discussion

In this evaluation, the proposed method could detect smaller pedestrians than the other methods with smaller false positives, however, as we can see in Fig. 9, even the proposed method could not detect the smallest pedestrian in the examples. It is because the proposed method used the HOG feature with smaller sliding window and the detected pedestrian candidates did not include all pedestrians. We need to modify how to detect pedestrian candidates further to cover all the pedestrians in an image.

We assumed that all pedestrians had been detected by other vehicles in this evaluation. However, such assumption does not necessarily hold true in real situations. Some of the pedestrians detected by the forward vehicles may disappear and new pedestrians may appear. When new pedestrians appear, the proposed method will filter them out because they are not included in the prior knowledge. However, we can detect them by combining a general pedestrian detector as shown in Fig. 5 after approaching them.

VI. Conclusion

In this paper, we proposed a pedestrian detection paradigm named person "re-detection", which utilize higher-resolution pedestrian images detected by forward vehicles as prior knowledge. We introduced an implementation of the "re-detection" paradigm for pedestrian detection from an in-vehicle camera by detecting pedestrians with small sliding windows and filtering the detection results comparing to the images provided by the forward vehicles. In the evaluation, we confirmed that the proposed method shows higher accuracy and lower false positives per image than a general pedestrian detection method.

Since the introduced detection method is a filtering based method, when the pedestrian detector with small sliding windows fail to detect small pedestrians, it could not find the pedestrians. For future work, we need to extend the method to find smaller pedestrians who cannot be detected by the pedestrian detector with small sliding windows. Selective search based method can be a candidate for this purpose. Further evaluation on public datasets such as Caltech Pedestrian Detection Benchmark [2] can also be a future work.

ACKNOWLEDGEMENT

Parts of this research were supported by JST, Nagoya University COI and MEXT, Grant-in-Aid for Scientific Research.

REFERENCES

[1] T. Ogawa, H. Sakai, Y. Suzuki, K. Takagi, and K. Morikawa, "Pedestrian detection and tracking using in-vehicle LIDAR for automotive application," in *Proc. 2011 IEEE Intelligent Vehicles Symposium*, June 2011, pp. 734–739.

[2] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, April 2012.

[3] L. Oliveira and U. Nunes, "Pedestrian detection based on LIDAR-driven sliding window and relational parts-based detection," in *Proc. 2013 IEEE Intelligent Vehicles Symposium*, June 2013, pp. 328–333.

[4] C. Premebida, J. Carreira, J. Batista, and U. Nunes, "Pedestrian detection combining rgb and dense LIDAR data," in *Proc. 2014 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS 2014)*, Sept. 2014, pp. 4112–4117.

[5] R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Ten years of pedestrian detection, what have we learned?" in *Proc. 2nd workshop on computer vision for road scene understanding and autonomous driving*, 2014.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 886–893 vol. 1.

[7] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. 2008 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[8] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, Sept. 2010.

[9] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in *Proc. 2013 IEEE Intl. Conf. on Computer Vision*, Dec. 2013, pp. 2056–2063.

[10] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning," in *Proc. 2013 IEEE Conf. on Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 3626–3633.

[11] X. Chen, P. Wei, W. Ke, Q. Ye, and J. Jiao, "Pedestrian detection with deep convolutional neural network," in *Proc. 2014 IEEE Intl. Workshop on Deep Learning on Visual Data*, 2014.

[12] J. Hosang, M. Omran, R. Benenson, and B. Schiele, "Taking a deeper look at pedestrians," in *Proc. 2015 IEEE Conf. on Computer Vision and Pattern Recognition*, 2015.

[13] X. Wang, M. Yang, S. Zhu, and Y. Lin, "Regionlets for generic object detection," in *Proc. 2013 Intl. Conf. on Computer Vision*, 2013, pp. 17–24.

[14] C. Long, X. Wang, G. Hua, M. Yang, and Y. Lin, "Accurate object detection with location relaxation and regionlets relocalization," in *Proc. the 12th Asian Conf. on Computer Vision*, 2014.

[15] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders, "Segmentation as selective search for object recognition," in *Proc. 2011 IEEE Intl. Conf. on Computer Vision*. IEEE, 2011, pp. 1879–1886.

[16] S. Gong, M. Cristani, S. Yan, and C. C. Loy, *Person re-identification*. Springer, 2014.

[17] C. Madden, E. D. Cheng, and M. Piccardi, "Tracking people across disjoint camera views by an illumination-tolerant appearance representation," *Proc. 2007 IAPR Conf. on Machine Vision and Applications*, vol. 18, no. 3-4, pp. 233–247, 2007.

[18] M. Piccardi and E. D. Cheng, "Multi-frame moving object track matching based on an incremental major color spectrum histogram matching algorithm," in *Proc. Workshop in 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. IEEE, 2005, pp. 19–24.

[19] M. Enzweiler and D. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.