MOBILE ROBOT WITH VISION BASED NAVIGATION AND PEDESTRIAN DETECTION

ROBOT MÓVIL CON NAVEGACIÓN BASADA EN VISIÓN Y DETECCIÓN DE PEATONES

Marco A. Luna[1], Julio F. Moya[1], Wilbert G. Aguilar[2,3,4,*], Vanessa Abad[5]

Abstract

This article proposes the design and implementation of a low-cost vision based navigation mobile robot that tracks pedestrians in real time using an IP camera onboard. The purpose of this prototype is the navigation based on people tracking keeping a safe distance by PID and on-off controllers. For the implementation we evaluate two pedestrian detection algorithms: HOG cascade classifier and LBP cascade classifier off-line and onboard the robot. In addition, we implement a communication system between the robot and the ground station. The metrics of evaluation for the pedestrian detection proposals were precision and sensibility, obtaining better results with HOG. Finally, we evaluate the communication system, computing the delay of the controller response; the results show that the system works properly with a transmission rate of 115200 bauds.

Keywords: AdaBoost, HOG, LBP, Pedestrian Detection, Urban Navigation.

Resumen

Este artículo propone el diseño e implementación de un robot móvil con navegación basada en visión, de bajo costo, que sigue la trayectoria de peatones en tiempo real usando una cámara IP a bordo. El propósito de este prototipo es la navegación basada en el seguimiento de personas conservando una distancia segura a través de controladores PID y $on-off$. Para la puesta en marcha se evalúan dos algoritmos de detección de peatones: cascada de clasificadores HOG y cascada de clasificadores LBP, tanto fuera de línea como a bordo del robot. Adicionalmente, se implantó un sistema de comunicación entre el robot y una estación de tierra. Las métricas de evaluación para las propuestas de detección de personas fueron la precisión y sensibilidad, obteniendo mejores resultados con HOG. Al final, se evaluó el sistema de comunicación, calculando el retraso de la respuesta del controlador. Los resultados mostraron que el sistema trabaja adecuadamente para una tasa de transmisión de 115200 baudios.

Palabras clave: AdaBoost, detección de peatones, HOG, LBP, navegación urbana.

[1]Departamento de Eléctrica y Electrónica DEEE, Universidad de las Fuerzas Armadas ESPE, Sangolquí – Ecuador.
[2]Departamento de Seguridad y Defensa DESD, Universidad de las Fuerzas Armadas ESPE, Sangolquí – Ecuador.
[3]Centro de Investigación Científica y Tecnológica del Ejército CICTE, Universidad de las Fuerzas Armadas ESPE, Sangolquí – Ecuador.
[4,*]Grup de Recerca en Enginyeria del Coneixement GREC, Universitat Politècnica de Catalunya UPC, Barcelona – España. Autor para correspondencia ✉: wgaguilar@espe.edu.ec
[5]Departament de Genètica, Universitat de Barcelona UB, Barcelona – España.

1. Introduction

Autonomous navigation is a research line with growing interest from DARPA Grand Challenge 2004. Several vision based vehicles have been developed [1–3]. There are a lot of challenges in urban navigation like perception [4, 5], obstacles avoidance [6, 7], people detection [8–11] and video stabilization [12–14].

Multiple pedestrian detection algorithms have been developed due their different applications that include: surveillance [15–17], driver assistance systems [11, 18, 19], robotics [20, 21], and others. According to [22] most approaches use detectors based on HAAR-like features [23], HOG [24, 25], LBP [26] and some combinations like: HOG-HAAR [27] or HOG-LBP [28]. In [29], HAAR-like features have a low performance in pedestrian detection task.

In this work, we experimentally test Histogram of Oriented Gradients (HOG) [24] and Local Binary Patterns (LBP) [26], feature extraction algorithms for pedestrian detection with Adaptive Boosting (AdaBoost), a supervised learning algorithm, applied in mobile robot navigation. The video is transmitted via Wi-Fi to a ground station to determine the motion direction of the robot using image processing, and the control actions are sent by Bluetooth from the computer to the robot.

The rest of the article is distributed as follows: In the next section, we do a review of the related work. In the section III, we make a review of two algorithms used for pedestrian detection. We describe our approach for robot development in the section IV. Finally the results obtained and the conclusions are presented in the sections V and VI.

2. Related Works

In the literature, there are several applications that integrate people tracking with mobile robots. A ground autonomous vehicle designed to track people based on a vision was proposed in [30], in this work the person to be tracked should wear a discriminable rectangle.

In [20], a mobile platform that uses a multiple sensor fusion approach and combines three kinds of sensors in order to detect people using RGB-D vision, lasers and a thermal sensor is presented. The mobile robot in [21] uses an omnidirectional camera and laser range finder to detect and track people. In [31] is introduced a robot to assist elderly people with a Kinect device and ROS packages. The detection and localization of people is an important aspect in robotic applications for interaction. The work presented in [32] deals with the task of searching for people in home environments with a mobile robot. The method uses color and gradient models of the environment and a color model of the user. Evaluation is done on real-world experiments with the robot searching for the user at different places.

We propose a simple structure robot (equipped only with a smartphone camera for navigation) capable of detect and track people orientation and translation based on pedestrian detection algorithms. In contrast with some words presented in the literature, our robot performs the navigation task using a PI control to achieve relatively fast speed, avoiding braking issues.

3. Pedestrian Detection Algorithms

3.1. Histogram of Oriented Gradients (HOG)

This algorithm is a feature descriptor for object detection focused on pedestrian detection and introduced in [24]. The image window is separated into smaller parts called cells. For each cell, we accumulate a local 1-D histogram of gradient orientations of the pixels in the cell. The gradient orientation is given by equations 1 and 1:

$$I_X = I \times D_X \qquad I_Y = I \times D_Y$$

$$D_X = [1, 0, 1] \quad D_Y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \qquad (1)$$

$$\phi = \alpha \tan(I_Y, I_X) \qquad (2)$$

Where: I is the image and I_X, I_y are the derivatives of the image in x,y.

Each cell is discretized into angular bins according to the gradient orientation and each pixel of the cell contributes with a gradient weight to its corresponding angular bin. The adjacent cells are grouped in special regions called blocks and the normalized group of histograms represents the block histogram. Finally, the set of these block histograms represents the descriptor.

3.2. Local Binary Patterns (LBP)

This algorithm was presented like a texture descriptor for object detection. This compares a central pixel with the neighbors. The central pixel value is taken as threshold and a value of "1" is assigned if the neighbor is greater or equal to the central pixel, otherwise the value is "0". In each pixel we set a weight of 2n according to the position respect to the central pixel [26]. The parameters of LBP operator are R and P, where R is the distance to the central pixel and P is the number of pixels [14]. LBP is defined mathematically as:

$$LBP_{P,R}(X_C, Y_C) = \sum_{P=0}^{P-1} s(g_P - g_C) \cdot 2^P \qquad (3)$$

In the equation , g_P is the value of the central pixel, g_C the value of the neighbors and 2P is the weight stablished for each operation. And $s(g_P - g_C)$ is given by:

$$s(g_P - g_C) = \begin{cases} 1 & g_P - g_C \geq 0; \\ 0 & g_P - g_C < 0 \end{cases} \quad (4)$$

3.3. Adaboost

Adaboost is a machine learning algorithm [33] that initially keeps uniform distribution of weights in each training sample. In the first iteration the algorithm trains a weak classifier using a feature extraction methods or mix of them achieving a higher recognition performance for the training samples.

In the second iteration, the training samples, misclassified by the first weak classifier, receive higher weights.

The new selected feature extraction methods should be focused in these misclassified samples. The final result is a cascade of linear combinations with selected weak classifiers:

$$g(x) = \sum_{t=1}^{T} \wp^t h^t(x) \quad (5)$$

where: h^t is a classifier, and $\wp^t$ is a coefficient.

4. Our Approach

4.1. Mobile Robotic Platform

We are using the Ackerman steering configuration for the robot [19], i.e. two wheels with rear haulage and two with front direction presented in the Fig. 1. The system is designed so that the interior front wheel in a gyre has an angle slightly sharper than the exterior, in order to avoid skidding; the normal wheels intersect at a point on extension of the axis of the rear wheels and make the toolpath for constant rotation angles.
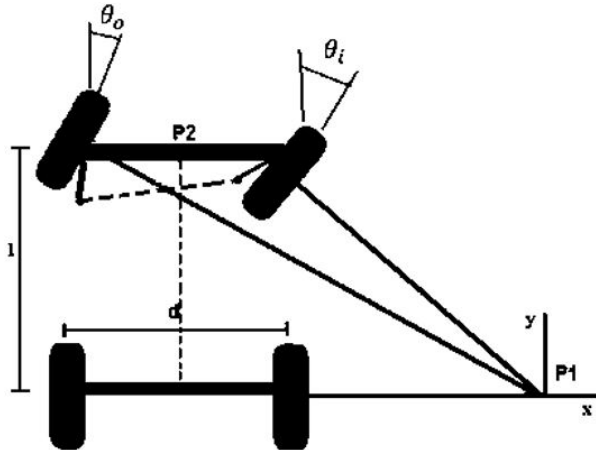

Figure 1. Position of wheels in Ackerman steering configuration.

The relationship between the angles of the wheels of direction is in equation 6:

$$\cot \theta_i - \cot \theta_o = \frac{d}{l} \quad (6)$$

where:

θ_i= Relative angle of the inner wheel.
θ_o= Relative angle of an outer wheel.
l= Longitudinal separation between wheels.
d= Lateral separation between wheels.

For traction and steering control the system use Pulse Width Modulator (PWM) with Chopper configuration [34]. Additionally, there is a smart device (smartphone) camera for navigation.

The robot has two motors, one for forward/reverse displacement and other that controls direction. An H bridges circuit for motor control is implemented. The camera of smartphone is used as IP camera for image capture.

4.2. Pedestrian Detection

The feature extraction algorithms previously described are applied on the mobile robot to obtain an autonomous navigation system based on pedestrian tracking. We use Adaboost as machine learning method. It is an algorithm in machine learning based on the notion of creating a highly accurate prediction rule by combining many relatively weak and inaccurate rules [33]. Fig. 2 shows images of the HOG algorithm performance in the mobile robot.
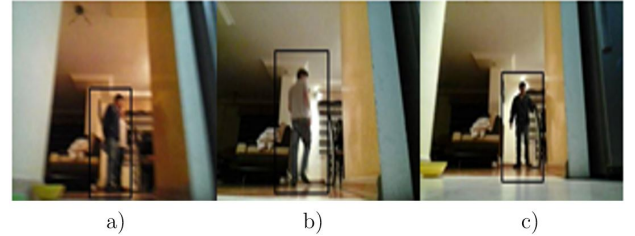

a) b) c)

Figure 2. Pedestrian detection with HOG algorithm. (a) Side detection of a person. (b) Back detection of a person. (c) Frontal Detection of a person.

Performance of LBP algorithm in images captured from the mobile robot is presented on Fig. 3.
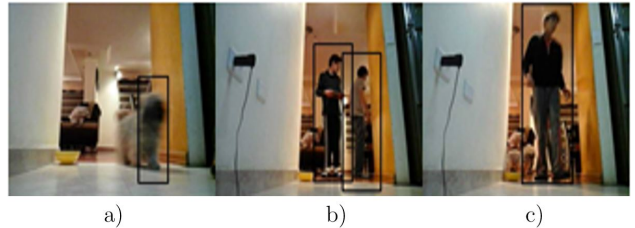

a) b) c)

Figure 3. Fig. 3. Pedestrian detection with LBP algorithm. (a) Detection of a false positive. (b) Detection of two people. (c) Frontal detection of a person.

4.3. Controller

The mobile robot uses two main inputs for the controller (needed for 2D navigation [35]): horizontal position and distance obtained from the bounding box of the pedestrian detection. Horizontal position depends of x coordinate of the bounding box centroid and the distance depends of the width of the bounding box. We use PI control for the distance that reduces overshoot in order to have a soft controller. For horizontal position we use an on-off controller with hysteresis, i.e. with a band of maximum and minimum values in the output due to the variation range of the actuators is small and precision is not necessary (Fig. 4).
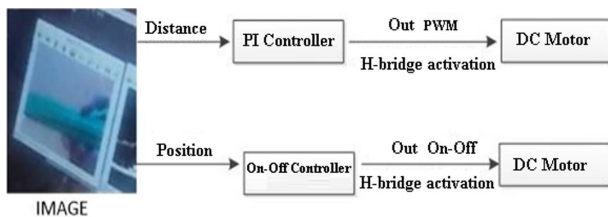

Figure 4. Robot navigation control scheme.

4.4. Communication

Wi-Fi communication is implemented for transmitting data from the IP camera to the computer. This is because Wi-Fi is able to send image data. In the other side we use Bluetooth communication to send control command from the computer to the robot. The transmission speed obtained from experimentation is 115200 baud. (See section 5).

5. Experimentation and Results

5.1. Pedestrian Detection

The performance of pedestrian detection algorithms was compared using online and offline videos. We are using recall and precision as metrics of evaluation. The formulas used are presented in equations 7 and 8 respectively.

$$recall = \frac{VP}{VP + FN} \times 100\% \qquad (7)$$

$$precision = \frac{VP}{VP + FP} \times 100\% \qquad (8)$$

where:
TP: true positives.
FN: false negative.
FP: false positives.

We realized two experiments, offline and online. For the online test, we used four different security cameras videos obtained from Internet. We captured 1000 frames separated in groups of 30 frames. For each group, we determined true positives, true negatives, false positives and false negatives using HOG and LBP algorithms. The image processing was realized with a 2.4 GHz processor and RAM memory of 4 GB. In the online test we recorded 7 minutes of video, i.e. 12600 frames, and the same procedure in offline test was performed. The results are presented in Table 1 and Table 2.

The Table 1 presents the results of recall of the detectors tested offline and online.

Table 1. Detection recall of the algorithms previously trained

Algorithm	Recall	
	Offline	*Online*
HOG +Adaboost	78%	74,20%
LBP+Adaboost	76,50%	75%

Table 2. Detection precision of the algorithms previously trained.

Algorithm	Accuracy	
	Offline	Online
HOG + Adaboost	76%	74%
LBP + Adaboost	73%	72%

The HOG precision test is better than LBP because LBP performance give more false positives in both cases (offline and online). The LBP algorithm compares image textures and confuses other objects with persons easily. HOG is a descriptor based on objects shape and focused principally on pedestrians [24], thus it has better performance but with higher computational cost. Based on the results, we choose HOG because the computational cost problem can be reduced using a processor with better characteristics.

5.2. Controller response and communication

During implementation of the robot, it was found that the response time of the controller varied according to Bluetooth transmission speed, the results of the test are presented in Table 3.

Table 3. Delay in the controller's response to different transmission speeds.

Transmission Speed (bauds)	Delay(s)
9600	≈ 2,5
14400	≈ 2
19200	≈ 1,5
56000	1
115200	≈ 0

It should be noted that the delays were measured according to the perceived response of the controller. A transmission rate of 115200 baud was chosen based on results.

6. Conclusiones

The HOG algorithm has better performance off-line; however, the implementation represents a higher computational cost. LBP had a lower performance offline but was not affected for online experiments. The false positive detection rate involves low precision.

Classical method of PI control provided a smooth braking control to keep a safe distance between the mobile robot and the pedestrian. On-Off control for the horizontal position is a simple controller that had a good performance maintaining the direction of the navigation.

A slow controller response endangered the safety of pedestrians, so it was necessary to increase the transmission speed obtaining better results. The response time also was affected by the parameters of the PI controller. People tracking algorithms have benefits in several applications. We can improve the algorithms for pedestrian detection, based on obtained results, with some techniques like specialized training or video treating, but this involves an increased computational cost.

Acknowledgments

References

[1] M. Bertozzi, A. Broggi, and A. Fascioli, "Vision-based intelligent vehicles: State of the art and perspectives," *Robotics and Autonomous Systems*, vol. 32, no. 1, pp. 1–16, 2000.

[2] W. Aguilar and C. Angulo, "Compensación de los Efectos Generados en la Imagen por el Control de Navegación del Robot Aibo ERS 7," *Memorias del VII Congreso de Ciencia y Tecnología ESPE 2012*, no. JUNE, pp. 165–170, 2012.

[3] ——, "Compensación y Aprendizaje de Efectos Generados en la Imagen durante el Desplazamiento de un Robot," *X Simposio CEA Ingeniería de Control 1,2 Marzo, 2012*, pp. 165–170, 2012.

[4] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," *Computer Vision and Pattern Recognition*, pp. 3354–3361, 2012.

[5] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler, "Autonomous driving goes downtown," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 6, pp. 40–48, 1998.

[6] U. Franke and I. Kutzbach, "Fast stereo based object detection for stop&go traffic," *IEEE Intelligent Vehicles Symposium*, pp. 339–344, 1996.

[7] W. Aguilar and S. Morales, "3D Environment Mapping Using the Kinect V2 and Path Planning Based on RRT Algorithms," *Electronics*, vol. 5, no. 4, p. 70, 2016.

[8] W. G. Aguilar and C. Angulo, "Estabilización de vídeo en micro vehículos aéreos y su aplicación en la detección de caras," pp. 155–160, 2014.

[9] D. Gavrila, J. Giebel, and S. Munder, "Vision-based pedestrian detection: the PROTECTOR system," *IEEE Intelligent Vehicles Symposium, 2004*, pp. 13–18, 2004.

[10] a. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: single-frame classification and system level performance," *IEEE Intelligent Vehicles Symposium, 2004*, pp. 1–6, 2004.

[11] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, 2010.

[12] W. G. Aguilar and C. Angulo, "Real-Time Model-Based Video Stabilization for Microaerial Vehicles," *Neural Processing Letters*, vol. 43, no. 2, pp. 459–477, 2016.

[13] ——, "Real-time video stabilization without phantom movements for micro aerial vehicles," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 46, 2014.

[14] ——, "Robust Video Stabilization based on Motion Intention for Micro Aerial Vehicles," *Systems Signals and Devices (SSD), 2014 International Multi-Conference on*, pp. 1 – 6, 2014.

[15] T. Teixeira, D. Jung, and A. Savvides, "Tasking networked cctv cameras and mobile phones to identify and localize multiple people," *Proceedings of the 12th ACM . . .*, pp. 213–222, 2010.

[16] H. Torresan, "Advanced surveillance systems: combining video and thermal imagery for pedestrian detection," *Proceedings of SPIE*, vol. 5405, pp. 506–515, 2004.

[17] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian Detection in Infrared Images based on Local Shape Features," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 0–7, 2007.

[18] D. M. Gavrila, "Pedestrian Detection from a Moving Vehicle," *System*, pp. 37–49, 2000.

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

[20] O. H. Jafari, D. Mitzel, and B. Leibe, "Real-Time RGB-D based People Detection and Tracking for Mobile Robots and Head-Worn Cameras," *IEEE International Conference on Robotics & Automation (ICRA)*, no. May 2016, pp. 5636–5643, 2014.

[21] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia, "People tracking and following with mobile robot using an omnidirectional camera and a laser," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2006, no. May, pp. 557–562, 2006.

[22] Benenson, Rodrigo; Omran, Mohamed; Hosang, Jan; Schiele, Bernt, "Ten Years of Pedestrian Detection, What Have We Learned?" *Proceedings of the Computer Vision-ECCV 2014 Workshops*, pp. 613–627, 2014.

[23] P. Viola and M. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[24] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

[25] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1491–1498, 2006.

[26] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," *Advances in Pattern Recognition*, vol. 2013, pp. 399–408, 2001.

[27] C. Wojek and B. Schiele, "A Performance Evaluation of Single and Multi-feature People Detection," *Pattern Recognition*, vol. 11, no. 2, pp. 82–91, 2008.

[28] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," *Computer Vision, 2009 IEEE 12th International Conference on*, no. Iccv, pp. 32–39, 2009.

[29] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: an evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–61, 2012.

[30] Y. Aleksandrovich, "Mobile Robot Navigation Based on Artificial Landmarks with Machine Vision System," *World Applied Sciences . . .*, vol. 24, no. 11, pp. 1467–1472, 2013.

[31] H. M. Gross, C. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, C. Martin, T. Langner, and M. Merten, "Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment," *IEEE International Conference on Intelligent Robots and Systems*, pp. 2430–2437, 2011.

[32] M. Volkhardt, S. Müller, C. Schröter, and H.-M. Groß, *Detection of Lounging People with a Mobile Robot Companion.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 328–337.

[33] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[34] M. Hagiwara and H. Akagi, "PWM control and experiment of modular multilevel converters," *Power Electronics Specialists Conference, 2008. PESC 2008. IEEE*, pp. 154–161, 2008.

[35] W. G. Aguilar, R. Costa-castelló, and C. Angulo, "Control autónomo de cuadricopteros para seguimiento de trayectorias," *IX Congreso de Ciencia y Tecnologia ESPE 2014*, pp. 144–149, 2014.