

# Real Time Pedestrian Detection-based Faster HOG/DPM and Deep Learning Approaches

R. Khemmar  
Normandie Univ, UNIROUEN,  
ESIGELEC, IRSEEM,  
76000 Rouen, France

Li. Delong  
Normandie Univ, UNIROUEN,  
ESIGELEC, IRSEEM,  
76000 Rouen, France

B. Decoux  
Normandie Univ, UNIROUEN,  
ESIGELEC, IRSEEM,  
76000 Rouen, France

## ABSTRACT

The work presented aims to show the feasibility of scientific and technological concepts in embedded vision dedicated to the extraction of image characteristics allowing the detection and the recognition/localization of objects. Object and pedestrian detection are carried out by two methods: 1. Classical image processing approach, which are improved with Histogram Oriented Gradient (HOG) and Deformable Part Model (DPM) based detection and pattern recognition. We present how we have improved the HOG/DPM approach to make pedestrian detection as a real time task by reducing calculation time. The developed approach allows us not only a pedestrian detection but also calculates the distance between pedestrians and vehicle. 2. Pedestrian detection based Artificial Intelligence (AI) approaches such as Deep Learning (DL). This work has first been validated on a closed circuit and subsequently under real traffic conditions through mobile platforms (mobile robot, drone and vehicles). Several tests have been carried out in the city centre of Rouen in order to validate the platform developed.

## Keywords

Pedestrian detection, pattern recognition, object detection, HOG, DPM, ADAS Systems, tracking, deep learning.

## 1. INTRODUCTION

The work presented in this paper is a part of M2SiNUM<sup>1</sup> project (Advanced Mathematical Modeling and NUMerical Simulation for Innovation in Environment and Health), which is dedicated to the development of generic platform with tools accessible to scientific community. One of the objective of the project is to develop an Advanced Driver Assistance System (ADAS) platform for pedestrian detection and tracking for application of road security and autonomous vehicle. Other objectives of the project are related to advances both in fundamental and technological research. The practical applications are clearly identified and the expected consequences for the project fall within: industrial applications, academic results and national/international developments. The work presented in this paper is dedicated to pedestrian detection in road traffic applications. In general, ADAS is used to improve safety and comfort in vehicles. ADAS is based on the combination of sensors (radar, LIDAR, cameras, etc.) and algorithms that ensure vehicle, driver, passenger's and pedestrian's safety [26]. Here in this project, ADAS aims to detect pedestrian. Our contribution aims on the development of perception system based object detection with

different approaches such as HOG, DPM, and DL. This paper is organized as follows: Section 1 introduces the motivation of the paper. Section 2 presents the state of the art of pedestrian detection. Section 3 presents the architecture of the system through the different algorithms used and the results obtained. Section 4 illustrates our pedestrian detection algorithm-based Faster-HOG/DPM in which we present not only how to improve the quality of HOG/DPM pedestrian detection but also how to estimate the distance separating pedestrians to vehicles. Results are presented in Section 4 too. Finally, in Section 5, we will conclude this paper.

## 2. STATE OF THE ART AND RELATED WORK

### 2.1 State of the Art

Object detection is key problem in computer vision with several applications like automotive, manufacturing industry, mobile robotics, assisted living, etc. Pedestrian detection is a particular case of object detection that improve security road and is considered as an important ADAS in the autonomous vehicle. In [13], a very good state of the art for pedestrian detection is presented. Detection is disappointing at low resolutions and for partially occluded pedestrians. In [1], a real time pedestrian detection with DPM algorithm applied to automotive application is presented. The system present superior detection performance based on multiresolution pedestrian model in order to detect small pixel-sized pedestrians when classical DPM approach missed them. The system was evaluated under Caltech Pedestrian benchmark [2]. The DPM is presented as an algorithm between generative and discriminative model [7, 28]. The algorithm has different steps: build a pyramid for any scale under given image, use different filter and part filters to get responses. The algorithm combines these different responses in star-like model, uses a cost function, and train classifiers by SVM. This algorithm is still a widely used algorithm particularly in combination with DPM [8]. As another method of object detection, the Integral Channel Features (ICF) [1], we can find a combination of multiple registered image channels, which are computed by linear and nonlinear transformations [9]. Integration of some features like HOG and training by AdaBoost in cascade way can lead to pedestrian detection with good accuracy [9]. The sliding window methods (pyramid method) are used in object detection with a high cost of detection time. In a recent work, proposing high-recall important regions is widely used [10]. In another way, the approaches based on artificial intelligence based deep learning like Convolutional Neural Network (CNN) become very successful methods for features extraction in image classification task [11]. We can give as an example the Rich Feature Hierarchies for CNN (RCNN) paper [12], which combine CNN and selective search. Next session presents some project in relation with these algorithms.

<sup>1</sup> This project was co-financed by the European Union with the European regional development fund (ERDF, 18P03390/18E01750/18P02733) and by the Haute-Normandie Regional Council via the M2SiNUM project.

## 2.2 Related Work

In the literature, for similar tasks, several approaches were applied based on object detection and pattern recognition such as HOG/ DPM, KLT/RMR and Deep Learning [1]. For example, in Google Robot's Project [17, 21], we can find a deep learning model applied to articulated robot for how to pick things up. In Kitti Vision Benchmark Suite Project [18], an object detection and orientation estimation benchmark is carried out. In [19], as an example of end-to-end object detection from Microsoft is shown. For the task of detecting objects in images, recent methods based on CNN like SSD [18, 20] allows multiple objects detection. In many applications of object detection, it is important to estimate their 3D orientation. In outdoor environments, solutions based on DL has been recently shown to outperform other monocular approaches for detecting cars and estimating their 3D orientation [22,23]. In indoor environments, it has been shown that using synthetic 3D models of objects to be detected in the learning process of a CNN can allow simplifying it [23]. In [24,25], we can find an evaluation of the state of the art based object (pedestrian) detection which uses HOG/DPM approaches.

## 3. OBJECT DETECTION ALGORITHMS

### 3.1 Introduction

In order to identify the most suitable approach to our object/pedestrian detection application for autonomous vehicle, we need to make the feasibility of several scientific concepts in pattern recognition such as the KLT/RMR, SIFT/SURF, HOG/DPM but especially recent approaches related to the AI such as DL. In this paper, we will focus on pedestrian detection-based HOG/DPM approaches and how to improve them.

### 3.2 Classical Approaches

We started our experimentation by implementing point of interest approaches like Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) developed by D. Lowe [14]. The method makes it possible to transform visual features, which are invariant to scale, rotation and illumination. Despite their robustness to perspectives changing and lighting conditions, we found that the approach is not robust for pedestrian detection. SURF is better up to twice as fast as SIFT. However, SIFT is better when scale changing or an increase in lighting is applied to the image. For pedestrian detection, SIFT and SURF therefore remain insufficient. We have experimented Kanade-Lucas Tomasi (KLT) approach [15] for extraction of points of interest and tracking between an image taken at  $t-1$  and image taken at  $t$ . The approach is highly accurate and is fast but not robust to perturbations. This is why we have decide to experiment the Robust Multiresolution Estimation of Parametric Motion Models (RMR) approach [16], which has very low precision but is considered as very robust. We have obtained the same results as with KLT approach. A better solution could be a hybrid approach KLT/RMR in order to have the benefits of each approach.

### 3.3 Pattern Recognition-based HOG/DPM

The pattern recognition algorithms calculate the image features. They use classifiers under databases of positive images (with pedestrian) and negative images (without pedestrian). They have the advantage of being accurate and gives relatively good results; however, the calculating time is high. We therefore experimented three approaches: HAAR,

HOG and DPM:

- **Pseudo-HAAR Features.** HAAR classifiers use pseudo-HAAR features [27]. Instead of using pixel intensity values, the pseudo-HAAR features use the contrast variation between rectangular and adjacent pixel clusters. To construct a pseudo-HAAR feature, two or three adjacent clusters with relative contrast variation are required. It is possible to change the size of features by increasing or decreasing the pixel clusters. Thus, it makes possible the use of these features on objects with different sizes.
- **HOG Algorithm.** HOG is a descriptor containing image key features. These features are represented by the distribution of image gradient directions. HOG is the best-known algorithm in the field of pattern recognition. It consists of five steps: 1. Sizing the calculation window, 2. Calculation of image gradients using simple masks, 3. Image division of 64x128 into 8x8 cells. For each cell, HOG algorithm calculates the histogram, 4. Normalization of histograms by 16x16 blocks, and 5. Calculation of the size of the final descriptor. This descriptor is then given to SVM classifier.
- **DPM Algorithm.** The hardest part for object detection is that there are many variance, which could arise from illumination, viewpoint, non-rigid deformation, occlusion, and intra-class variability [4]. The DPM is trying to capture those variances. The detector will first find a match by coarser root filter, and then using its part models to fine-tune the result. Recently, new algorithms have been developed in order to make DPM Faster and more efficient [4, 8].

### 3.4 Pedestrian Detection-based HAAR/HOG/DPM

We carried out a wide range of experiments under six different databases dedicated to the pedestrian detection: ETH, INRIA, TUD Brussels, Caltech, Daimler and our own ESIGELEC dataset dedicated to the pedestrian detection. The Figure 1 shows examples of results obtained under the 3 datasets.



Fig 1. Comparison between HAAR (left), HOG (middle) and DPM (right) applied under three different datasets: TUD-Brussels (top row), ETH (middle row) with 640x480 image resolution, and ESIGELEC (bottom row) with 960x540 image resolution.

Table 1 gives the comparative study.

**Table 1. HAAR/HOG/DPM pedestrian detection.**

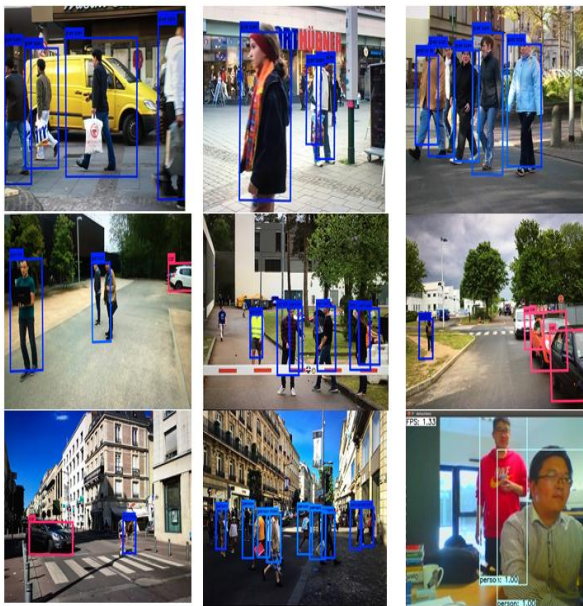
Dataset	HAAR	HOG	DPM
TUD-Brussels	Computing time: 51ms, quality detection: +	Computing time: 154ms, quality detection: ++	Computing time: 308ms, quality detection: +++
ETH	Time: 37ms, detection: +	Time: 108ms, detection: ++	Time: 247ms, detection: +++
ESIGELEC	Time: 50ms, detection: +	Time: 197ms, detection: ++	Time: 342ms, detection: +++

### 3.5 Pedestrian Detection-based Deep Learning

**3.5.1 Pattern Recognition-based Deep Learning**  
Recently, many detection algorithms using selective search [4,5] as region proposals, have been used to avoid exhaustive sliding window method. In addition, Deep Learning (CNN) [4,6] has proved to be successful in ImageNet classification challenge. Since the progress of DPM model is now restrictive (i.e. next session), Deep learning is trying to do detection in different way. We have implanted You Only Look Once (YOLO) as a real time object detection [14]. Prior detection, systems repurpose classifiers or localizers to perform detection. They apply the model eto an image at multiple locations and scales. High scoring regions of the image are considered detections [4,6].

#### 3.5.2 Experiment Results

The whole of experiments has been carried out under six datasets. Constrained by pages limit number, we present results obtained on two datasets: Caltech and ESIGELEC. The object detection rate used is calculated as Intersection over Union Ratio (IoU). The IoU represents the relationship between the area of overlap and union of ground truth and predicted bounding box [29]. The results have been obtained under GPU Nvidia Quadro K4000M Machine. Figure 2 shows some results obtained under YOLO and SSD algorithm.



**Fig 2. Pedestrian detection-based deep learning (YOLO/SSD).**

## 4. PEDESTRIAN DETECTION AND

## TRACKING BASED FASTER HOG/DPM ALGORITHM

### 4.1 Improved HOG Algorithm

We have identified four methods to improve HOG/DPM algorithm, which corresponds to the five algorithm HOG steps. In order to be able to make comparison, a classifier was trained from INRIA and TUD-Brussels dataset images. The number of available images is low which has an impact on the pedestrian detection quality. The objective is to improve the quality of HOG/DPM pedestrian detection and to reduce, significantly, the computation time for pedestrian detection. The tests were carried out under the ESIGELEC dataset:

1. Gamma Correction: in order to improve the pedestrian image quality, we performed a gamma processing on the images to be trained and tested. We found fewer parasites in the processed images.
2. Image Resizing: HOG algorithm uses in its calculations images of 64x128 sizes. We performed calculation with windows of 128x256. By doubling the size of the calculation window, we have improved the accuracy but also doubled the computation time (for example from 58ms in the classic HOG to 115ms for the Faster-HOG).
3. Negative Gradients: when calculating gradients, Faster-HOG uses signed gradients (0 to 180°) and unsigned gradient (0 to 180°) like classic HOG. This allows the calculation of histogram with 18 values (9 values in classic HOG). The calculation time does not vary, however, by taking into account, the negative gradients, and we found a presence of parasites, which does not improve the classic HOG.
4. Normalization Vector: as a last step, HOG performs a standardization of 2x2 cells, i.e. 32x32 pixels with a pitch of 16x16. The object detection is degraded and the computation time doubles, which cannot be considered as a classic HOG improvement.

### 4.2 Faster DPM Algorithm

The DPM algorithm is applied on the entire image and this is done at each iteration of the pedestrian video sequence. In order to reduce the computing time, the idea is to apply DPM algorithm only in a Region of Interest (RoI) in which the target is located. This will drastically reduce the computing time and will also isolate the object. Firstly, the DPM is applied all over the image once to locate the object. Secondly, and after obtaining a RoI surrounding the object to be detected, we built a New RoI (NRoI) by providing a tolerance zone (a new rectangle which is larger than the first one). Starting from the second iteration, DPM algorithm is applied only in this new image, which is represented by NRoI. If the target to be detected is lost, the DPM algorithm is re-applied over the entire image.

### 4.3 Pedestrian Distance Estimated

For better detection of pedestrian, it is necessary to estimate their distance from the vehicle. Our system is based on monocular camera, which makes it very difficult to measure the distance. The law called “Inverse Squares” [29], used in astronomy to calculate the distance between stars, inspired us: “physical quantity (energy, force, etc.) is inversely proportional to the square of the distance of stars”. By analogy, the physical quantity represents the area of our RoI. We have used a parametric approach. Measurements, taken at an interval of 50cm, were taken and the object RoI surface was recorded at each time.

## 4.4 Experiment Results

We have carried out several tests under 6 datasets dedicated to pedestrian detection. In this section, we present the tests carried out on the ESIGELEC dataset including tests scenarios under real traffic conditions in the shopping center car park and Rouen city center. The results obtained are shown in Figure 3 in which we can find the distance estimated on the top of each pedestrian box.



**Fig. 3. Pedestrian detection with Faster-DPM under ESIGELEC Dataset. Top row: ESIGELEC and Rouen City center, Middle row: Rouen mall center car park, Bottom row: Rouen city center pedestrian detection.**

## 5. CONCLUSION

In this paper, we have presented a comparison study between different approach dedicated to pattern recognition and object detection. We have presented our contribution for improving both HOG and DPM algorithms. We have also implemented DL based Yolo and SSD algorithm for pedestrian detection. An innovative method for distance calculation of pedestrians is presented in this paper. The approach was validated within experiments and the accuracy is good with only 5% of errors (see [30] for more details). The system detects and recognize not only pedestrian but also cars. The system is generic so that it is "applicable" on any type of platform and environment. This work aims to show the feasibility of scientific and technological concepts that affect HOG/DPM based detection and pattern recognition, but also AI approaches such as DL (like YOLO and SSD). We have validated the developments with a vehicle in real traffic conditions (Mall Center Car Park, Rouen City Center).

## 6. ACKNOWLEDGEMENT

This project was co-financed by the European Union with the European regional development fund (ERDF, 18P03390/18E01750/18P02733) and by the Haute-Normandie Regional Council via the M2SINUM project. Many thanks to SEGULA Engineering France company for his help in the development of this work. Many thanks to Technological Resources Center (TRC) of ESIGELEC/IRSEEM for his help in testing phase. his work was performed in part on computing resources provided by CRIANN (Centre Régional Informatique et d'Applications Numériques de Normandie, Normandy, France).

## 7. REFERENCES

[1]. H. Cho, P. E. Rybski, A. B-Hillel, W. Zheng.: Real-time

Pedestrian Detection with Deformable Part Models.

- [2]. Caltech Pedestrian Detection Benchmark Homepage, [http://www.vision.caltech.edu/Image\\_Datasets/CaltechPedestrians/](http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/), last accessed 2018/01/14.
- [3]. Felzenszwalb, P., McAllester, D., & Ramanan, D. A discriminatively trained, multiscale, deformable part model. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1-8). IEEE (June 2008).
- [4]. Jong-Chyi Su.: State of the Art object Detection Algorithms. University of California, San Diego, 9500 Gilman Dr. La Jolla, CA.
- [5]. Koen E. A. van de Sande, Jasper R. R. Uijlings, Theo Gevers, Arnold W. M. Smeulders, Segmentation As Selective Search for Object Recognition, ICCV, 2011.
- [6]. Alex Krizhevsky , Ilya Sutskever , Geoffrey E. Hinton. Imagenet classification with deep convolutional neuralnetworks, , NIPS, 2012.
- [7]. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR, 2014.
- [8]. Piotr Dollar, Christian Wojek, Bernt Schiele.: Pedestrian Detection: An Evaluation of the State of the Art. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 34, Issue: 4, April 2012.
- [9]. D. G. Lowe. Distinct Image Features from Scale-Invariant Keypoints. Computer Science Department. University of British Columbia. Vancouver, B. C. Canada. January, 5, 2004 28
- [10]. Kanade-Lucas-Tomasi. KLT Feature Tracker. Computer Vision Lab. Jae Kyu Suhr. Computer Vision (EEE6503) Fall 2009, Yonsei Uni.
- [11]. J. M. Odobez and P. Bouthemy. Robust Multiresolution Estimation of Parametric Motion Models. IRIS/INRIA Rennes, Campus de Beaulieu, February, 13, 1995.
- [12]. Object Detection Homepage, <http://cseweb.ucsd.edu/~jcsu/reports/ObjectDetection.pdf> , last accessed 2018/01/14.
- [13]. Yan, J., Lei, Z., Wen, L., & Li, S. Z. (2014). The fastest deformable part model for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2497-2504).
- [14]. Yolo Homepage, <https://pjreddie.com/darknet/yolo/>, last accessed 2018/01/19.
- [15]. Koen E. A. van de Sande, Jasper R. R. Uijlings, Theo Gevers, Arnold W. M. Smeulders, Segmentation As Selective Search for Object Recognition, ICCV, 2011.
- [16]. Alex Krizhevsky , Ilya Sutskever , Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. NIPS, 2012.
- [17]. Dave Gershgon, Google's Robot Are Learning How to Pick Things Up. Popular Science, March 8, 2016.
- [18]. Andreas Geiger, Philip Lenz, Christoph Stiller, Raquel Urtasun. Object Detection Evaluation. The Kitti Vision Benchmark Suite. Karlsruhe Institute of Technology.
- [19]. Eddie Forson. Understanding SSD MultiBox-Real Time Object Detection in Deep Learning. Towards Data

Science.

- [20].Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. C., (2016, déc.). Single-Shot Multibox Detector. <https://arxiv.org/abs/1512.02325>.
- [21].Levine S., Pastor P., Krizhevsky A., Ibarz J., Quillen D. (2017, June). Learning Hand-Eye Coordination for Robotic Grasping with DeepLearning and Large-Scale Data Collection. *The International Journal of Robotics Research*.
- [22].Chabot F., Chaouch M., Rabarisoa J., Teulière C., Château T.,. (2017, July). Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image. *IEEE Conference on Computer Vision and Pattern Recognition*.
- [23].Mousavian A., Anguelov D., Flynn J., Kosecka J. (2017, July). 3D Bounding Box Estimation Using Deep Learning and Geometry. *IEEE Conference on Computer Vision and Pattern Recognition*.
- [24].Piotr Doll, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [25].Junjie Yan, Xucong Zhang, Zhen Lei, Shengcai Liao, Stan Z. Li. Robust Multi-Resolution Pedestrian Detection in Traffic Scenes. *Computer Vision Foundation. CVPR2013*.
- [26].[http://www.bmw.com/com/en/insights/technology/efficientdynamics/phase\\_2/](http://www.bmw.com/com/en/insights/technology/efficientdynamics/phase_2/), last accessed 2018/01/14.
- [27].P. Viola, M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition Conferencies*. 2001.
- [28].Georgakis G., Mousavian A., Berg A. C., Kosecka J. (2017, July). Synthesizing Training Data for Object Detection in Indoor Scenes. *IEEE Conference on Computer Vision and Pattern Recognition*
- [29].Multibox SSD Homepage, <https://towardsdatascience.com/understanding-ssd-multibox-real-time-object-detection-in-deep-learning-495ef744fab>. Last accessed 2018/01/20.
- [30]. Z. Chen, R. Khemmar, B. Decoux, A. Atahouet, and J-Y. Ertaud. Real Time Pedestrian and Object Detection and Tracking-based Deep Learning: Application to Drone Visual Tracking. *27th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, 2019, Plzen, Czech Republic, May 27–30, 2019.