PEOPLE DETECTION SYSTEM FOR AN AUTONOMOUS ROBOTIC PLATFORM INTEGRATED WITH SIMULTANEOUS LOCALIZATION AND MAPPING

Alberto Torres Angonese, Paulo Fernando Ferreira Rosa*

*Instituto Militar de Engenharia
Pç General Tibúrcio, 80
Rio de Janeiro, RJ, Brasil

Email: angonesealberto@gmail.com, rpaulo@ime.eb.br

Abstract— This paper presents the implementation of a people detection system for a robotic platform, allowing the exploration and navigation of the robot considering human detection interaction. The robotic platform consists of a Pioneer 3DX robot equipped with an RGB-D camera, a Sick Lms200 sensor laser and a computer using the robot operating system ROS. The idea is to integrate the people detection system to the simultaneous localization and mapping (SLAM) system of the robot using ROS. In this paper, people identification technique is based on HOG (Histogram of Oriented Gradients) detectors. The technique's accurateness was evaluated using two different libraries. The PCL library (Point Cloud Library) implemented in C ++ and MatLab VLFeat library with two variants of HOG, the original one, and the DPM (Deformable Part Model) variant. Tests were made objecting the evaluation of losses and false positives in the people's detection process. It allows evaluating the people detection system during the navigation and exploration of the robot, considering the real time interaction of people recognition in a semi-structured environment.

Keywords— People Detection, HOG detector, Deformable Part Model, SLAM, Robot Operating System ROS

1 Introduction

The human and robot interaction is a reality presented nowadays. For performing tasks as search and rescue, domestic services or elderly assistance is expected that an autonomous robot is capable of locating itself while exploring the environment. This issue characterizes the research area known as SLAM (Simultaneous Localization and Mapping). For both domestic and rescue services, it is interesting that the robot considers the detection of people while performing SLAM and thereby, modifying its decision-making actions based on this detection. This paper presents a people detection method that runs simultaneously with a SLAM method, already implemented and presented in a previously paper (Silveira Vidal et al., 2015).

The human recognition has been the focus of research for a long time. Despite the impressive results achieved by a wide range of techniques, it still presents many challenges to overcome, especially considering its application in mobile autonomous robotic. In (Munaro and Menegatti, 2014), the authors present a methodology for people detection and tracking from RGB-D images extracted by a mobile robotic platform. The paper describes a real-time people detection and tracking application that is based on Histograms of Oriented Gradients (HOG). Notwithstanding positive results presented by the authors, scenarios and people's position should have "well-behaved" settings for satisfactory detection, limiting its application in unstructured environments. It is necessary the manual identification of a ground plan so the system can be robust to scale, meaning that it can detect people at different distances. Another limitation is that humans should be in a vertical position.

The different methods to discover more robust detection of variations has been motivating the interest of many researchers. In (Parizi et al., 2012), it is applied the DPM (Deformable Part Models) method, whose proposal is to perform a detection using HOG descriptor and from it, identifying deformable parts of structures, enabling the human's recognition in different positions.

Even with the presented advances, for satisfactory results, the HOG, and the DPM techniques require efficient mechanisms of computational learning. In this scenario, a widely used method of machine learning is SVM (Support Vector Machine). However, its results depend on the quality of the training dataset.

Figure 1: People Detection Challenges

In figure 1, it is possible to observe the green bounding box representing that the person at the upright position was detected and the person in crouched position was not.

The majority of researches dealing with SLAM do not take into account the presence of people interacting in the environment which is being mapped by the robot, even with exploration and mapping of multiple cooperatives robots as in (de Melo Neto et al., 2012) or considering dynamics environments as in (Silveira Vidal et al., 2013). People detection and SLAM systems were integrated by robot operating system ROS (Quigley et al., 2009) embedded on the robot. The general idea is to change the robot navigation once a human is detected. For this, it will be developed specific ROS nodes working simultaneously to SLAM system, which is being currently developed at the laboratory of Military Institute of Engineering.

The people detection system is implemented as a C++ application based on the Point Cloud Library (PCL) (Rusu and Cousins, 2011). It is embedded in a robotic platform able to perform SLAM, allowing that the exploration and navigation of the robot can be made considering the interaction with detected people in a semi-structural environment.

Another implementation was applied, in order to evaluate the accuracy of the two variant of HOG descriptor. The original HOG method proposed by (Dalal and Triggs, 2005) and the DPM (Deformable Part Model) HOG variant, proposed by (Felzenszwalb et al., 2010). It was implemented using the MATLAB VLFeat library (Vedaldi and Fulkerson, 2010) and compared with the C++ application embedded in the robot.

The contribution deals with the searching of computational methods enabling robust detection to variations on the position and partial occlusions, allowing its use in unstructured environments.

Besides introduction, this paper is also composed of the following sections. In Section II is described the implementation of people detection system. Furthermore, it will also be described comparative tests and results to evaluate its accuracy. Section III consists of the description of the robot hardware composition and ROS nodes system integration. In Section IV is presented a real experiment that uses the robotic platform described. Finally, in Section V are presented some final considerations.

2 People Detection

The people detection system was implemented using two different libraries. The Point Cloud Library (PCL)(Felzenszwalb et al., 2010) and the MATLAB VLFeat library (Vedaldi and Fulkerson, 2010).

2.1 PCL People Detection System

The PCL people detection system implementation has been adapted from the Ground Based People Detection App application available on the PCL. In this paper it was adapted the presented proposal in (Munaro and Menegatti, 2014), implementing ROS nodes that integrate with SLAM system robot. Figure 2 provides an overview of the system stages. In accordance with figure, as the robot starts exploring the environment, it extracts RGB-D images that are processed by the people detection node, returning the person's pose (θ_H, X_H, Y_H) in relation to the robot's pose (θ_R, X_R, Y_R) and to the map, that is being generated by the SLAM process.

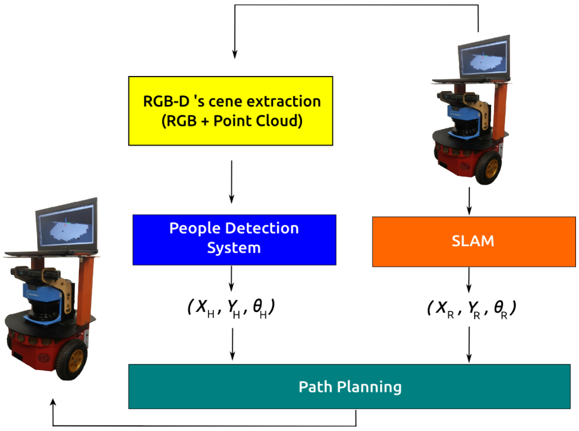

Figure 2: System Overview

The current detection process implemented is based on HOG descriptors, using a labeled dataset from the Kinect Tracking Precision (KTP) Dataset [1] only for SVM classification process. It is important to highlight that the training stage was not processed in our implementation.

During the HOG experimentation, it was verified, as shown in figure 3, a satisfactory real time multiple detections, even with partial occlusions in structured environments.

Figure 3: Multiple and partial occlusion people detection

However, the system was not robust to unstructured environments with many false positives detection. Moreover, another characteristic is that the system presents the detection of people in upright position, with a frequent loss when different

[1] http://www.dei.unipd.it/ munaro/KTP-dataset.html

ISSN 2525-8311

positions from the upright. Figure 4a demonstrates the loss of detection when person lightly starts to bend. Furthermore, on figure 4b is presented a false detection when the systems recognize a squeegee as a person. This fact continuously occurs in unstructured environments.

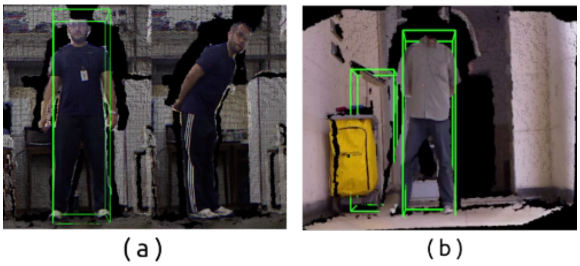

Figure 4: People Detection Loss (a) and False positives (b)

2.2 VLFeat People Detection System

The Histogram of Oriented Gradients (HOG) was implemented by the MATLAB VLFeat library, as in PCL system for people detection. However, the VLFeat library implements the original Dalal and Triggs HOG method and the DPM variant. In both methods, the image is divided into square cells typically with 8 pixels. Posteriorly, the gradient $\nabla l(x, y)$ is calculated and attributed to one of the orientations in the range of 0 to 2π. The gradients magnitudes are accumulated using bilinear interpolation of 4 neighbors cells, equally to the SIFT (Scale Invariant Feature Transform) descriptor (Lowe, 2004).

The first oriented histogram is generated taking into account gradient's direction and orientation. The quantity of directed orientations is correspondent to $\frac{K\pi}{N_{or}}$, where N_{or} is the number of orientations and K is an index that varies from $0, ..., 2N_{or} - 1$. A second histogram is obtained in the range of 0 to π, without taking into account the gradient orientations.

For the Dalal and Triggs variant implemented on VLFeat library, each oriented histogram is copied 4 times, normalized and finally stored as a descriptor vector of the cell with dimension 4 times the number of orientations. Typically, $4 * 9$ bins $= 36$. In DPM model variant, the same is done but using the two types of histograms (oriented and non-oriented), resulting in a descriptor vector with dimension of $(4 * (2 + 1) * N_{or})$ nevertheless it is decreased store by averaging of the corresponding histograms, resulting in a vector of $(2 + 1) * N_{or}$. Additionally, the normalization of non-oriented histograms is calculated and stored as additional 4 dimensions, totaling $(4 + 3 * N_{or})$ dimensions, providing an improved representation of descriptors elements.

2.3 Comparative Tests and Results

The comparative tests were performed in order to evaluate the accuracy of HOG variants implemented with the VLFeat library, compared to the PCL Library implementation. For the tests, we used different scenarios extracted from the KTP dataset and our own dataset acquired from images reproducing the KTP dataset scenarios and situations.

Firstly, the scenario 1 demonstrates a person walking forward and backward directions. For this, it was performed 3 tests, the first scenario was captured by the robot's RGB-D camera and the people detection was made by the PCL library implementation using the labeled KTP dataset; the second and third scenarios were captured by an RGB camera and the people detections were respectively made by the VLFeat DalalTriggs original HOG implementation and DPM HOG variant.

The second scenario revealed three people walking randomly. As seen in the previous test, people detection system was evaluated with PCL library implementation using labeled KTP dataset and VLFeat DalalTriggs original HOG implementation and DPM HOG variant.

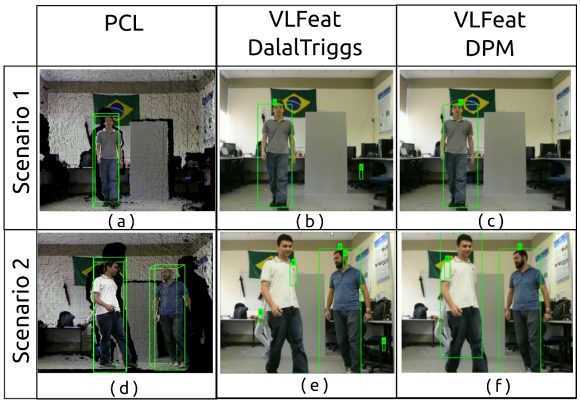

Figure 5: Scenario 1: PCL People Detection (a) DalalTriggs original HOG (b) and DPM HOG variant (C); Scenario 2: PCL People Detection (d) DalalTriggs original HOG (e) and DPM HOG variant (f)

Figure 5 illustrates the scenarios of the tests. Figures a, b and c shows the first scenario, and the figures d, e and f shows the second scenario. In both scenarios, it was noticed an improved accurateness of PCL implementation, followed by Matlab VLFeat DPM HOG Variant and DalalTriggs original HOG. Table 1 indicates the accurateness results of the implementations.

Scenarios 3 and 4 utilized images from KTP dataset in two different situations, as in scenarios 1 and 2, respectively. Nevertheless, for these tests, the KTP dataset was used to train the learning algorithms of the people detection system implemented with the Matlab VLFeat library.

Table 1: Accuracy results of scenarios 1 and 2

Scenarios	PCL	VLFeat DalalTriggs	VLFeat DPM
1	87%	69%	73%
2	70%	54%	61%

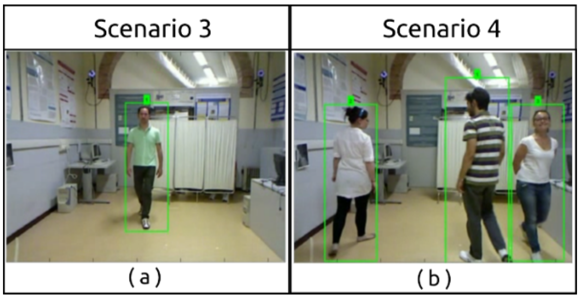

Figure 6: (a) Scenario 3: KTP dataset with one person walking forward and backward; (b) Scenario 4: KTP dataset with three people randomly walking.

Figure 6 illustrates scenarios of the tests 3 and 4. Figure a, illustrates the first scenario, and figure b, the second one. The table 2 shows the accurateness results of the implementations.

Table 2: Accuracy results of scenarios 3 and 4

Scenarios	VLFeat DalalTriggs	VLFeat DPM
3	64%	70%
4	58%	64%

2.4 *Test and results considerations*

Concerning efficacy, PCL People Detection System exhibited better accuracy compared to VLFeat, because as described in KTP dataset, the ground truth of people position was acquired using a BTS motion capture system [2] for the training process.

3 Robot System

In this section it is described the robot's platform and the ROS nodes components developed for the integration of the people detection system and the SLAM system.

3.1 *Robot Plataform*

The robot illustrated in figure 7 is composed of the described components as it follows:

- Robotic Platform Pioneer 3DX: It is a widely used platform in research related to mobile robots. It is composed of encoders on wheels, a battery and a set of sonar already embedded. Additionally, it enables the angular and linear speeds control.

- Laser Sick LMS200: It is used for the 2D feature extraction during the SLAM process.

- RGB-D Camera: Besides being used in SLAM process, it is the principal component of scenes capture that will be used in people detection.

- Computer: Core I7 notebook with robot operating system installed for SLAM applications processing and people detection.

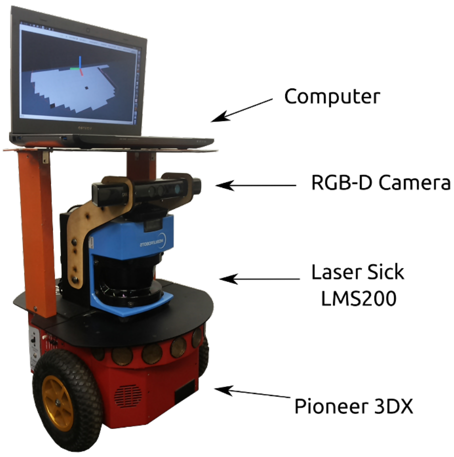

Figure 7: Robotic Platform

3.2 *People Detection and SLAM ROS Nodes Components Integration*

Robot operating system ROS integrates the applications of People Detection and SLAM. As suggested in figure 8, the physical devices as the sensor laser and the RGB-D camera provides information to SLAM applications and People Detection trough ROS which implements an efficient exchange messaging system. It allowed us to integrate People Detection to SLAM system with the minimum of possible adaptations. Another interesting characteristic is that ROS implements a hardware abstraction set, providing the possibility of sensors substitution or even the entire robotic platform without any significant impact on applications already developed. It allows us to continuously maintain an updated system, testing more efficient and robust components.

The ROS node components for people detection and SLAM system are illustrated in figure 9. Each of the robot hardware devices has an abstraction on ROS, as showed in the figure. The

[2]http://www.btsbioengineering.com/

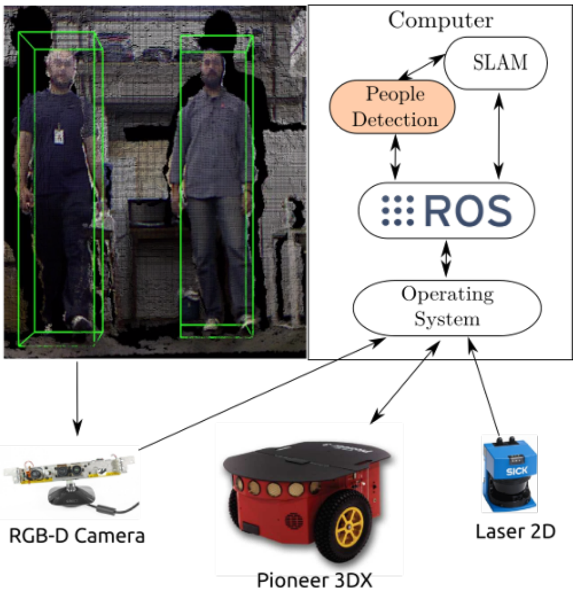

Figure 8: ROS Components Overview

orange rectangles illustrate the physical parts coupled to the robot. Odometry, Left Motor, and Right Motor are components of the robot pioneer described in this paper. The RGB-D Camera and the Laser Sensor are components coupled to the robot and are responsible for visual sensing and features extraction, respectively. The blue ellipses are ROS abstractions related to physical components. The topics, illustrated by the blue rectangles, are responsible for the messages coming from the ROS nodes, specifically the encoders for odometry, laser and camera RGB-D, that feeds the SLAM application node and People Detection node. The people detection system uses ROS components that are highlighted by the yellow rectangle, which the main sensor is the RGB-D camera, responsible for scene extraction detection. After people detection processing, a message containing the people's pose (X_H, Y_H, θ_H) is transmitted to the path planning node that along with the robot's pose (X_R, Y_R, θ_R) are processed and returned to Pioneer node for robot move control.

4 Experiments

As a way to evaluate the integration of detection system with the robot operation in real time, it was implemented a simple follower application in which whether detecting a person, the robot turns and it follows the detected person, maintaining a 50cm distance. The experiment was conducted at the Military Institute of Engineering facilities in a semi-structured environment. In figure 10, it is represented the robot following a person and the person detected HOG features. The application of the follower robot comprises the following steps. People detection using HOG implementation method described in PCL library.

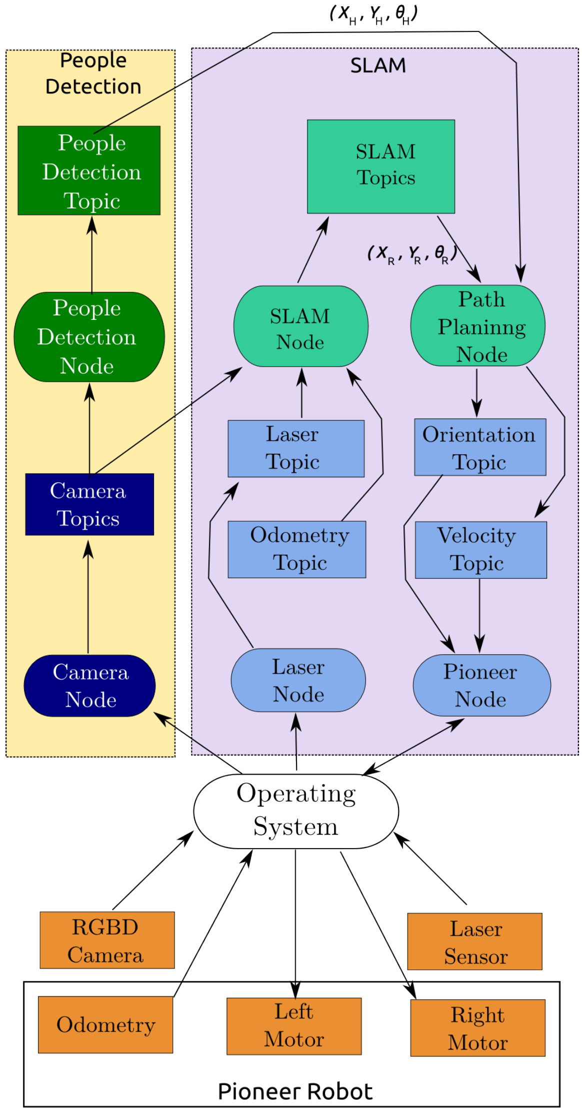

Figure 9: People Detection and SLAM ROS nodes components

After positive detection of a person, it is calculated the angle $\theta = \tan^{-1}\left(\frac{y_H - y_R}{x_H - x_R}\right)$ and distance $d = \sqrt{(x_H - x_R)^2 + (y_H - y_R)^2}$ from the center of the person detected in relation to the RGB-D sensor coupled to the robot.

Posteriorly, it is calculated the robot angular velocity $w = \theta - (\frac{\pi}{2})$. The following step verifies whether the angle $\theta \approx 90$ degrees and distance $d \geq 50$ cm, if yes it is attributed a constant linear velocity of $10cm/s$ on the robot x axis, moving the robot direction towards the detected person. Otherwise, it is attributed a linear velocity $= 0$ and angular velocity w, turning the robot around $\theta \approx 90$ degrees.

The implantation of the follower application presents a simple setting. However, it supported to evaluate the detection process functioning in real time jointly with the execution of SLAM application.

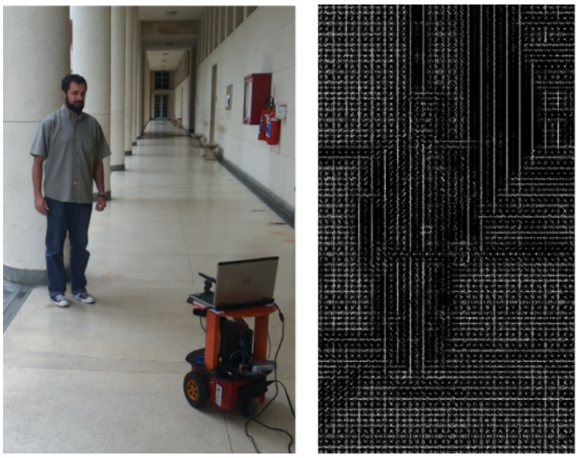

Figure 10: Robot folowing a people and HOG detection

5 Conclusion and Future Work

This paper presented a people detection system for a robotic platform capable of performing SLAM, where the detection method is based on HOG descriptor. The system was implemented in a robot consisting of a 2D laser sensor and an RGB-D camera coupled to robotic platform Pioneer 3DX, using the robot operating system ROS. The application of detection was developed as a ROS node that behaves seamlessly with SLAM application nodes allowing that the robot movement model starts to consider people detected. The test results of people detection application embedded on the robot demonstrated that HOG descriptors are efficient for people detection in a real-time. However, the system is not robust to unstructured environments and detection of positions different from the upright.

As further steps of this research, we propose the utilization of a deep learning architecture based on a particular model of a convolutional neural network (CNN) to improve the process of people recognition in extracted images by an RGB-D sensor attached to the robot.

6 Acknowledgements

The authors gratefully acknowledge the financial support from *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (**CAPES**)* – Edital *Pro-Estratégia.*

References

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, IEEE, pp. 886–893.

de Melo Neto, A., Rosa, P. F. F., de Oliveira, T. E. A. and Pellanda, P. C. (2012). Multiple Robots in a Cooperative Task: Exploration and Mapping., *IROS12 4th International workshop on Planning, Perception and Navigation for Intelligent Vehicles*, Vilamoura, pp. 79–86.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D. and Ramanan, D. (2010). Object detection with discriminatively trained part-based models, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(9): 1627–1645.

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

Munaro, M. and Menegatti, E. (2014). Fast rgb-d people tracking for service robots, *Autonomous Robots* **37**(3): 227–242.

Parizi, S., Oberlin, J. and Felzenszwalb, P. (2012). Reconfigurable models for scene recognition, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2775–2782.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R. and Ng, A. Y. (2009). Ros: an open-source robot operating system, *ICRA workshop on open source software*, Vol. 3, p. 5.

Rusu, R. B. and Cousins, S. (2011). 3d is here: Point cloud library (pcl), *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, IEEE, pp. 1–4.

Silveira Vidal, F., de Oliveira Palmerim Barcelos, A. and Ferreira Rosa, P. (2015). Slam solution based on particle filter with outliers filtering in dynamic environments, *Industrial Electronics (ISIE), 2015 IEEE 24th International Symposium on*, pp. 644–649.

Silveira Vidal, F., Ferreira Rosa, P. F., de Melo Neto, A. and Alves de Oliveira, T. E. (2013). Cooperating robots for mapping tasks with a multilayer perceptron, *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, IEEE, pp. 4073–4078.

Vedaldi, A. and Fulkerson, B. (2010). Vlfeat: An open and portable library of computer vision algorithms, *Proceedings of the international conference on Multimedia*, ACM, pp. 1469–1472.