

Visual control system for grip of glasses oriented to assistance robotics

Robinson Jiménez-Moreno, Astrid Rubiano, José L. Ramírez

Faculty of Engineering, Militar Nueva Granada University, Colombia

Article Info

Article history:

Received Jan 22, 2020

Revised May 28, 2020

Accepted Jun 10, 2020

Keywords:

Assistive robotics

Faster R-CNN

Fuzzy system

Glass grip

Visual control

ABSTRACT

Assistance robotics is presented as a means of improving the quality of life of people with disabilities, an application case is presented in assisted feeding. This paper presents the development of a system based on artificial intelligence techniques, for the grip of a glass, so that it does not slip during its manipulation by means of a robotic arm, as the liquid level varies. A faster R-CNN is used for the detection of the glass and the arm's gripper, and from the data obtained by the network, the mass of the beverage is estimated, and a delta of distance between the gripper and the liquid. These estimated values are used as inputs for a fuzzy system which has as output the torque that the motor that drives the gripper must exert. It was possible to obtain a 97.3% accuracy in the detection of the elements of interest in the environment with the faster R-CNN, and a 76% performance in the grips of the glass through the fuzzy algorithm.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Robinson Jiménez-Moreno,
Mechatronics Engineering Program, Faculty of Engineering,
Militar Nueva Granada University,
Carrera 11 #101-80, Bogotá D.C., Colombia.
Email: robinson.jimenez@unimilitar.edu.co

1. INTRODUCTION

Assistance robotics has gained importance in research in the last decade. An example of this, the international federation of robotics mentions in [1], which is estimated by the end of 2019 that about 31 million robots will perform tasks in homes around the world. In [2], the development of assistant robots stands out, especially in areas related to medicine, where robotic systems are used, among multiple applications, in tasks focused on patient care. Currently, several pathologies and/or musculoskeletal problems affect daily life of people, whether caused by an accident or by their advanced age, causing them to not be self-sufficient in the feeding process, e.g. Guillain – Barré syndrome which affects the peripheral nerves of a person [3], or because of some disease derived from a partial cerebral palsy [4].

Taking into account the possible motor limitations that people can present to feed themselves, various research have focused on the development of robotic support systems to assist them in this process. An example is shown in [5], where a robotic arm that establishes a trajectory to feed a user from an algorithm that is based on the 3d capture of the environment by means of a Kinect sensor is presented. With this, it is possible to identify the position of the user's mouth and a PID controller is implemented where its constants are calculated from a multilayer neural network. In [6], based on electroencephalography (EEG) signals and their analysis by implementing steady state visual evoked potentials (SSEVP), the intentions of the food that the patient wishes to consume are estimated, and by implementing a machine vision system, a cascading Haar classifier is used, the user's mouth is detected so that a robotic agent takes the food to it. In [7], the development of a robotic system to bring the food to the user is presented, which is based on an RGB-D camera and discriminative optimization (DO) method for locating the mouth of the person, managing to operate said system in real time.

In the state of the art, it can be identified that current systems to assist people in the feeding assistance process are supported by machine vision techniques. Nowadays, the most robust machine vision systems implement deep learning (DL) algorithms [8], within which are convolutional neural networks (CNN) [9], that are used primarily for image processing tasks focused on pattern recognition. CNNs were introduced in [10] and initially used for the recognition of handwritten numbers, but given the high computational cost in hardware and software for the equipment of the time, they were not viable for applications with greater complexity. In 2012, in the ImageNet [11], which is a challenge focused on implementing machine vision techniques for the classification of more than one million images, a CNN called AlexNet [12] was presented. This was supported by a GPU for network calculations, reducing training and classification times, and surpassing the traditional machine learning techniques in percentage accuracy. Since then, they have been implemented in various applications, for example, in [13] they are used for the recognition of food products in a refrigerator, obtaining an accuracy of 94% in their classification. In [14], CNNs for the identification of people are implemented, from being trained with face images obtaining a 97.3% accuracy in the detection of more than 2600 people. In [15], CNNs are used for the identification of actions in people from UFC100 database [16], obtaining a 78.76% accuracy in the identification of the action in video sequences.

An important stage that is not commonly addressed in food systems is the supply of beverage to the user. Although there are investigations such as the one presented in [17], which focuses on the detection of the glass and on giving the drink to the user by means of a robot, the amount of liquid that can be in the glass is not foreseen so that the torque that the gripper actuator must exert is also established. The above observation raises a problem that can cause the glass to fall due to lack of exerted torque or, on the contrary, due to excess torque, the structure of the glass is affected. This article presents the development of a system in which two types of artificial intelligence techniques are implemented together, specifically a faster R-CNN [18] and a fuzzy system [19]. With them, it is estimated the torque that a motor must exert for a gripper to make the grip of a glass, in order to be able to lift it without it slipping, or the internal fluid spills. Thus, an alternative to the variation of liquid level is provided when it is supplied to a user by means of an assistance robot, by means of a visual control algorithm.

This article is divided into five main parts. In the first part, a general outline of the process for the manipulation of glass and the established virtual environment is presented. The second part shows the acquisition of the database, the training parameters of the network and the results obtained with it. The third part shows the fuzzy system proposed to estimate the torque that must be exerted by the motor that drives the gripper. In the fourth part, a user interface is presented that allows the acquisition of the image database, the training of the faster R-CNN, the test section and the performance of the system as a whole for the manipulation of the glass. Finally, the conclusions derived from the results obtained are presented.

2. RESEARCH METHOD

In Figure 1, a general scheme of the designed system is shown, where from an RGB-D supervisory camera, the capture of the scene of interest (gripper + glass) is taken. This is used as input for a faster R-CNN, which is trained to detect the existence of the glass with liquid and the gripper, indicating in turn the region of the image where each one is located. With the detection of the position of the glass in the image, its centroid is calculated, and from this value on the camera's depth map, the distance to the fluid is calculated and its mass is estimated. To calculate the torque to be exerted, a fuzzy system is implemented, which receives as inputs the estimated mass and a delta of the distance of the fluid from the gripper. These two parameters were chosen since, depending on the estimated mass, a different torque must be exerted and, if the delta increases, it means that the torque is not sufficient and the glass is slipping when it is held by the gripper.

In the development of the system, the tests are implemented in a virtual environment, in order not to put the user at risk. Currently there are several virtual environments, among the most common are Gazebo [20] or V-REP [21]. In [22], a comparison between each software is described, highlighting V-REP for its physics, allowing simulate situations similar to reality. In Figure 2, the test environment in V-REP is presented, in which a robotic agent equipped with a gripper and an RGB-D camera is shown. It should be noted that the architecture of the robotic agent is focused on facilitating the testing of the system, therefore, once the glass has been grasped, there will only be displacement in the coordinate axis Z.

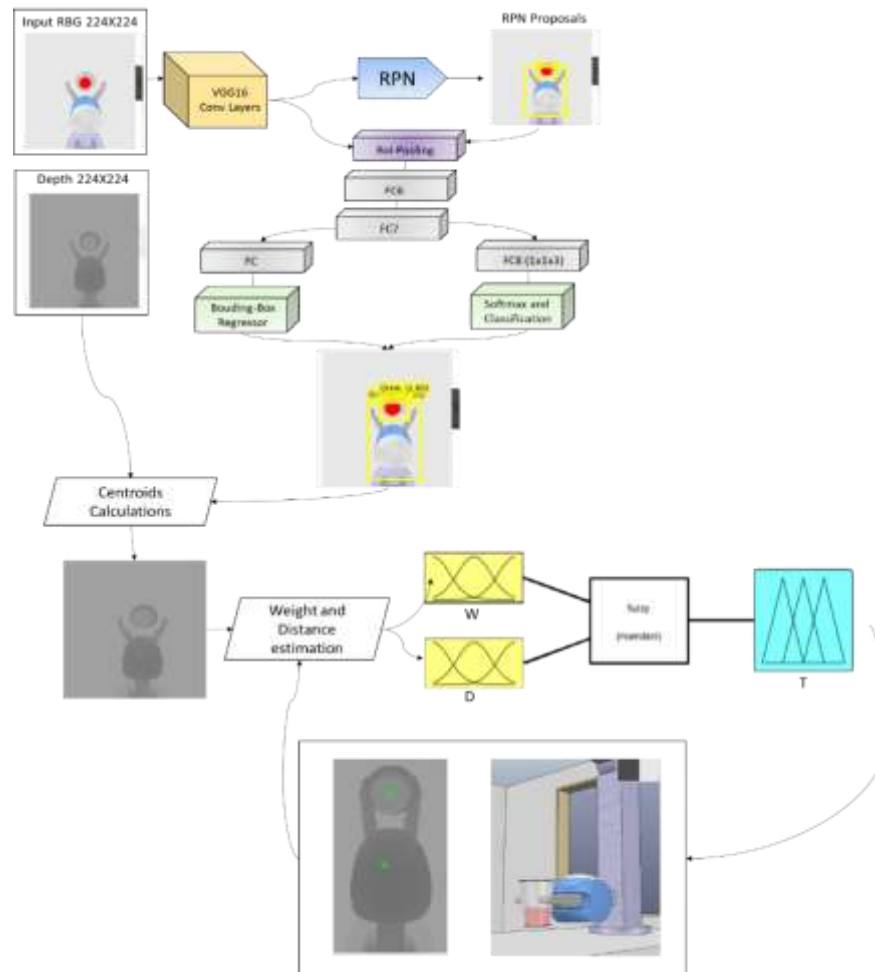


Figure 1. General flowchart of the system designed for the manipulation of a glass with liquid



Figure 2. Virtual test environment for the designed system

3. GLASS AND GRIPPER POSITION DETECTION SYSTEM

Currently, faster R-CNN has gained importance in various research focused on the detection of the position of elements in images. Faster R-CNN, unlike a traditional CNN, generates bounding boxes in the image, where each of these indicates where an object of interest is located and to which category it belongs. To achieve this, the faster R-CNN implements a region proposal network (RPN), which is an additional branch in the architecture. In this, anchors of different sizes are generated throughout the image and it is detected in which of them there may be an object. Then, with the proposed regions detected, the CNN classified to which category it belongs. In [23], a faster R-CNN is implemented for the detection of malaria infected cells, obtaining 72% accuracy in detection. In [24], it is implemented for the detection of small elements in satellite images, obtaining a 78.9% accuracy in the detection of ships and airplanes. In [25],

a faster R-CNN with a VGG-16 architecture [26] is used, where this architecture was one of the winners of the ImageNet. For the training of the network, 2 categories are established (drink and gripper), a database of 1000 images as shown in Figure 3 is built in which the position of the gripper or the glass can vary slightly, in addition, the fluid color and its amount vary. From the total data in the database, 80% equivalent to 800 images are set for training and the remaining for tests after training.

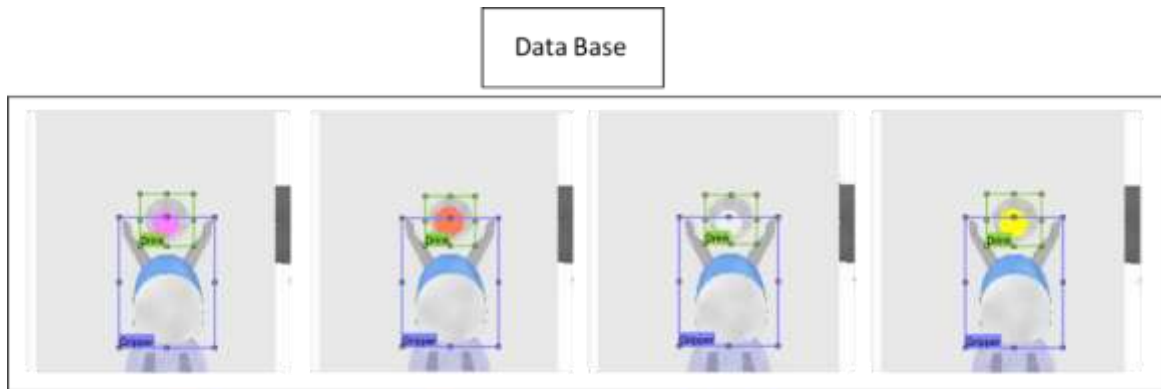


Figure 3. Examples of the database

The training parameters of the network used are presented in Table 1, where these values were chosen based on iterations in the training, identifying which showed a greater accuracy in the classification of the test images. In stage 1 and 2, a more aggressive learning factor and a greater number of epochs than in stage 3 and 4 are set, since in the last ones, a fine tuning is performed between the weights of the RPN and CNN.

Table 1. Training parameters of the faster R-CNN

	Stage 1	Stage 2	Stage 3	Stage 4
<i>Learning Rate</i>	1×10^{-5}	1×10^{-5}	1×10^{-6}	1×10^{-6}
<i>Epoch</i>	30	70	15	40

3.1. Results of the faster R-CNN for the detection of the drink and the gripper

In Figure 4, some examples of the network activations in the convolution layers are shown. These activations represent features that were learned, allowing the detection of the elements and differentiate them from others in the environment. Mainly, there are activations in the detection of the drink, the gripper, edge recognition and from the user that is sit (right side).

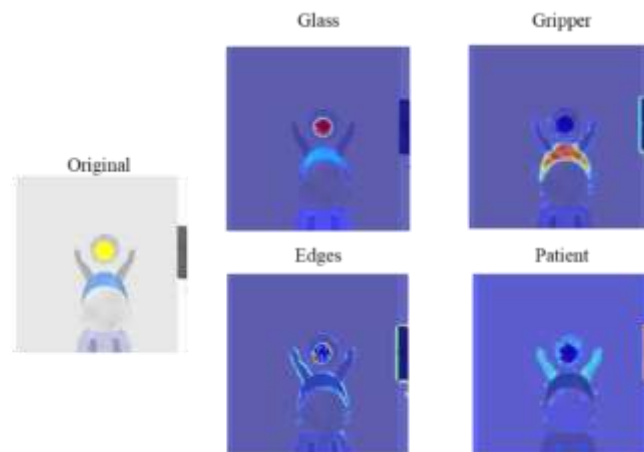


Figure 4. Activations of the faster-RCNN

In Figure 5(a), the confusion matrix of the classification of the test images is shown, where, in the green colored diagonal, the images that were correctly classified are located and, outside this, the incorrect classifications are presented. In Figure 5(b), the recall vs precision graph is shown, where the area under these curves corresponds to the average accuracy (Av. P). This value represents how precise were the regions of interest generated by the network compared to the ground truth. Both results were very close one from another, making the lines look almost overlapping.

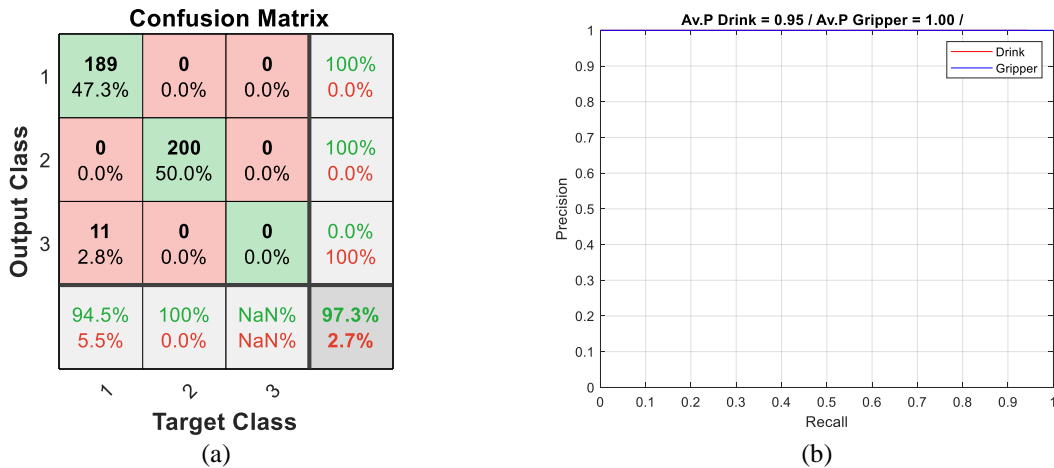


Figure 5. (a) Confusion matrix of the test images (a), where 1 = drink, 2 = gripper and 3 = background, (b) recall vs. Precision graph using the test images

It can be evidenced that 97.3% accuracy was obtained in the classifications of the test images, with average precision of 100% and 95% for gripper and drink, respectively. This favors that the calculation of the centroids from the regions of interest (RoI) generated are right and an adequate grip of the drink is achieved. Figure 6 shows some examples of the detections and classifications of the network, where the true positives are visualized, which represent the categories that were in the image and were correctly detected. False negatives are also shown, these indicate examples of categories that were in the image but were not detected. Taking into account the false negatives given in the tests, it can be evidenced that these cases only occurred when the amount of fluid particles was low. This does not represent an error in the system, since there is practically no fluid, it would not be necessary for the gripper motor to be operated since there would be no drink to be supplied to the user.

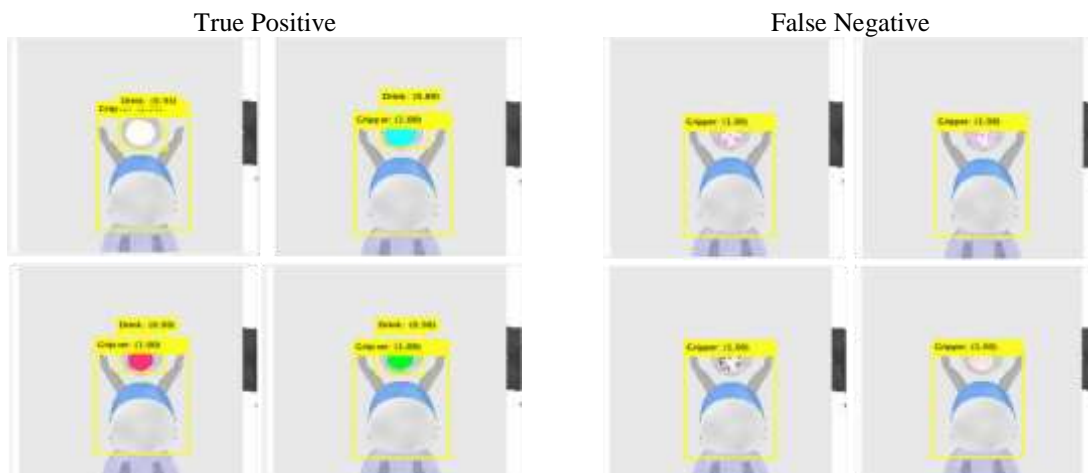
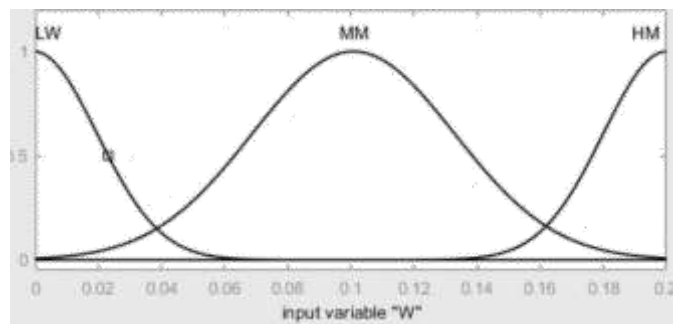


Figure 6. Faster R-CNN classifications for drink and gripper detection

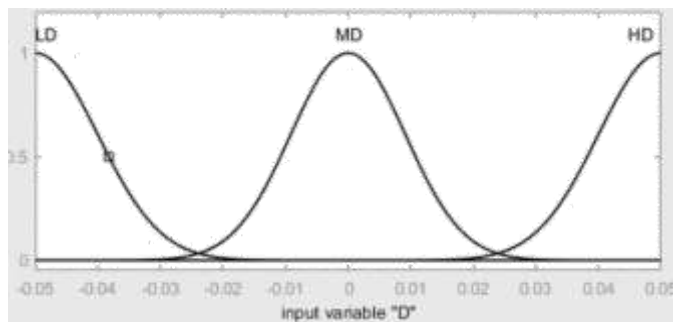
4. FUZZY SYSTEM

Fuzzy systems have been widely used in research in which there is no mathematical certainty of the response of a system, such as highly nonlinear systems. As examples of application associated with this work, in [27], a fuzzy system is implemented together with a CNN to estimate the quality of fruits, taking into account, as system inputs, the estimation of its mass, number of defects and equatorial diameter. In [28], they are used to control the force exerted by a hand prosthesis, concluding that the implementation of fuzzy systems is effective for this type of tasks. In this article, it is decided to implement a fuzzy system, since it is sought to generate a grip control where the weight of the glass varies. Since the feedback of the action is visual, it is not necessary to establish the mathematical model of the general system if visual control techniques are used that usually integrate fuzzy systems [29].

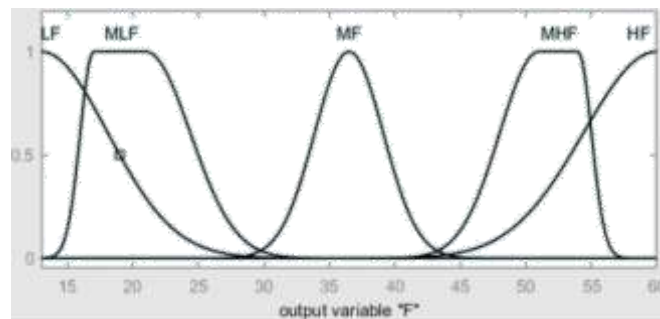
The first input parameter for the fuzzy system is the mass of the fluid. For this, the centroid of the RoI of the drink is calculated and, from the depth map, the height of the liquid in the glass can be calculated and finally, according to its geometry with the known volume, the mass is calculated. The second parameter is a distance delta between the gripper and the fluid. In Figure 7, the membership functions established for fuzzy system inputs and output are shown.



(a) Where $w = \text{mass}$ / LW = low mass / MM = Medium Mass / HM = High Mass



(b) Where $D = \text{distance delta}$ / LD = negative distance delta / MD = neutral distance delta / HD = positive distance delta



(c) Where LT = Low torque / MLT = Medium low torque / MT = Medium torque / MHT = Medium high torque / HT = High torque

Figure 7. (a) Membership functions for the entry of estimated weight, (b) Membership functions for the delta distance between the gripper and the drink, (c) Membership functions of the force that the motor must exert

As can be seen, Gaussian-type membership functions were implemented which allow a smooth change in the system without presenting over breaks. The Mamdani method is used for the fuzzy system controller [30]. This method has been used in works such as the one presented in [31], where a fuzzy system focused on the detection of photovoltaic failures is implemented, or in [32], for the detection of software failures. Table 2 shows the fuzzy associative matrix (FAM) used for the system.

Table 2. FAM of the fuzzy system designed

W	D	T
LW	MD	LF
LW	LD	MF
LW	HD	LF
MM	MD	MF
MM	LD	MHF
MM	HD	MLF
HM	MD	HF
HM	LD	HF
HM	HD	MHF

5. GRAPHIC USER INTERFACE AND PERFORMANCE TESTS

Figure 8 shows the section of the designed user interface, which focuses on the acquisition of training and testing databases. For this, the maximum number of fluid particles that can be generated in the environment must be set, in addition to their density and the number of images to be acquired. The images that are stored are those of the local camera and the depth map. The external camera is presented only for viewing the work environment.



Figure 8. Section of training and testing databases acquisition

Figure 9 shows the section in which the training parameters for each Stage of the faster R-CNN are set, this stage is used once the acquisition of the databases has already been made. Once the network is trained, the confusion matrix and activations of the trained model can be generated. In "Train", *dstrain* corresponds to the image database that will be used for training and in "Test", *dstest* contains the test images.

In Figure 10, the test section is shown, in which the user can establish the number of particles, their density, the network that will be used for the detection of the gripper and the drink, as well as the fuzzy system that will calculate the torque that the motor that drives the gripper must exert. With "Fill", the color of the fluid is selected and the filling of the glass begins and with the "Pickup" option, the manipulation of the glass is performed, generating the grip and lifting it.

Although the interface for the virtual environment is presented, if it is wanted to extrapolate this system to a real environment, it would be enough to design a special automatic filling system for the glass in

order to automatically acquire the databases. For the calculation of the performance in the manipulation of the glass with fluid, 100 tests were performed, in which the density of the particles and their number were varied. Figure 11 shows the different cases presented during the tests, where 76% corresponded to correct manipulations and 24% to cases in which there was a slip. It was identified that in cases where there is slippage, they occur when the densities of the fluid have no values that resemble that of a drink.

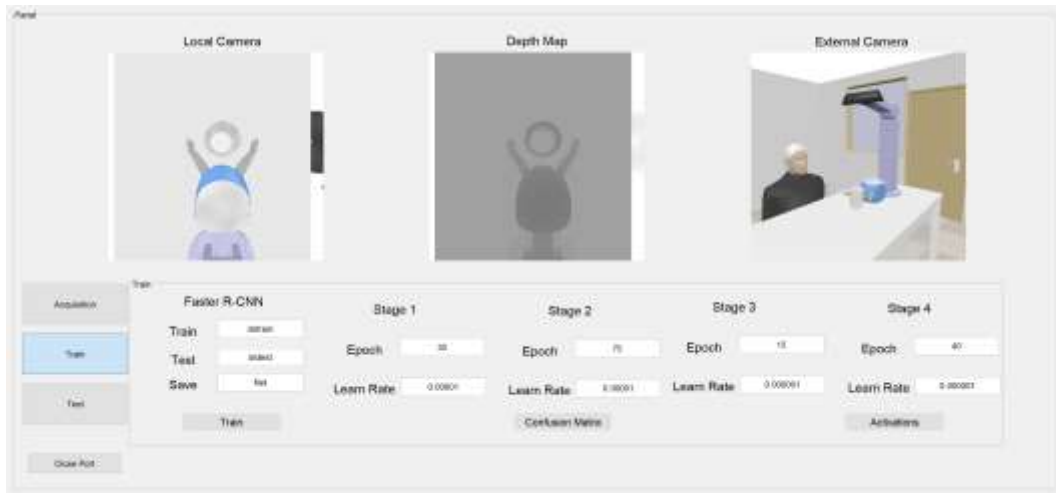


Figure 9. Section of training parameters selection of the faster R-CNN



Figure 10. Real-time testing section for glass manipulation

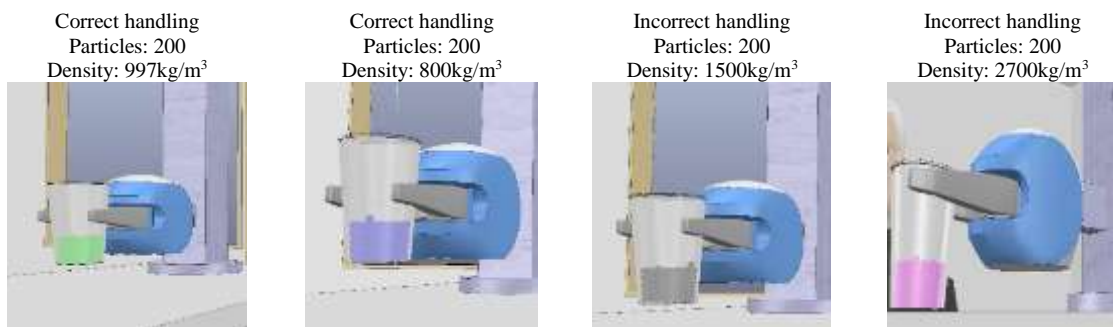


Figure 11. Examples of grip tests

6. CONCLUSION

It was possible to show that the integration of a faster R-CNN with a fuzzy system allows visual control tasks aimed at the proper manipulation of the glass with variations in weight due to changes in the level of liquid it contains, applicable to assistance robots for assisted feeding. The developments obtained are based on liquid density parameters similar to water. In general, the faster R-CNN obtained a 97.3% accuracy in the detection of the elements of interest and coupled, with the tests, 76% accuracy was obtained in proper handling of the glass, without loss of grip, making them applicable to real environments.

The results obtained allow to identify that, although the performance of the system is adequate under realistic fluid parameters, no action forces other than that of the gripper and that caused by the normal were taken into account. Therefore, it is necessary to evaluate the system for possible disturbances that may destabilize it or making trajectories that are not on a single coordinate axis.

ACKNOWLEDGEMENTS

The authors are grateful to the Universidad Militar Nueva Granada, which, through its Vice chancellor for research, finances the present project with code IMP-ING-2935 (being in force 2019-2020) and titled "prototipo robótico flexible para asistencia alimentaria", from which the present work is derived. Likewise, acknowledgement is given to the participation of the engineer Cesar Pachón who, through the service provision order number 0543/2019, was hired for the development of the exposed algorithms, with the financing of the project.

REFERENCES

- [1] Carsten H., Susanne B., "31 million robots helping in households worldwide by 2019," *IFR Press*, 2019. [Online]. Available: <https://ifr.org/ifr-press-releases/news/31-million-robots-helping-in-households-worldwide-by-2019>.
- [2] Karabegović, I. and Doleček, V., "The role of service robots and robotic systems in the treatment of patients in medical institutions," *Advanced Technologies, Systems, and Applications*, pp. 9-25, 2017.
- [3] World Health Organization, "Guillain-Barré syndrome," 2016. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/guillain-barr%C3%A9-syndrome>.
- [4] Das, N., Bezboruah, G. and Das, I., "Study on the Clinical Profile of Patients with Cerebral Palsy," *International Organization of Scientific Research Journal of Dental and Medical Sciences*, vol. 15, no. 7, pp. 54-58, 2016.
- [5] Sento, A., Srisuk, P. and Kitjaidure, Y., "An intelligent system architecture for meal assistant robotic arm," *9th International Conference on Knowledge and Smart Technology (KST)*, pp. 166-171, 2017.
- [6] Perera, C. J., Lalitharatne, T. D. and Kiguchi, K., "EEG-controlled meal assistance robot with camera-based automatic mouth position tracking and mouth open detection," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1760-1765, 2017.
- [7] Candeias, A., Rhodes, T., Marques, M. and Veloso, M., "Vision Augmented Robot Feeding," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [8] LeCun, Y., Bengio, Y. and Hinton, G., "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [9] Kim, Y., "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.
- [10] LeCun, Y., et al., "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541-551, 1989.
- [11] 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.
- [12] 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.
- [13] Pachón, C. G., Pinzón, J. O. and Moreno, R. J., "Product Detection System for Home Refrigerators implemented through a Region-based Convolutional Neural Network," *International Journal of Applied Engineering Research*, vol. 13, no. 12, pp. 10381-10388, 2018.
- [14] Parkhi, O.M., Vedaldi, A. and Zisserman, A., "Deep face recognition," Visual Geometry Group Department of Engineering Science University of Oxford, 2015.
- [15] Peng, X. and Schmid, C., "Multi-region two-stream R-CNN for action detection," *European conference on computer vision*, pp. 744-759, 2016.
- [16] Soomro, K., Zamir, A. R. and Shah, M., "UCF101: A dataset of 101 human actions classes from videos in the wild," *arXiv preprint arXiv:1212.0402*, 2012.
- [17] Schröder, S., Killmann, I., Frank, B., Voelker, M., Fiederer, L., Ball, T. and Burgard, W., "An autonomous robotic assistant for drinking," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6482-6487, 2015.
- [18] 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.
- [19] Lee, C. C., "Fuzzy logic in control systems: fuzzy logic controller," *IEEE Transactions on systems, man, and cybernetics*, vol. 20, no. 2, pp. 419-435, 1990.
- [20] Furrer, F., Burri, M., Achtelik, M. and Siegwart, R., "RotorS-A modular Gazebo MAV simulator framework," *Robot Operating System (ROS)*, pp. 595-625, 2016.

- [21] Rohmer, E., Singh, S. P. and Freese, M., "V-REP: A versatile and scalable robot simulation framework," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1321-1326, 2013.
- [22] Nogueira, L., "Comparative analysis between gazebo and v-rep robotic simulators," *Seminario Interno de Cognicao Artificial-SICA*, 2014.
- [23] Hung, J. and Carpenter, A., "Applying faster R-CNN for object detection on malaria images," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 56-61, 2017.
- [24] Ren, Y., Zhu, C. and Xiao, S., "Small object detection in optical remote sensing images via modified faster R-CNN," *Applied Sciences*, vol. 8, no. 5, 2018.
- [25] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [26] A. Selimović, B. Meden, P. Peer and A. Hladnik, "Analysis of Content-Aware Image Compression with VGG16," *IEEE International Work Conference on Bioinspired Intelligence (IWOBi)*, pp. 1-7, 2018.
- [27] Enciso-Aragón, C. J., Pachón-Suescún, C. G. and Jimenez-Moreno, R., "Quality control system by means of CNN and fuzzy systems," *International Journal of Applied Engineering Research*, vol. 13, no. 16, pp. 12846-12853, 2018.
- [28] Zhang, Y., Deng, H., Zhuo, W., and Zhong, G., "Stiffness-gradient-based Grasping Force Control for an Underactuated Prosthetic Hand," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 11, no. 4, pp. 460-464, 2018.
- [29] F. Abadianzadeh, V. Derhami and M. Rezaeian, "Visual servoing control of robot manipulator in 3D space using fuzzy hybrid controller," *4th International Conference on Robotics and Mechatronics (ICROM)*, pp. 61-65, 2016.
- [30] Weihong, Z., Shunqing, X. and Ting, M., "A fuzzy classifier based on Mamdani fuzzy logic system and genetic algorithm," *IEEE Youth Conference on Information, Computing and Telecommunications*, pp. 198-201, 2010.
- [31] Dhimish, M., Holmes, V., Mehrdadi, B. and Dales, M., "Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection," *Renewable energy*, vol. 117, pp. 257-274, 2018.
- [32] Erturk, E. and Sezer, E. A., "Software fault prediction using Mamdani type fuzzy inference system," *International Journal of Data Analysis Techniques and Strategies*, vol. 8, no. 1, pp. 14-28, 2016.

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. M.Sc. in Industrial Automation from the Universidad Nacional de Colombia - 2012 and PhD in Engineering at the Francisco José de Caldas District University - 2018. 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.
E-mail: robinson.jimenez@unimilitar.edu.co



Astrid Rubiano, Bogotá Colombia. PhD degree in Mechatronics, with specialty in control of soft robotics, Nanterre University, Paris, France, 2016. M.Sc. degree with specialty in automatic control systems, Technologic 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.



Jose Luis Ramirez, Bogotá Colombia. PhD degree in Mechanics, with specialty in artificial muscles based on smart materials, Nanterre University, Paris, France, 2016. M.Sc. degree with specialty in automatic control systems, Technologic 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.