Real Time Pedestrian and Object Detection and Tracking-based Deep Learning. Application to Drone Visual Tracking

Redouane Khemmar, Matthias Gouveia, Benoit Decoux, Jean-Yves y Ertaud

Real Time Pedestrian and Object Detection and Tracking-based Deep Learning: Application to Drone Visual Tracking

R. Khemmar
Institute for Embedded Systems Research, ESIGELEC*, UNIRouen, Normandy University*, Saint Etienne du Rouvray, 76800, France
redouane.khemmar@esigelec.fr

M. Gouveia
Institute for Embedded Systems Research, ESIGELEC*, UNIRouen, Normandy University*, Saint Etienne du Rouvray, 76800, France
gouveia.matthias@hotmail.com

B. Decoux
Institute for Embedded Systems Research, ESIGELEC*, UNIRouen, Normandy University*, Saint Etienne du Rouvray, 76800, France
benoit.decoux@esigelec.fr

JY. Ertaud
Institute for Embedded Systems Research, ESIGELEC, UNIRouen, Normandy University, Saint Etienne du Rouvray, 76800, France
jean-yves.ertaud@esigelec.fr

ABSTRACT

This work aims to show the new approaches in embedded vision dedicated to object detection and tracking for drone visual control. Object/Pedestrian detection has been carried out through two methods: 1. Classical image processing approach through improved Histogram Oriented Gradient (HOG) and Deformable Part Model (DPM) based detection and pattern recognition methods. In this step, we present our improved HOG/DPM approach allowing the detection of a target object in real time. The developed approach allows us not only to detect the object (pedestrian) but also to estimates the distance between the target and the drone. 2. Object/Pedestrian detection-based Deep Learning approach. The target position estimation has been carried out within image analysis. After this, the system sends instruction to the drone engine in order to correct its position and to track target. For this visual servoing, we have applied our improved HOG approach and implemented two kinds of PID controllers. The platform has been validated under different scenarios by comparing measured data to ground truth data given by the drone GPS. Several tests which were ca[1]rried out at ESIGELEC car park and Rouen city center validate the developed platform.

Keywords

Object detection, object recognition, visual tracking, tracking, pedestrian detection, deep learning, visual servoing, HOG, DPM.

1. INTRODUCTION

The works presented in this paper are a part of ADAPT[1] project (Assistive Devices for empowering disAbled People through robotic Technologies) which

[1]This work is carried out as part of the INTERREG VA FMA ADAPT project ”Assistive Devices for empowering disAbled People through robotic Technologies” http://adapt-project.com/index.php. The Interreg FCE Programme is a European Territorial Cooperation programme that aims to fund high quality cooperation projects in the Channel border region between France and England. The Programme is funded by the European Regional Development Fund (ERDF).

focuses on smart and connected wheelchair to compensate for user disabilities through driving assistance technologies. One of the objectives of the project is to develop an Advanced Driver-Assistance System (ADAS) platform for object detection, recognition, and tracking for wheelchair applications (object detection, obstacle avoidance, etc.). The work presented in this paper is related to object/pedestrian detection. 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 safety of vehicle, driver, passenger and pedestrian based on different parameters such as traffic, weather, etc. [26]. Here in this project, ADAS aims to detect pedestrian. Our contribution aims to develop a perception system based on object detection with different approaches such as HOG, DPM, and Deep Learning. This paper is organized as follows: Section 1 introduces the motivation of the paper. Section 2 presents the state of the art about object detection/tracking and visual control. Section 3 presents a comparison between the adopted object detection algorithms and some results obtained. Section 4 illustrates our improved HOG/DPM approach applied to object/pedestrian detection and tracking. In the same section, we present an innovative solution to estimate the distance separating objects to the vehicle or to the drone. The visual control system-based multi approach controller will be presented in Section 5. Finally, in Section 6, we will conclude this paper.

2. STATE OF THE ART AND RELATED WORK

State of the Art

Object detection is a 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 can improve road security and is considered as an important ADAS component in the autonomous vehicle. In [13], a very in-depth state of the art for pedestrian detection is presented. Three main contributions are developed by the authors: 1. Study of the statistics of the size, position, and occlusion patterns of pedestrians in urban scenes in a large pedestrian detection image dataset, 2. A refined per frame evaluation methodology that allows to carry out informative comparisons, including measuring performance in relation to scale and occlusion, and 3. Evaluation of 16 pre-trained state of the art detectors [13]. As a main conclusion of the study, 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 is based on a multiresolution pedestrian model and shows superior detection performance than classical DPM approaches in the detection of small pixel-sized pedestrians. The system was evaluated with the Caltech Pedestrian benchmark [2], which is the largest public pedestrian database. The practicality of the system is demonstrated by a series of use case experiments that uses Caltech video database. The discriminatively trained, multiresolution DPM is presented as an algorithm between generative and discriminative model [3][7]. The algorithm has different steps: building a pyramid of images at different scales, using several filters and part filters to get responses. The algorithm combines these different responses in a star-like model then uses a cost function, and trains classifiers by Support Vector Machine (SVM) classifier. 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], can find a combination of multiple registered image channels, which are computed by linear and nonlinear transformations [9]. Integrating some features like HOG and do a training by AdaBoost in a cascade way can lead to pedestrian detection with good accuracy [9]. The sliding window methods (also called pyramid methods) 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 Deep Learning, also called Convolutional Neural Networks (CNN), become very successful for feature extraction in the image classification task [33][34]. Rich Feature Hierarchies for Convolutional Neural Networks (RCNN) model [12], that combines CNN and selective search can be taken as an example. This algorithm has made a huge progress on object detection task like PASCAL VOC. It will be presented in the next section.

Related Work

In the literature, for similar tasks, several approaches have been used for object detection and pattern recognition, such as HOG/DPM, KLT/RMR (Kanade-Lucas-Tomasi/Robust Multiresolution Estimation of Parametric Motion Models) and Deep Learning [1]. For example, in Google Robot's Project [17], a deep learning model is applied to articulated robot arm for the picking up of objects. In Kitti Vision Benchmark Suite Project (KIT and Toyota Technological Institute) [18], an object detection and orientation estimation benchmark is carried out. The system allows not only localization of objects in 2D, but also estimation of their orientation in 3D [18]. In [19], an example of end-to-end object detection from Microsoft is described. For the task of detecting objects in images, recent methods based on convolutional neural networks (CNN, Deep-Learning) like SSD [18][20] allow detection of multiple objects

in images with high accuracy and in real-time. Furthermore, the SSD model is monolithic and relatively simple compared to other models, making it easier to use for various applications. For the task of grasping objects, people from the AI-Research of Google have recently used Deep Learning to learn hand-eye coordination for robotic grasping [21]. The experimental evaluation of the method demonstrates that it achieves effective real-time control, and it can successfully grasp new objects, and correct mistakes by continuous servoing (control). In many applications of detection of objects like pedestrian, cyclists and cars, it is important to estimate their 3D orientation. In outdoor environments, solutions based on Deep Learning have been recently shown to outperform other monocular state-of-the-art 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 simplify it [23]. In [24] and [25], we can find an evaluation of the state of the art object (pedestrian) detection approaches based on HOG/DPM. In [31], B. Louvat *et al*. have presented a double (cascade) controller for drone-embedded camera. The system is based on two aspects: target position estimation-based KLT/RMR approaches and control law for the target tracking. The developed platform was validated on real scenarios like house tracking. In [32], B. Hérissé has developed in his PhD thesis an autonomous navigation system for a drone in an unknown environment based on optical flow algorithms. Optical flow provides information on velocity of the vehicle and proximity of obstacles. Two contributions were presented: automatic landing on a static or mobile platforms and field following with obstacle avoidance. All algorithms have been tested on a quadrotor UAV built at CEA LIST laboratory.

3. OBJECT DETECTION APPROACH

In order to identify the most suitable approach to our object and/or pedestrian detection application for the autonomous vehicle (a drone in this study), we need to establish the feasibility of several scientific concepts in pattern recognition approaches such as the KLT/RMR, SIFT/SURF, HOG/DPM but especially recent approaches of the artificial intelligence field, such as Deep-Learning. In this paper, we will focus on object detection and tracking-based through improved HOG/DPM approaches.

Classical Approaches

We have started out our experimentations by implementing point of interest approaches like Scale Invariant Feature Transform (SIFT) [9] and Speeded-Up Robust Features (SURF) [35], based on descriptors that are very powerful to find matching between images and to detect objects. These methods allow the extraction of visual features which are invariant to scale, rotation and illumination. Despite their robustness to changing perspectives and lighting conditions, we found that the approach is not robust for object/pedestrian detection. SURF is better as it is up to twice as fast as SIFT. However, SIFT is better when a scale change or an increase in lighting is applied to the image. For pedestrians detection, SIFT and SURF therefore remain insufficient for our application. We have also experimented KLT approach [15] for extraction of points of interest and tracking them between an image taken at t-1 and another image taken at t. The approach has high accuracy and is fast, but on the other hand, it is not robust to perturbations, for example when the target is displaced too much, and in case of highly textured images. This is why we have decided to experiment the RMR approach [16], which has very low precision but is considered to be very robust. We have obtained results which are same as KLT approach. A hybrid KLT/RMR approach would be a better solution where we can have benefits of both the approaches.

Object Detection-based HOG/DPM

HOG and DPM algorithms first calculate image features. They apply classifiers on databases of positive images (with pedestrian) and negative images (without pedestrian). They have the advantage of being accurate and give relatively good results; however, their calculating time is high. We have therefore experimented three approaches: HAAR, HOG and DPM.

3.1.1 Pseudo-HAAR Features

HAAR classifiers use features called pseudo-HAAR [36][27]. Instead of using pixel intensity values, the pseudo-HAAR features use the contrast variation between rectangular and adjacent pixel clusters. These variations are used to determine light and dark area. 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 it possible to use these features on objects with different sizes.

3.1.2 HOG Algorithm

HOG is a descriptor containing key features of an image. These features are represented by the distribution of image gradient directions. HOG is an algorithm frequently used in the field of pattern recognition. It consists of five steps:

1. Sizing of the calculation window ; by default the size of the image to be processed is 64x128 pixels
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 (*ie.* 4 cells)
5. Calculation of the size of the final descriptor.

The obtained descriptor is then given to a SVM classifier. The classifier needs many positive and negative images. The mixture HOG/SVM gives good results with limited computing resources.

3.1.3 DPM Algorithm

The hardest part for object detection is that there are many variances. These variances arise from illumination, change in viewpoint, non-rigid deformation, occlusion, and intra-class variability [4]. The DPM method is aimed at capturing those variances. It assumes that an object is constructed by its parts. Thus, the detector will first find a match by coarser (at half the maximum resolution) root filter, and then using its part models to fine-tune the result. DPM uses HOG features on pyramid levels before filtering, and linear SVM as a classifier with training to find the different part locations of the object in the image. Recently, new algorithms have been developed in order to make DPM faster and more efficient [4][8].

Pedestrian Detection-based HAAR/HOG/DPM

We have carried out all the experiments with 6 different databases dedicated to the pedestrian detection: ETH, INRIA, TUD Brussels, Caltech, Daimler and our own ESIGELEC dataset dedicated to the pedestrian detection. Fig. 1 and Fig.2 shows respectively results obtained under ETH and ESIGELEC datasets.

Figure 1. Comparison between HOG and DPM applied under ETH dataset (640x480-image resolution). Left Column: HOG algorithm, Right Column: DPM algorithm

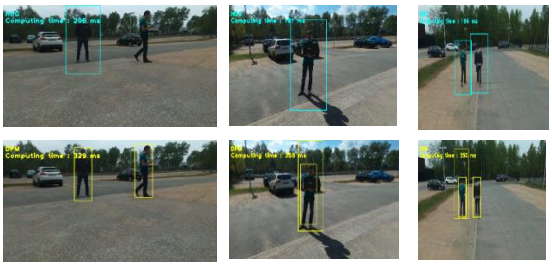


Figure 2. Comparison between HOG and DPM applied under ESIGELEC dataset (640x480-image resolution). Top row: HOG algorithm, Bottom row: DPM algorithm.

Object Detection-based Deep Learning

Many detection systems repurpose classifiers or localizers to perform detection. They apply the classification to an image at multiple locations and scales. High scoring regions of the image are considered as positives detections [4][6]. CNN classifier, like RCNN, can be used for this application. This approach gives good results but require many number of iterations to process a single image [4][6]. Many detection algorithms using selective search [4][5] with region proposals have been proposed to avoid exhaustive sliding window. With deep learning, the detection problem can be tackled in new ways, with algorithms like YOLO and SSD. We have implemented those two algorithms for object detection in real time [14], and obtained very good results (qualitative evaluation). Fig. 3 shows an example of our deep learning model applied in pedestrian detection within ESIGELEC dataset at the city center of Rouen.

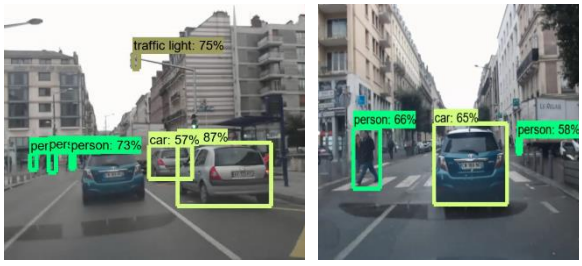


Figure 3. Object/Pedestrian detection-based deep learning (Yolo and SSD). The results have been obtained under GPU Nvidia Quadro K4000M machine.

4. OBJECT DETECTION AND TRACKING-BASED FASTER DPM ALGORITHM

Improved HOG Algorithm

We have identified four methods to improve HOG/DPM algorithm, which correspond to the five step of the HOG algorithm. In order to be able to make comparisons with the original version, a classifier was trained on the INRIA and TUD-Brussels image datasets. The number of available images in these datasets is relatively small compared to some other datasets, which has an impact on the pedestrian detection quality. The objective is to improve the quality of HOG/DPM pedestrian detection and to significantly reduce the computation time for the detection in road scenes. The tests were carried out on the ESIGELEC dataset; 1. Gamma Correction: in order to improve the pedestrian image quality, we performed a gamma pre-processing on the images to be trained and tested. Gamma correction can illuminate more the dark areas in images if gamma coefficient is greater than 1 or, in the contrary, darken them if the coefficient is less than 1. We found fewer parasites in the processed images. 2. Image Resizing: HOG algorithm performs calculations on images of size 64x128 pixels. We performed calculation with windows size 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 improved HOG). 3. Negative Gradients: when calculating gradients, improved HOG uses negative gradients (180° to 0) and positive gradient (0 to 180°) like classical HOG. This allows the calculation of histogram with 18 values (9 values in classic HOG). The calculation time does not vary, however, by taking the negative gradients (signed gradient) into account, a small improvement was carried out but considered as not significant. However, we found presence of noise, which does not improve the classic HOG. 4. Normalization Vector: as a last step, HOG performs a standardization of 2x2 cells, *ie*. 32x32 pixels with a pitch of 16x16. The object detection is degraded and the computation time doubles, which cannot considered as an improvement of classic HOG.

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 HOG algorithm only in a Region of Interest (RoI) in which the target (here the pedestrian) is located. This will drastically reduce the computing time and will also better 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 as a bounding box (yellow box in figure 4), we built a New RoI (NRoI) by providing a tolerance zone (a new rectangle like the green one in figure 4 which is larger than the first one). Starting form 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. In Fig. 4, we can see that Faster-DPM improves target detection by reducing time from 4 to 8 times less than classic DPM.

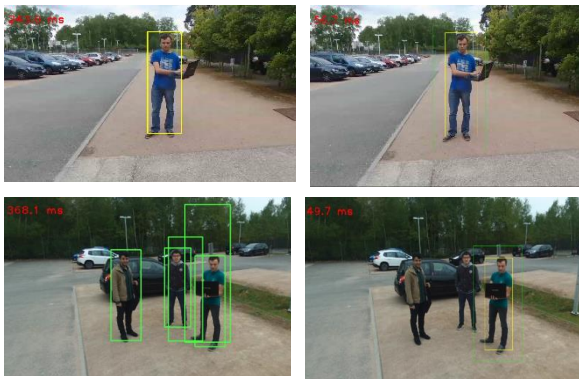


Figure 4. Comparison between classic DPM and Faster-DPM. Left top and bottom image: object detection-based DPM, Right top and bottom image: object detection-based Faster-DPM with adaptive RoI (green box).

Pedestrian Distance Estimation

To have a better information on the detected pedestrians, it is necessary to estimate the distance separating the pedestrians from the vehicle. As our system is based on a monocular camera, the measurement of this distance has to be estimated. The law called “Inverse Squares”, 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 (bounding box surrounding the target). We have used a parametric approach. We have taken measurements at intervals of 50 cm, and the corresponding object RoI surfaces have been recorded. Fig. 5 shows the result obtained.

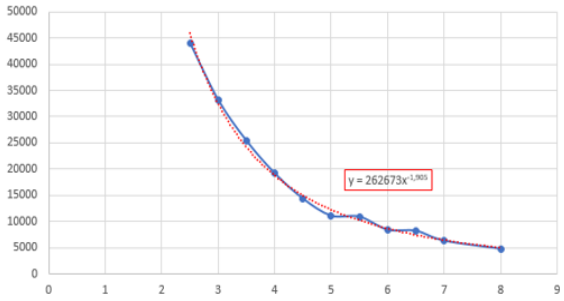


Figure 5. Detected RoI surface (blue curve) *vs* distance from the object (red curve): y axis represents the surface (pixels^2), and x axis represents distance (m).

The blue curve looks like a square root. This is validated by the trend curve, whose as equation (1) is:

$$y = 262673 * x^{-1.905} \quad (1)$$

where x is the abscissa and y the ordinate.
The equation of blue curve is (2):

$$S = A * d^{-2} \quad (2)$$

where S is the surface in pixel^2, d is the distance to be estimated, and A = 262673.

The ideal detection zone is between 3 and 7 meters with an error of 5%. The Tab. 1 illustrates the calibration process carried out for the measurement environment.

Distance (m)	RoI Surface (pixel^2)	K = S * d^{-2}
2.5	44044	275275
3	33180	298620
3.5	25392	311052
4	19200	307200
4.5	14360	290790
5	11102	277550
5.5	10920	330330
6	8427	303372
6.5	8216	347126
7	8348	311052
8	4800	307200

Table 1. Measurement environment calibration.

5. VISUAL CONTROL-BASED MULTI APPROACH CONTROLLER

Drone Double Controller

The speed servo control of the drone allows it to track the target continuously and in real time. The image processing as a closed loop gives the coordinates of the target in (x,y,z) plan. Using this information, the system send instruction to the Drone engines corresponding to the real time position of the target in order to correct the position of the drone. We have developed two kinds of controllers: Proportional-Integral-Derivative (PID) and Adaptive Super Twisting (AST) controller [28]. The controller system is based on three corrections: 1. Drone altitude correction, 2. Drone rotation speed, and 3. Drone navigation. Under Robot Operating System (ROS) system, the instructions are considered as speeds sent to drone engines.

Firstly, we have applied a Proportional (P) controller:

- Altitude : $K_p = 0.004$
- Forward translation : $K_p = 0.08$

Secondly, we have applied a classic PID controller (3):

$$V_{rotation} = K_p * erreur + K_i * \int_0^t erreur * dt + K_d * \frac{d}{dt}(erreur) \quad (3)$$

With Rotation: $K_p = 0.0041$, $K_i = 0.0003$, $K_d = 0.0003$

Lastly, in (4), we have applied an AST (nonlinear PID) controller:

$$V_{rotation} = K_p * \left(\frac{erreur}{|erreur| + A}\right) * \sqrt{|erreur|} + K_i * \int_0^t \frac{erreur}{|erreur| + A}\, dt \quad (4)$$

with $A = 0.001$, $K_p = 0.013$, and K_i=0.02.

The P controller is amply enough to enslave the altitude of the drone and its translation. However it is the rotational slaving that predominates; we need to keep the target in the center of the image and the target moves strongly from right to left and vice versa.

Overall, the visual servoing that we have adopted uses the data extracted during the HOG/DPM or deep learning image processing (visual controller) and sends commands to the drone actuators. This is a double controller: speed control (internal loop) to control the motors and servo positioning (external loop) to position the drone according to the target position to track. The correction can be done with classical correctors (P, PI, PID) or more advanced commands like for example AST controller. Fig. 6 illustrates the architecture of our cascade control used.

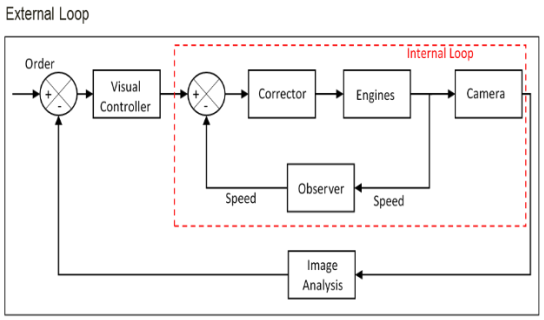


Figure 6. Double controller with external (black) and internal loop (in red).

Test & Validation

Indoor and outdoor environment tests were carried out throughout the project. For the final test phases, scenarios have been established. In order to validate the final platform, several scenarios has been defined. The platform gives very good results. Better results are obtained when the target (person) is alone in the environment. Performance is reduced when the target is in a group, despite the use of an adaptive region of interest. In addition, there is always the risk of losing the target, object detection algorithms do not have a detection rate of 100% and sometimes the person to track is no longer detected. If the target is no longer detected, the drone is ordered to switch to stationary mode as a filled situation. We are correcting this problem through two different methods: 1. An estimation of the future position of the target based on improved Kalman filter that gives better results; from now on, we are able to predict the position of the target to be followed and thus to minimize a possible confusion of this target with another target of the same nature (as for example the case of several people who walk together). A second approach is also under development which concerns deep learning not only for object/people detection and tracking, but also for object distance and orientation estimation. To compare the performance of the two implemented control laws (PID controller and AST controller), the GPS coordinates of each trajectory were recorded and compared. In order to illustrate the results obtained, the target (here a person) has made a reference trajectory in the form of a square. Fig. 7 shows the performances obtained on a square shape trajectory. We can see that the PID controller is more accurate than the AST controller.

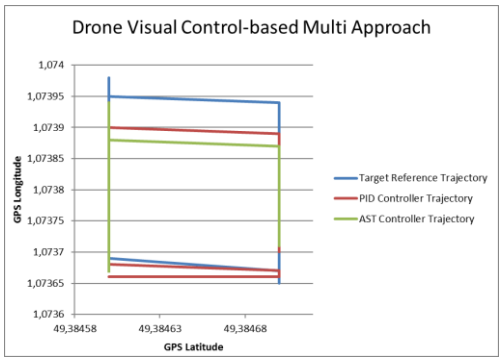
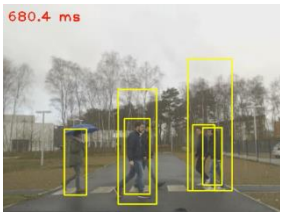


Figure 7. PID and AST comparison in Object/Person detection and tracking-based on our improved HOG/DPM. Top image: target square trajectory (blue line) carried out but the mobile target (person). Bottom image: comparison for the different trajectory carried out by person (ground truth data), PID controller and AST controller. x and y axis represents the geographic coordinate system coordinates.

6. EXPERIMENTAL RESULTS

The developments were carried out on an Ubuntu Linux platform with ROS, which is a set of computer tools for the robotics environment that includes a collection of tools, libraries and conventions that aim to simplify the process of operation, thus allowing more complex and robust robot behavior [29]. The ROS architecture developed comprises 4 different nodes: Node 1: image acquisition from the drone's embedded camera, Node 2: image processing and calculation of the position of the target to follow, Node 3: visual servoing-based on PID or AST controller, and Node 4: Sending commands to the drone (manual and/or autonomous steering). The drone used is a Parrot Bebop 2 with 22 minutes of autonomy, weight 420g, range of 300m, camera resolution of 14 Mpx, and maximum speed of 13m/s. We have carried out several tests with six 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 Rouen shopping center car park and Rouen city center. The results obtained are shown in Fig. 8.



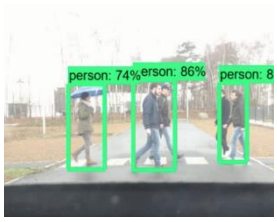


Figure 8. Pedestrian Detection with improved HOG/DPM approach under ESIGELEC Indoor/Outdoor dataset. Left top image: Pedestrian detection with calculated distance under DPM, Right top image: Pedestrian detection-based SSD deep learning model, Left bottom image: Pedestrian detection-based DPM, Right bottom image: Pedestrian detection-based Faster-DPM.

7. CONCLUSION

In this paper, we have presented a new approach for object detection and tracking applied for drone visual control. A comparison study between different approaches dedicated to pattern recognition and object/pedestrian detection has been carried out. We have present our contribution to improve the quality of both HOG and DPM algorithms. We have also implemented deep learning based Yolo and SSD algorithm for pedestrian detection. An innovative method for distance calculation of pedestrian/object is presented in this paper. The approach were validated within experiments and the accuracy is 5% of errors. The system detects not only object/target to follow but also their distance from the drone. The system is generic so that it is "*applicable*" on any type of platform and/or environment (pedestrian detection for autonomous vehicle and smart mobility, object detection for smart wheelchair, object detection for autonomous train, etc.). This work aims to show the feasibility of scientific and technological concepts that affect object detection and tracking-based classic approaches like HOG/DPM, but also deep learning approaches-based Yolo or SSD. We have validated the whole development under several scenarios by using both parrot drone platform and real vehicle in real traffic conditions (city center of Rouen in France).

8. ACKNOWLEDGMENTS

This research is supported by ADAPT project (This work is carried out as part of the INTERREG VA FMA ADAPT project ”Assistive Devices for empowering disAbled People through robotic Technologies” http://adapt-project.com/index.php. The Interreg FCE Programme is a European Territorial Cooperation programme that aims to fund high quality cooperation projects in the Channel border region between France and England. The Programme is funded by the European Regional Development Fund (ERDF). Many thanks to the engineers of Autonomous Navigation Laboratory of IRSEEM for the support during platform tests.

9. REFERENCES

[1]. H. Cho, P. E. Rybski, A. B-Hillel, W. Zheng.: Real-time Pedestrian Detection with Deformable Part Models, 2012

[2]. Caltech Pedestrian Detection Benchmark Homepage, 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., 2014.

[5]. Koen E. A. van de Sande, Jasper R. R. Uijlings, TheoGevers, 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 neural networks, NIPS, 2012.

[7]. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR, 2014.

[8]. P. Dollar, C. Wojek, B. 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. The fastest deformable part model for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2497-2504). 2014.

[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 Se-lective 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. IEEE CVPR, 2012.

[19]. Eddie Forson. Understanding SSD MultiBox-Real Time Object Detection in Deep Learning. Towards Data Science. November, 18th. 2017.

[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]. 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

[25]. P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: A Benchmark. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Piscataway, NJ, pp. 304-311. ISBN 978-1-4244-3991-1. 2009.

[26]. Junjie Yan, Xucong Zhang, Zhen Lei, Shengcai Liao, Stan Z. Li. Robust Multi-Resolution Pedestrian Detection in Traffic Scenes. Computer Vision Foundation. CVPR 2013.

[27]. http://www.bmw.com/com/en/insights/technology/efficientdynamics/phase_2/, last accessed 2018/01/14.

[28]. P. Viola, M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. Computer Vision and Pattern Recongition Conferencies. 2001.

[29]. Mohamed, G., Ali, S. A., & Langlois, N. (2017). Adaptive Super Twisting control design for manufactured diesel engine air path. The International Journal of Advanced Manufacturing Technology, 1-12. 2017.

[30]. ROS homepage: http://www.generationrobots.com/blog/fr/2016/03/ros-robot-operating-system-3/. Last accessed 2017/12/12.

[31]. B. Louvat, "Analyse d'image pour l'asservissement d'une camera embarquée dans un drone". Gipsa-lab, February, 5th, 2008, Grenoble. 2008.

[32]. B. Hérissé, "Asservissement et navigation autonome d'un drone en environnement incertain par flot optique. PhD thesis, Université Sophia Antipolis, UNSA et I2S CNRS, November 19th, 2010.

[33]. R. Girshick, "Fast R-CNN," ICCV, 2015

[34]. S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," arXiv:1506.01497. 2015.

[35]. H. Bay, T. Tuytelaars, and L. V. Gool. "SURF: Speeded Up Robust Features", Computer Vision – ECCV 2006, pp. 404-417. 2006.

[36]. D. Gerónimo, A. López, D. Ponsa, A.D. Sappa. "Haar wavelets and edge orientation histograms for on–board pedestrian detection". In: Pattern Recognition and Image Analysis, pp. 418-425. 2007.