International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 6, December 2020, pp. 6292~6299
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i6.pp6292-6299

Object gripping algorithm for robotic assistance by means of deep leaning

Robinson Jiménez-Moreno, Astrid Rubiano Fonseca, José Luis Ramírez
Faculty of Engineering, Militar Nueva Granada University, Colombia

Article Info

Article history:

Received Jul 17, 2019
Revised May 28, 2020
Accepted Jun 7, 2020

Keywords:

CNN regression
Convolutional network
Object gripping
Robotic
Virtual environment

ABSTRACT

This paper exposes the use of recent deep learning techniques in the state of the art, little addressed in robotic applications, where a new algorithm based on Faster R-CNN and CNN regression is exposed. The machine vision systems implemented, tend to require multiple stages to locate an object and allow a robot to take it, increasing the noise in the system and the processing times. The convolutional networks based on regions allow one to solve this problem, it is used for it two convolutional architectures, one for classification and location of three types of objects and one to determine the grip angle for a robotic gripper. Under the establish virtual environment, the grip algorithm works up to 5 frames per second with a 100% object classification, and with the implementation of the Faster R-CNN, it allows obtain 100% accuracy in the classifications of the test database, and over a 97% of average precision locating the generated boxes in each element, gripping successfully the objects.

Corresponding Author:

Robinson Jiménez-Moreno,
Department of Mechatronics Engineering,
Militar Nueva Granada University,
Cra. 11 No. 101-80, Bogotá D.C., Colombia.
Email: robinson.jimenez@unimilitar.edu.co

1. INTRODUCTION

Currently, deep learning (DL) techniques present greater robustness than other types of techniques in pattern recognition in both temporary signals and image analysis [1]. An example of this, in [2], AlexNet presents a convolutional neural network (CNN), which surpasses all the classic Machine Learning techniques implemented so far in the ImageNet challenge [3], obtaining a TOP 5 error of 16.4% in the classification of more than one million images in a thousand different categories. In [4], the CNNs are implemented by regression to estimate the trajectory distance of a robotic arm that allows grabbing an element of the environment. However, CNN can be used for several kinds of signal like speech recognition [5] or correlated feature data analysis [6].

At the same time, several CNN-based DL techniques allow the detection of objects in an image, among them the DAG-CNN [7], the R-CNN [8], fast R-CNN [9], Yolo [10] and faster R- CNN [11]. The latter has been implemented in various investigations, an example of which in [12], it is implemented for the detection of vehicles in real-time, obtaining an accuracy between 85%-95%. In [13], a Faster R-CNN is used for the detection of occluded objects for unmanned aerial vehicles, in which the performance of this network is compared with other CNN architectures, obtaining that the Faster R-CNN reached a greater accuracy in the detection of the elements with 83.9%. The R-CNN allows to identify an object in a region of the image, the fast increases the speed in the processing of the network and the faster optimizes said performance. This makes it possible to identify that the faster R-CNN is the best DL option for the detection and identification of elements in a region-based environment, as is the case proposed. To estimate the rotation angle of an element, there are techniques based on CNNs with a regression layer, for example, in [14], the authors implement a CNN regression to return the grip coordinates from an RGB-D camera.

Another example is presented in [15], where the CNN regression is used to estimate the joint velocities that a robot must have to launch and catch objects.

On the other hand, DL techniques are integrated into the control of robots in aspects based on object recognition [16] and allow applications oriented to the grip of objects, for example, the work presented in [17], where an application of conventional use for garment grip is oriented. More specialized applications involve human-robot interaction [18], where the DL is used to identify the intention of movement. For applications that involve the use of robots in integrated environments as assistants, it is necessary to identify the object of interest and grasp it, for which CNN has already shown their versatility [19, 20]. This work presents an advance in the use of DL for the recognition and grip of objects in multi-objective environments oriented to assistive robots using recent techniques of variation of conventional CNN architectures such as fast-RCNN and CNN regression [21].

CNN is usually used for detecting grasping objects with final robotics effectors with a tweezer form, as it is exposed in [22, 23]. But it presents an unstable grip requiring additional time to find the best way to grip, close to the gravity center to the object. This article presents the development of a new algorithm based on Faster R-CNN and CNN regression to provide to robotic agents, equipped with three-finger grippers, the ability to grasp objects that are in the environment, increasing the stability, but its training implies an additional complexity, reason for use both of the CNN kinds. This problem emerges as a necessity to integrate an efficient grip method with previous works based on human-robot interaction [24] with assistant robots [25].

This article is divided into three sections. The first section presents the environment of the application that focuses on the acquisition and adaptation of the databases and the neural architecture implemented for the detection of the elements to be grasped. In the second section, the graphic user interface that facilitates the acquisition of databases, training of networks and results obtained in a virtual environment are exposed. Finally, the conclusions of the developed system and possible improvements for future work are presented.

2. RESEARCH METHOD

To evaluate a grip algorithm using DL using CNN, 3 objects are established in a virtual environment. Since the aim is to use a gripper for a robotic agent, the characteristics of the grip object must be defined. To generalize the geometries, two types of objects are proposed; the first type have infinite symmetry axes, where two objects with that characteristic are established: a cylinder and a toroid; the second type is defined with a finite number of symmetry axes, using a parallelepiped geometry. By having elements with a finite number of axes of symmetry, their rotation could affect the way it grabs when changing their orientation with respect to the Z axis, requiring a test of this type. With the defined objects, two databases are established: the first one, to train a network for its detection and the second one, to estimate the angle at which the parallelepiped is rotated.

2.1. Database for networks training

In Figure 1, some examples of the database for detection and localization are shown, in which the position and orientation of the elements in the environment are changed. A total of 2195 RGB images with a resolution of 224x224 pixels are established, of which 200 images are separated for evaluation after training and the rest are used to train the network.

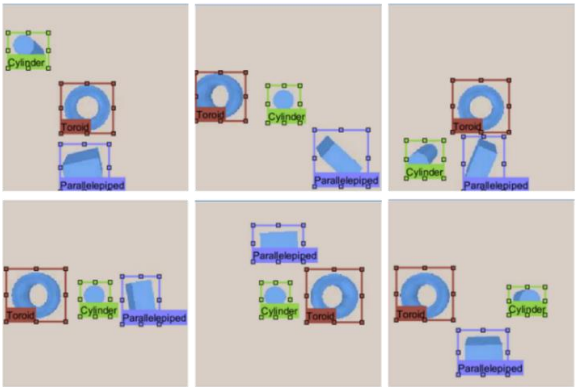

Figure 1. Sample of the database for detection and localization of element

For the network trained to estimate the angle of rotation, two databases were established, one RGB and one binary, to evaluate which of the two databases manages to have a better learning. From the previous database, 2000 images were taken, from which the regions where the parallelepiped is located are extracted. Once the regions are obtained, edges are added to the images of both databases to make the image square. This is done since all the images must be resized to the same size, because network input size cannot be variable. It is set a size of 50x50 pixel for the images, which encompasses the size of the parallelepiped as shown in Figure 2. From 2000 images, 10% of them, corresponding to 200 images, are used for tests after training.

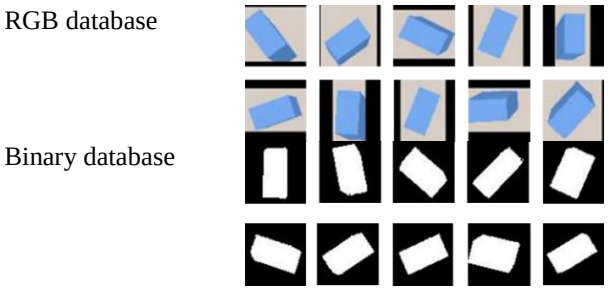

Figure 2. Proposed data bases for angle estimation

2.2. DL architectures

The proposed architecture is divided in two parts. In the first part, a Faster R-CNN is implemented to detect and to locate the elements in the environment. In the second part, an architecture proposed for a CNN with a regression layer that allows estimating the angle to which the parallelepiped is rotated as shown in Figure 3. The Faster R-CNN, unlike a conventional CNN, has an RPN (region proposal network), with which frames, called Anchors, are generated in the image, where it is identified in which an object could exist. The learned characteristics are linked with a RoI-Pooling, then passes through Fully Connected layers that allow to generate a learning on the extracted characteristics and finally, to identify in each detected object to which category it belongs. As an architecture for the Faster R-CNN, the VGG16 network [26] is implemented. Then, from the regions detected by the Faster R-CNN, only the one corresponding to the parallelepiped is extracted and adjusted as indicated in the database for the CNN regression. This architecture consists of four convolutional layers that will learn the features of the parallelepiped, to be then entered into the Fully Connected layers and finally, with the regression layer, to estimate the angle. Given the geometry of the object, the angle of rotation will be between 0° and 180°, from this the maximum angle is set at 175°, taking the range between 175° and 180° equal to zero, that is, variations of at least 5° of error will be validated.

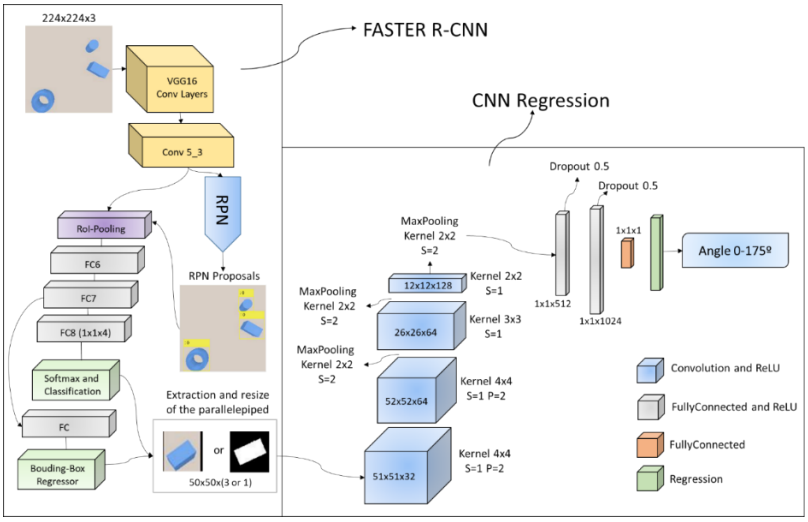

Figure 3. Proposed architecture for the detection of elements in the environment and calculation of the orientation angle of an element with finite number of symmetric axes

3. RESULTS AND ANALYSIS

3.1. Networks results

Faster R-CNN is trained with the training images, the confusion matrix is calculated in order to be able to demonstrate its performance in learning as shown in Figure 4. In Figure 4(a) and (b), the confusion matrices of the training and testing database are presented. The green diagonal represents the images that were correctly classified and out of this those that were classified in an incorrect category. Both the training database and the test database showed 100% accuracy in the classification of the images.

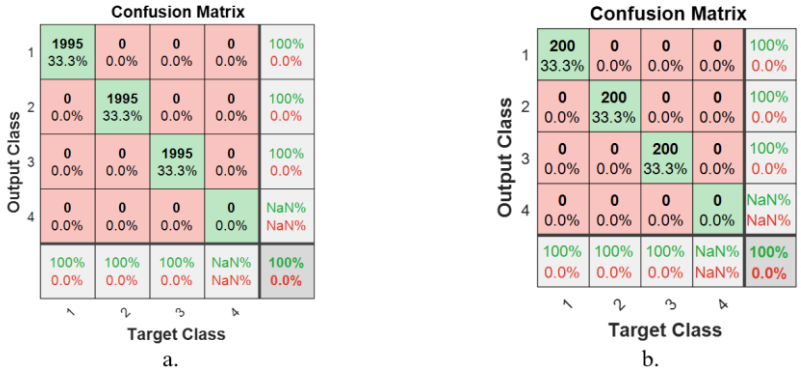

Figure 4. Confusion matrices for the training database, (a) and tests and (b) where 1 = cylinder / 2 = paralelepiped / 3 = toroid / 4 = background

Another factor to take into account when performing a system for the detection of elements in images is the Average Precision. Since this will indicate the precision of the overlap between the boxes that generate the network and those set in the ground truth. For both databases (training and test), values greater than 97% accuracy are presented, i.e. the boxes generated by the network are reliable and will allow the robotic agent to move correctly to grab them.

In Figure 5, some examples of classification and detection of the Faster R-CNN are shown. It can be seen cases in which the elements to be identified are partially obstructed and even so the network is able to detect them although with less reliability. The network is tested by placing strange objects that are unknown, identifying that, although there are elements similar to those trained, no false positives are presented.

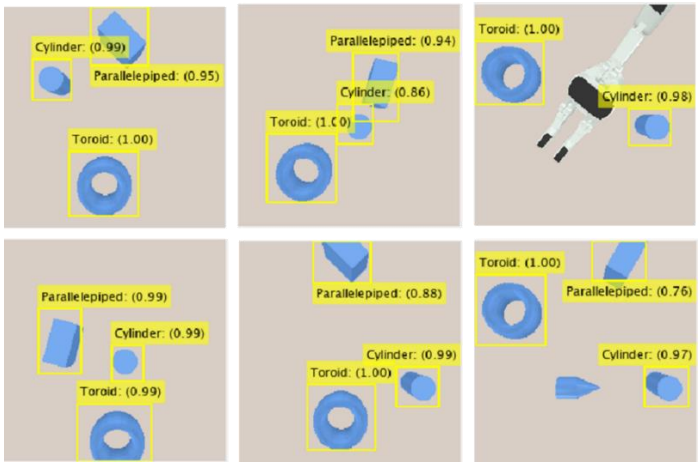

Figure 5. Tests of the trained network

For the CNN regression, through an iterative process, the training parameters of the network were established, obtaining a batch size for the training of 100, a learning factor of 1x10-6 and 100 epochs of training, allowing the network to estimate an approximate angle to the real one without going into overfitting. In Figure 6, the box diagrams for the RGB and Binary database are presented. The binary database obtained

19 outliers, while the RGB database only 10. Another factor to highlight is the mean error for both bases, for the binary it corresponds to 1.049º and for the RGB, 0.769º. The binary database also showed a greater error in the range of non-atypical values, being between 20.23º and -16.54º, while in the RGB images, of 8.49º and -11.76º. Therefore, it can be concluded that the RGB database presented a better performance than the binary one.

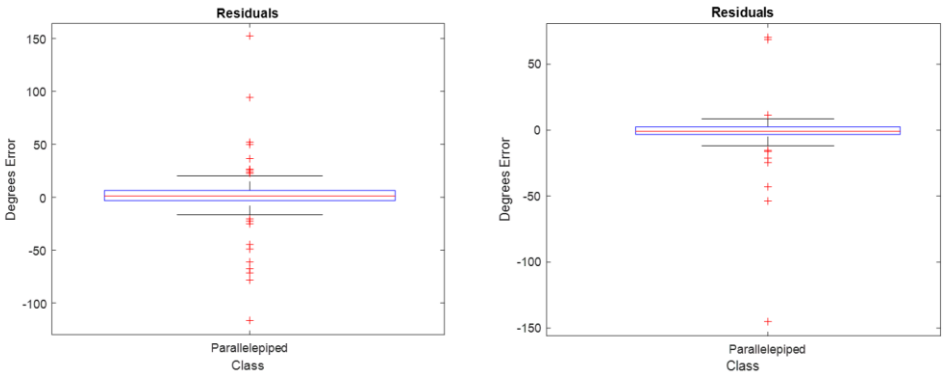

Figure 6. (a) Boxplot of the test database of one channel (binary) and (b) three channels (RGB)

In Figure 7, some examples of angle estimation with RGB input images are shown. It can be seen that the estimate agrees with the results calculated in the box diagram. Although in different positions and orientations of the object in the environment the error in the calculation of the angle is low, there are cases like the one presented in Figure 8, where there are some positions and angles of orientation that can generate atypical results.

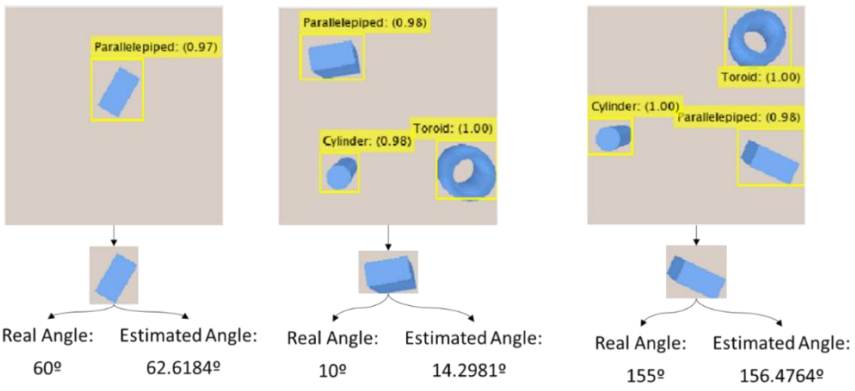

Figure 7. Examples of orientation angle estimation with respect to the Z axis of the parallelepiped with RGB images

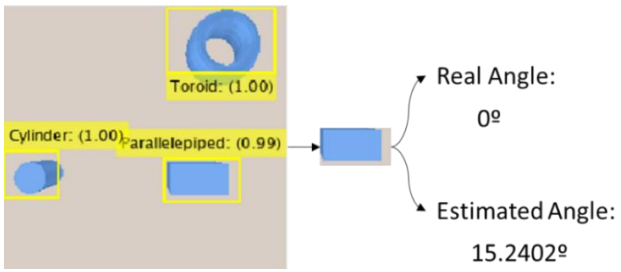

Figure 8. Error in the estimation of the orientation angle of the parallelepiped

3.2. Tests in virtual environment

Nowadays, there are several virtual environments that are used for the simulation of robotic systems, among the main ones are Gazebo and V-REP, where in [27], the main advantages and disadvantages of each are presented. V-REP is selected for its diversity of sensors and coupling with different programming languages. In Figure 9, an example of the virtual environment used is shown, where it can be seen a RRP robot, which has a gripper as final effector, and the three different types of objects in the environment (cylinder, parallelepiped and toroid), which can be visualized with a camera that is incorporated in the gripper. In [28], it is mentioned that using a virtual environment will facilitate the tests of the DL architectures for the detection of the elements. In this way, when implementing them in a real environment, it will not be necessary to train the network from scratch, it will only be necessary a fine tuning for the networks to be coupled to the real environment.

In the virtual environment, the gripper camera runs frame by frame while running the Faster R-CNN and it can reach up to 5fps with a seventh-generation i7 computer with an NVIDIA GTX 960M GPU and 16 GB of RAM. Once the system is running, it can be selected between the options "Cylinder", "Parallelepiped" and "Toroid", in order to tell the robotic agent which item to collect. It is highlighted that under the option of "Parallelepiped" the angle to which the element is rotated is shown. In case of significant errors in this angle, the option "Acquisition" makes a new database with the sections detected by the Faster R-CNN in which the parallelepiped is located, with which a new CNN network can be trained for tasks of regression of the orientation angle of the object as shown in Figure 10.

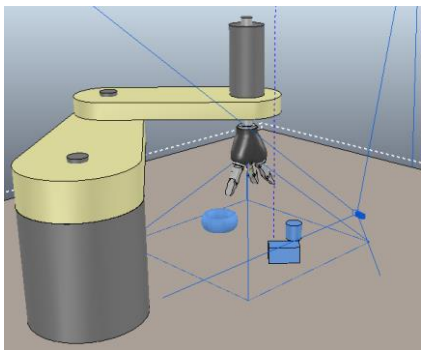

Figure 9. Virtual work environment in V-REP

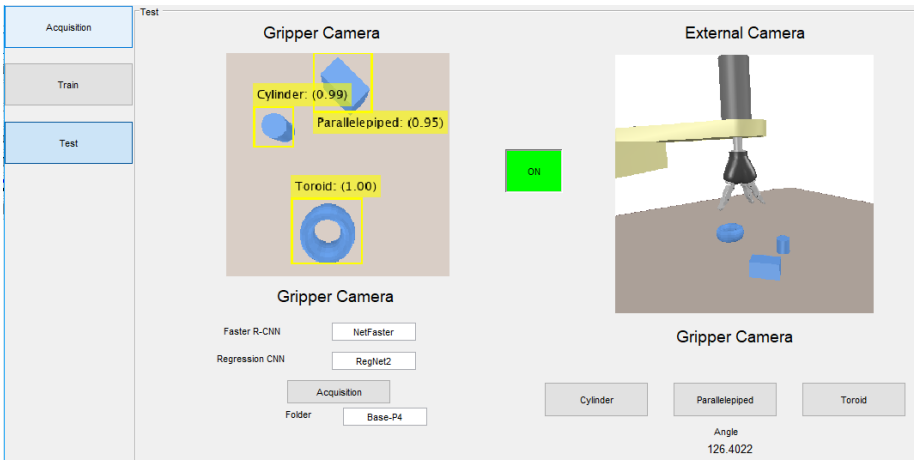

Figure 10. Test section of the networks with the virtual environment in the graphic user interface

In Figure 11, some grip tests are shown for each object with two and three finger grippers. In Figure 11, some grip tests are shown for each object with two and three finger grippers. In general, in the tests carried out, the three-finger gripper allowed a greater margin of error with respect to the two, but in some cases, one of the additional fingers of the three-finger gripper may not perform a useful function during the grip, as shown in Figure 11, when the three-finger gripper grips the parallelepiped.

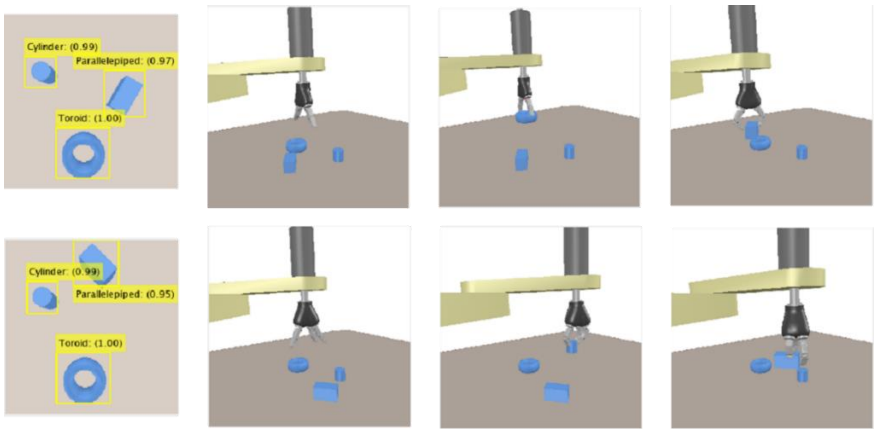

Figure 11. Examples of tests with two and three finger grippers

4. CONCLUSION

The implementation of the Faster R-CNN for the detection of elements of interest in the environment managed to obtain 100% accuracy in the classifications of the test database, and over a 97% of average precision locating the generated boxes in each element, thus allowing the robotic agent to have greater autonomy in the execution of their trajectories towards the objects that are wanted to be collected. The design of a CNN trained for regression tasks allowed to calculate an approximate angle to the real one in most cases, but from the results obtained in the boxplot, it is possible to identify that in case it is wanted to implement a system that requires high precision and accuracy to perform a task, errors in the gripping tasks may appear. For this reason, for future developments, other possible databases will be evaluated to reduce the error obtained.

Taking into account the trajectories made and the grips made by the robotic agent, it is proposed for future developments to calculate, by means of a CNN regression, the coordinates of the approximate points of grip for each of the fingers that the gripper may have, in order that all serve as support in the execution of the task. The comparison between two and three fingers gripping simulated, it stated three fingers gripping allow keep the object, while with two fingers it is gripping on close to the gravity center but not in it, the object is exposed to possible fallen. In every case, the object was successfully grabbed by the proposed algorithm.

ACKNOWLEDGEMENTS

The authors are grateful to the Universidad Militar Nueva Granada, which, through its Vice Rectory for researchs, finances the present project with code IMP-ING-2935 (2019-2020) and titled "Prototipo robótico flexible para asistencia alimentaria," from which the present work is derived.

REFERENCES

[1] Amine Boulemtafes, Abdelouahid Derhab, Yacine Challal, "A review of privacy-preserving techniques for deep learning," *Neurocomputing*, vol. 384, pp. 21-45, 2020.
[2] Krizhevsky, A., Sutskever, I. and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097-1105, 2012.
[3] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., "Imagenet: A large-scale hierarchical image database.," *IEEE conference on computer vision and pattern recognition*, pp. 248-255, 2009.
[4] Viereck, U., Pas, A.T., Saenko, K. and Platt, R., "Learning a visuomotor controller for real world robotic grasping using simulated depth images," *arXiv preprint arXiv:1706.04652*, 2017.
[5] Pinzón-Arenas Javier Orlando, Jiménez-Moreno Robinson, "Comparison between handwritten word and speech record in real-time using CNN Architectures," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 4313-4321, 2020.
[6] Yanmiao Li, et al., "Robust detection for network intrusion of industrial IoT based on multi-CNN fusion," *Measurement*, vol. 154, 2020.
[7] Jiménez-Moreno R., and Pinzón-Arenas J. Orlando, "Object sorting in an extended work area using collaborative robotics and DAG-CNN," *ARPN Journal of Engineering and Applied Sciences*, vol. 15, no. 2, pp. 192-202, 2020.
[8] Girshick, R., et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition.*, pp. 580-587, 2014.
[9] Girshick, R., "Fast R-CNN," *Proceedings of the IEEE international conference on computer vision*, pp. 1440-1448, 2015.

[10] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.
[11] Ren, S., He, K., Girshick, R. and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, pp. 91-99, 2015.
[12] Fan, Q., Brown, L. and Smith, J., "A closer look at Faster R-CNN for vehicle detection," *IEEE intelligent vehicles symposium (IV)*, pp. 124-129, 2016.
[13] Lee, J., Wang, J., Crandall, D., Šabanović, S. and Fox, G., "Real-time, cloud-based object detection for unmanned aerial vehicles," *First IEEE International Conference on Robotic Computing (IRC)*, pp. 36-43, 2017.
[14] Watson, J., Hughes, J. and Iida, F., "Real-world, real-time robotic grasping with convolutional neural networks," *Annual Conference Towards Autonomous Robotic Systems*, pp. 617-626, 2017.
[15] Zeng, A., Song, S., Lee, J., Rodriguez, A. and Funkhouser, T., "TossingBot: Learning to Throw Arbitrary Objects with Residual Physics," *IEEE Transactions on Robotics*, 2019.
[16] Xi Chen, Jan Guhl, "Industrial Robot Control with Object Recognition based on Deep Learning," *Procedia CIRP*, vol. 76, pp. 149-154, 2018.
[17] Enric Corona, Guillem Alenyà, Antonio Gabas, Carme Torras, "Active garment recognition and target grasping point detection using deep learning," *Pattern Recognition*, vol. 74, pp. 629-641, 2018.
[18] Zitong Liu, et al., "Deep Learning-based Human Motion Prediction considering Context Awareness for Human-Robot Collaboration in Manufacturing," *Procedia CIRP*, vol. 83, pp. 272-278, 2019.
[19] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1316-1322, 2015.
[20] Zhichao Wang, Zhiqi Li, Bin Wang, Hong Liu., "Robot grasp detection using multimodal deep convolutional neural networks," *Advances in Mechanical Engineering*, vol. 8, no. 9, 2016.
[21] Garima Devnani, et al., "Performance Evaluation of Fine-tuned Faster R-CNN on specific MS COCO Objects," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 2548-2555, 2019.
[22] Z. Wang, Z. Li, B. Wang, H. Liu, "Robot grasp detection using multimodal deep convolutional neural networks," *Advances in Mechanical Engineering*, vol. 8, no 9, 2016.
[23] G. Ghazaei, et al., "An exploratory study on the use of convolutional neural networks for object grasp classification," *2nd IET International Conference on Intelligent Signal Processing (ISP)*, pp. 1-5, 2015.
[24] Jiménez-Moreno Robinson, Pinzón-Arenas Javier Orlando, "New Hybrid Fuzzy-CNN Architecture for Human-Robot Interaction," IREACO, vol. 12, no. 5, 2019.
[25] Pinzón-Arenas J. Orlando, et al. , "Assistant robot through deep learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 1053-1062, 2020.
[26] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[27] Nogueira, L., "Comparative analysis between gazebo and v-rep robotic simulators," *Seminario Interno de Cognicao Artificial-SICA*, 2014.
[28] Tai, L., Paolo, G. and Liu, M., "Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 31-36, 2017.

BIOGRAPHIES OF AUTHORS

Robinson Jiménez Moreno was born in Bogotá, Colombia, in 1978. He received the Engineer degree in Electronics at the Francisco José de Caldas District University - UD - in 2002, respectively. M.Sc. in Industrial Automation from the Universidad Nacional de Colombia - 2012 and PhD in Engineering at the Francisco José de Caldas District University - UD. He is currently working as a Professor in the Mechatronics Engineering Program at the Nueva Granada Military University - UMNG. He has experience in the areas of Instrumentation and Electronic Control, acting mainly in: Robotics, control, pattern recognition and image processing.

Rubiano Astrid, Bogotá Colombia, PhD. degree in mechatronics, especiality in control of soft robotics, Nanterre University, Paris, France, 2016. M.Sc. degree especiality in automatics control systems, Tecnologic University of Pereira, 2012, Bachelor degree in Mechatronic Engineering, Nueva Granada University, 2006. She has publications related to features extraction from electromyographic signals, control of soft structures, images processing toward control systems. She has 20 patents in technology field, such mechatronics systems applied to medicine. Currently, she is interested in control based on electromyographic signals and control of soft bodies applied to robotic.

Ramirez Jose Luis, Bogotá Colombia, PhD. degree in mechanics, especiality in artificial muscles based on smart materials, Nanterre University, Paris, France, 2016. M.Sc. degree especiality in automatics control systems, Tecnologic University of Pereira, 2012, Bachelor degree in Mechatronic Engineering, Nueva Granada University, 2006. He has publications related to artificial muscles, modeling of smart materials, control of shape memory alloys, in among others. Currently, he is interested in smart material applied to soft robotics, dynamic of soft robots, modeling of smart structures.