25th ABCM International Congress of Mechanical Engineering
October 20-25, 2019, Uberlândia, MG, Brazil

COBEM-2019-2405
TEACHING A HUMANOID ROBOT THE ALPHABET IN LIBRAS

Marlete Maria da Silva
José Maurício S. T. Motta
Department of Mechanical Engineering/Faculty of Tecnology
University of Brasilia, Campus Darcy Ribeiro, Brasília-DF, Brazil | CEP 70910-900
marlete_eng@outlook.com; jmmotta@unb.br

Abstract. *The humanoid robot NAO can be used as an excellent social stimulus tool for recognizing objects, people faces, imitations, and other forms of non-verbal communication. In this context, the LIBRAS (Brazilian Sign Language) is used here in a work experiment using the NAO robot for recognition. The main objective of this article is to contribute to the study of the NAO Vision System with the resources provided by the robot in the context of inclusion of impaired people, as a proposal for human-machine interaction. This first work deals with the recognition of the alphabet in LIBRAS. The recognition of letters was developed in the Choreographe software, a powerful tool with several interaction applications by using animations, such as movements, robot speech and object and face recognition. As partial results, the robot NAO recognized some static letters on cards with a white background and was tested with hand-made letters. As results were used 2 methods in the Choreographe with total recognition of the alphabet.*

Keywords: *Humanoid Robot NAO, Image Recognition, Non-Verbal Communication, LIBRAS.*

1. INTRODUCTION

Technological advances have modified people's behavior using modern resources, such as: Social Media; Assistive Technology equipment to assist in daily activities; prostheses to help with locomotion; among others.

Assistive Technologies (TA's) have become an extremely important resource to rescue and provide life with more accessibility and approximation in the social and professional relations of these users. According to the Technical Assistance Committee: *"Assistive Technology is an area of knowledge, an interdisciplinary feature that encompasses products, resources, methodologies, strategies, practices and services that aim to promote functionality related to the activity and participation of people with disabilities, disabilities, reduced mobility, aiming at their autonomy, independence, quality of life and social inclusion"* (Brasil, 2009). In this way, TAs ensures the development of researches, strengthening the use of technologies, among other subjects related to its developed.

Among the technological resources robots have an important role, since they are used in several areas, for example: in the industries, due to their characteristics such as resistance, robustness and precision; in the medical area with robots that perform surgeries; drones (unmanned aerial vehicles) with various functions; those of domestic use; military use; the humanoid robots; mechanical prostheses; in the educational area, Assistive Robotics, among other classifications (Adade, 1992).

The contribution of robotics to the world is very wide depending on the robot design and applications. Another area of robotics, not as much explored as the others, is the Social Assistive Robotics as a therapy resource in the aid of socialization with elderly and autistic children with difficulties of interactions.

In the context of communication, the LIBRAS (Brazilian Sign Language) is approached here as a work experiment using the *NAO* robot for recognition. According to WHO (World Health Organization), more than 5% of the world's population - 466 million people - have incapacitating hearing loss (432 million adults and 34 million children). It is estimated that by 2050, more than 900 million people - or one in ten - will have disabling hearing loss. Worldwide, unmet hearing loss represents an annual cost of US $ 750 billion (OPAS, 2019).

In Brazil in 2002, LIBRAS was established as the 2nd official language, that is, Brazil is a bilingual country. The sign language differential of other languages is its visual-spatial modality. In all Sign Languages, including Libras, each word is represented by a sign. The Sign Language are sign-visual calls because the responsible for sending the communication are the hands through the signs, and the receiver are the eyes. These languages differ from oral-hearing (used by listeners) in which the sender is the voice and the receiver the ears.

Signals are formed from the combination of the shape and movement of the hands and the point on the body (or space) where these signals are performed. For example, the same formation of the hands, but in a different place in space or the body takes on another meaning, that is, it means another word (Pinheiro, 2010).

The alphabet in LIBRAS or Datylology is the identification of the combined letters that are used to enter names of people, localities and other words that aid in communication. Alphabet typing is used in cases of words that do not have a specific sign or do not remember the sign, so the word is spelled letter by letter. Learning the alphabet in LIBRAS is the first step to being fluent in deaf communication.

Brazilian deaf people have been interested in the study and learning of these technologies. Therefore, the importance to develop research that promotes the inclusion in both academic and professional environments of social robotics is plenty justified for these people.

2. THE USE OF HUMANOIDS ROBOTS

According to Wade (Wade et al., 2012), it is important that robots, in the context of its interaction with humans, behave predictably and even under failures it should be possible to anticipate its behavior. In this way, this type of robotic system should be robust and appropriate for this interface with humans. In the case of users with cognitive deficits, the interaction can be accomplished through the stimulation of speech and the direct relationship with other people seeking to help them in their interaction with the world, motivate them to socialize with other people, placing themselves as a facilitating interface for social interaction, and encouraging them to social and cognitive development. (Wade et al., 2012).

In the work of Shamsuddin (Shamsuddin et al., 2012) it is discussed a revision of the control and the exploration of the *NAO* robot workspace. The interaction of this robot with children is usually very helpful. The *NAO* robot has been introduced in a primary school environment, allowing children to learn programming in a fun and practical way. In this context it has been also used in the treatment of autistic children, encouraging the cognitive development and interactive capacities of these children (Shamsuddin et al.,2012).

Humans can easily express and recognize emotional information, which is transmitted primarily with the use of non-verbal communication such as facial expressions, hand gestures, and body movements.

Talking and time of talking is also an important means of communication between humans. In addition, the talking time, such as starting or stopping the conversation or changing the subject is something that humans can do very well but that the machines have difficulty to perform it easily (Feil-Seifer et al., 2011).

In the last decade, many works have been developed to make Human Machine Interaction (HMI) more natural and the robot more socially and contextually aware. During a 3-month HMI exposition at Trinity College Dublin's Science Gallery, researchers showed strong evidence of the usefulness of using nonverbal cognitive information in humans and interaction robots. The research also indicated that vision-based channels are becoming an important component in these interactions (Han et al., 2012).

2.1 *NAO* Robot

Currently, the *NAO* robot is probably the most widely used humanoid robot in research and education around the world. The *NAO* robot is able to sing, dance, walk and talk. It was manufactured by the French company Aldebaran Robotics. The *NAO* robot is considered as one of the most advanced robots of the present time. Its use is mostly to teaching and research in Robotics and Artificial Intelligence in institutions around the world, especially to interact with humans and objects (Softbankrobotics, 2019).

In Figure 1, the *NAO* robot is equipped with various sensors: tactile sensors, ultrasonic sensors, a gyroscope, an accelerometer, force sensors, infrared sensors, 2 HD cameras, 4 microphones and high precision digital encoders at each joint. The robot has 25 degrees of freedom, soft and precise motors, controlled by software with capacity of complex movements. The new version 6 of the *NAO* comes with a new Intel Atom processor E3845 (1.91Ghz) (Robotlab, 2019).

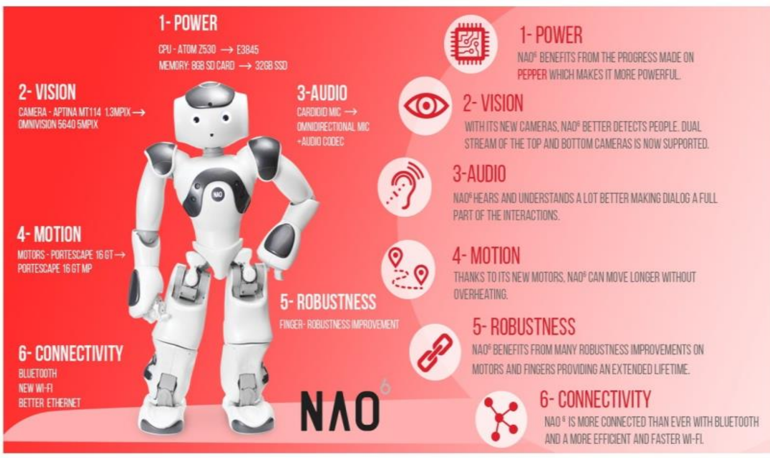


Figure 1. *NAO* Robot (Robotlab, 2019).

These robots are versatile and have high-performance, with advanced platforms for in-depth study, such as human-machine interaction, cognitive computing and stand-alone navigation among other characteristics (Softbankrobotics, 2019).

Concerning the programming for the implementation of activities, the head of the robot carries a processor with a software, allowing autonomous behavior, named Open*NAO*, a GNU/Linux distribution based on Gentoo, developed specifically for the robot, allowing the multi-platform development of programs for *NAO*. The C ++ and Python languages are supported, C ++ is the most complete and the one that writes in real time.

The additional software provided by the Aldebaran Organization for the *NAO* robot are the Choregraphe and Monitor tools. These tools provide access to the image streams produced by the *NAO* cameras and thus allow the user to see from the robot's point of view.

In the Monitor settings related to the camera parameters are: resolution, frame rate and contrast. Another useful feature is the ability to switch between the top and bottom cameras. The implementation part, which actually retrieves the images for classification, uses parameters independent of those defined in the Monitor (Miskam, 2014).

The *Choregraphe* can help users in the basic construction of cognitive interaction modules. Interaction activities include two-way communication, facial recognition, object tracking, dance, storytelling, among others. *Choregraphe* is one of the most important tools for creating behaviors for the *NAO* using a graphical user interface (Csala, 2013).

Another development resouces is *NAOqi*, which consists of the highest-level controller that describes the *NAO* software architecture. This controller can be run directly from the robot remotely with a computer, or proportionally among these (Duchi et al, 2011).

The heart of the system, called *NAO*, runs at runtime, and is responsible for all important tasks such as USB communication of the second CPU with the ARM processor in the body. *NAOqi* is generically and modularly designed to serve as a universal cross-platform foundation for other applications. The idea behind this configuration is to provide the same APIs and interfaces that robots use for application developers that do not run directly on the *NAO* robots (Miskam, 2014).

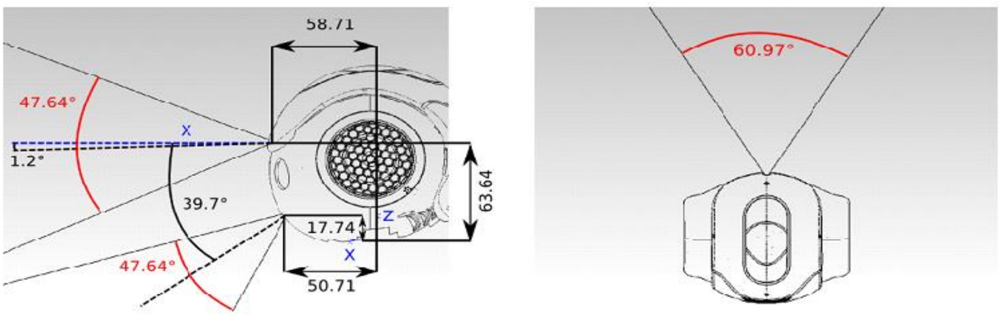
The *NAOqi* API is separated into several parts, each allowing access to distinct functionalities and robot systems: NAOqi contains the central methods for interacting with robot configurations, resources, network connections, etc. Particularly important for the current work, the data transferred to and from each robot module was performed through ALMemory. *NAOqi* Vision is another important part of the API because it is used to take pictures with ALVideoDevice (AllMemory, 2019).

2.2 *NAO* Robot Vision System

By using the cameras, the robot can recognize faces, objects, read books and perform imitations. The programming tools of the *Choreographe* software are powerful concerning developing creativity and sophisticated interactions. The *NAO* programming platform goes from the most basic level to the most complex, allowing it to be used as a teaching and learning tool for children, young people and adults.

The *NAO* robot has two identical High Definition (HD) video cameras on its face, which are used to look forward and down. The resolution of the cameras is up to 1280 * 960 pixels at 30 frames per second (fps). The camera on the forehead is used to look forward, and the other one, which is located in its mouth, is used to look down. The fixed cameras have limited viewing angles. The two cameras are used together to help the *NAO* robot to identify objects immediately (Robot Video, 2019).

The Figure 2 shows the location and the visual angles of the two cameras. As reported above, the fixed cameras cannot roll like human eyes, so the visual angles are limited. Figure 2a, on the left, shows the visual angle in the vertical direction and Fig. 2b shows the visual angle in the horizontal direction. The head joint is relevant to movement controls and offers 2 degrees of freedom (Aldebaran, 2019).



(a) (b)

Figure 2. Viewing angles of the NAO robot: (a) Vertical direction, (b) horizontal direction (Aldebaran, 2019).

Vision can be built through the process of capturing an image frame to "remember" what you see. It is then stored in memory for later recognition.

There are also two other ways to access the *NAO* ALVisionRecognition Python view module: (1) signing the AlMemory key that contains the ALVisionRecognition detection result, or (2) calling ALVisionRecognition APIs via the *Choregraphe* Script box (Aldebaran, 2019).

The ALVisionRecognition module is part of the *NAOqi* framework, which is the main Python module. It provides functionality to learn how to identify specific objects, which must be labeled by hand.

While ALVisionRecognition allows the user to teach instances of specific robot objects using visual key points, ALVisionRecognition, as provided by *NAOqi*, does not have the concept of object classes (AllVisionRecognition, 2019).

The main idea is that these features are extracted from an image. Then, they are used to detect properties in the image. The method produces a database of labeled images. When a new image is captured, its resources are compared with the labeled images in the database and the one with closest resources is chosen.

The AlPhotoCapture module allows one to take photos and use the cameras of the robot and save the images to a disk. If the parameters are not set manually, the following will be used: Camera ID: default (upper camera); Color space: BGR; Time between two images: 200 milliseconds (that is, 5 frames per second); Resolution: VGA (640 x 480) and image format: ".jpg" (AlPhotoCapture, 2019).

3. METHODOLOGY

The experiment was carried out in the Robot *NAO* version 4 (H25) whose configurations are: Intel Atom processors of 1.6GHz in the head and the Advanced RISC Machine (ARM) model in the body (for control and communication with motors and sensors). Random Access Memory (RAM) of 1GB; 802.11 A/B/G network, 10/100Mbit Ethernet; 4GB secure digital card (SD) hard drive for software; one set of lithium polymer batteries with 27.6 Watt-hours, for approximately 1 hour of use; 4 microphones on the head for sound detection; 2 loudspeakers on the left and right sides of the head; 2 cameras with 1.3 megapixel (one pointed forward, one pointed to the ground); 2 ultrasonic sensors for obstacle detection (15-70 cm range); 3 touch sensors in each hand; 3 touch sensors in the head; 8 pressure sensors per foot; acceleration sensors and gyroscopes in the body; infrared transmitter and receiver in the "eyes"; 25 degrees of freedom; Light Emitting Diode (LED's) colored eyes and ears, head and feet; size of 57 cm standing (Tozadore, 2016).

Learning an object can be performed through the "*NAO* Teaching" features for recognition. The first camera, located on the robot's forehead, is mainly used to look at the object and get information from it.

Two types of experiments were performed using the *Choregraphe* resources: the first with boxes with the "face reco" box and the second using the "visio reco" box. The procedures that were carried out for the 2 types of experiments will be presented.

Both experiments were done for the recognition of the alphabet in Brazilian Sign Language (LIBRAS) by the *NAO* robot. The first experiment for the recognition of the static image, the *Choregraphe* software was used.

To perform this activity, it was necessary that the *Choreographe* is connected to the real robot. Subsequently, the monitor video box was activated. In Figure 3, the steps performed for letter recognition are shown.

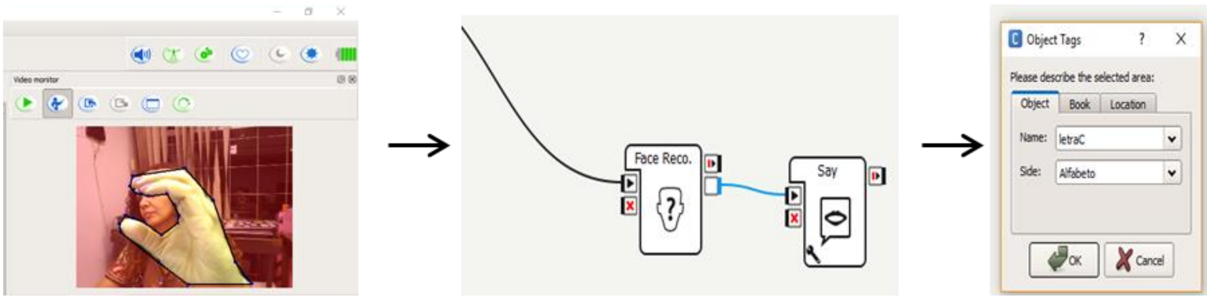


Figure 3. Steps for recognizing the alphabet in LIBRAS.

After activating the video monitor the signal corresponding to the letter "C" is made with a human hand. For this action 5 seconds is needed to conclude. Next, click on "learning", where it is bypassed with dots. The animation named "Vision Reco" is inserted in the main area of the screen, in the second step of Fig. 3. Then, a box opens to be named with the name of the letter and later sent to the database and saved in the robot´s computer memory.

First, the robot learns an object, in this case the static letters, then these images are sent to the database and loaded into the robot's memory. The program made here for the recognition of the letter was to have the speech module of the robot identifying the letter´s name after the recognition with the camera.

In the process of recognizing the image of any letter in LIBRAS, it is required first to feed the database with the corresponding images. In the robot learning process, image capturing can be performed with a white background (use a card with the letter image) or only using a color for the border of the static letter. And for the recognition of any letter in LIBRAS, one must first feed the database with the corresponding images, show the letter to the *NAO* and wait for him to speak the name.

In Figure 4, the second experiment is presented using the "vison recon" box, together with the "Switch Case" module. The letter format contour was performed manually according to Fig. 3. The "Switch Case" module searches the letter in the database as previously stored. The advantage of this experiment is that it centralizes all data in a simplified way, speeding the search and making identification using the conditional structure. It was also used the "say" box as output, where the robot identifies and speaks the letter shown, turning to be possible the recognition of all letters of the alphabet in LIBRAS in a single file extension ".pml".

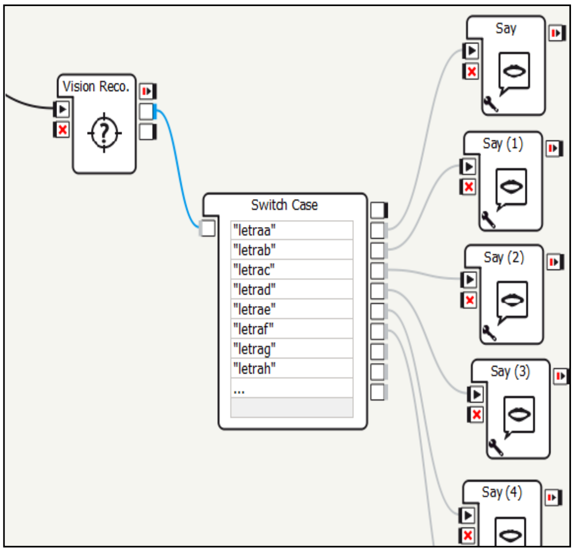


Figure 4. "Switch Case" module for letter identification in LIBRAS.

The recognition process is partially robust at a distance from half to twice the distance used for learning, with inclination angles of the camera up to 50°, where light and rotation conditions can influence results. In addition, learned objects may be partially hidden for the recognition step. For a better performance of the robot, the submitted database should contain only essential information for the detection task, since all additional data stored may be not necessary (AllVisionRecognition, 2019).

4. RESULTS

As results, the *NAO* robot recognized some static letters with the white background (card with the letter drawn). It is still being tested with the letter made with the human hand without the card, according to Fig. 5.

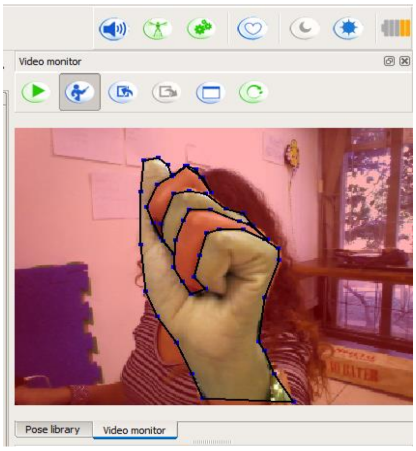


Figure 5. Learning the letter “A” in LIBRAS.

In the first experiment, the time for letter recognition is low, taking up to 40 sec. in average, but the modules still must be investigated for the letter position, brightness and luminosity in the tested environment.

The second experiment was more robust and faster for recognition of alternate letters, with a period of 5 to 10 sec. from one letter to another.

In Figure 6 it is shown the Python script of the "Switch Case" module, as well as a box for sensor search or kind commands.

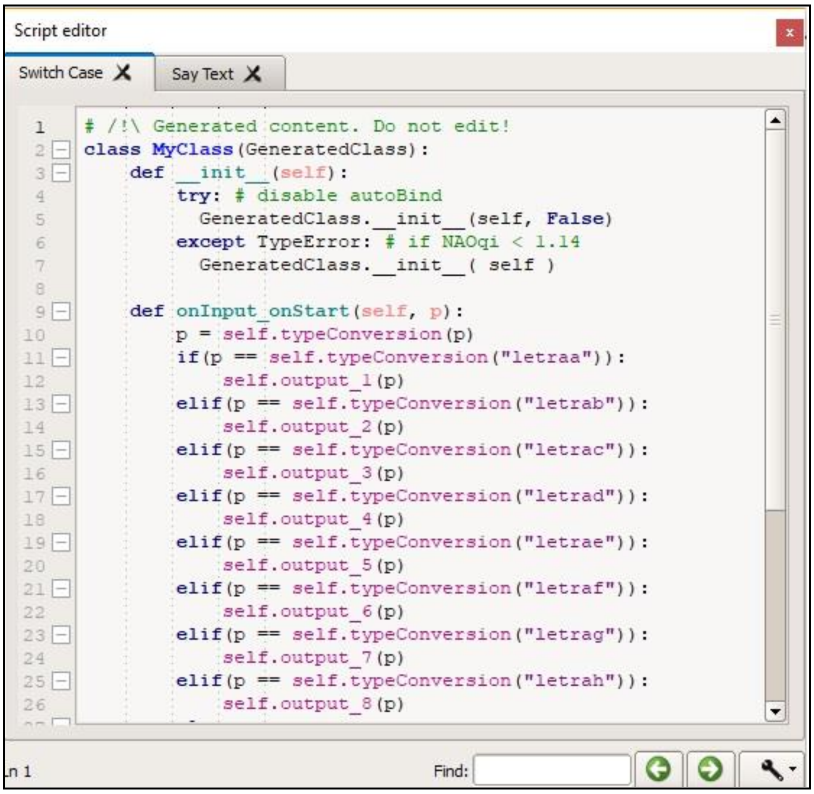


Figure 6. Python script of the conditional module "Switch Case".

5. DISCUSSION

In the Computer Vision System, the algorithm first finds features in the image. The resources are distinguishing parts of the image that aid the computational vision algorithm to decide which object is present. It typifies features and segment borders and texture. Edges occur because objects tend to contrast from the background. Some objects tend to have uniform textures. Another feature that usually comes to mind is color. However, it is difficult for an algorithm to use colors to be searched. This is because colors look very different under various lighting conditions. In most situations, our eyes adjust the color of the light source, so one can recognize that an object is red, but the computers are unable to do so robustly (AlPhotoCapture, 2019).

In the Figure 7, the left image is an example of an image that cannot be learned due to missing parameters, in such a way that Choreographe will send an error message canceling the saving process in the database. Differently, the right image presents borders in a more effective way for letter recognition.

That error message is a feedback confirming whether the learning is successful or not (due to the poor quality of the input). If it is, the image changes to QVGA resolution and it is captured.

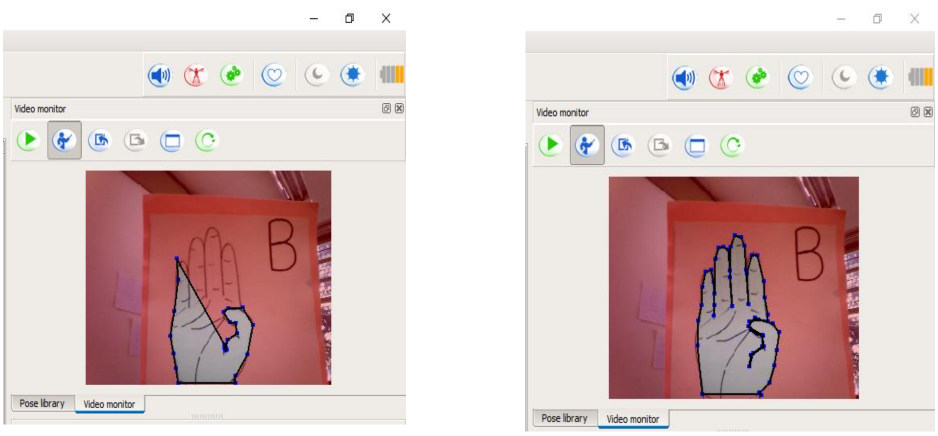


Figura 7. Analysis of image quality.

This vision module is based on the recognition of feature points other than on the internal shape of objects, so it cannot recognize non-textured objects, that is, it recognizes instances of objects. Currently, all the feature points detected in the current image are combined with only one feature point learned in the database. If the scores assigned to two objects are very close, the feature point will be associated with none of them. Learning twice the same area of an object will reduce its detection rate (AllVisionRecognition, 2019).

When the object is too close to the robot, it is no longer in the camera's field of view. Thus, the vision algorithm cannot detect any object and so the robot stops talking.

6. CONCLUSION

The object of this work was to explore the resources of the *Choreographe* aimed at human-machine interaction with a humanoid robot in the inclusive context.

Compared 2 types of implementation in modules in the *Choreographe* to get the images and save in robot memory. The second method proved to be more efficient in relation to the first in terms of time and also for creating a unique file for all letters. The exit of the recognition with the voice of the robbery was very precise, whereas the first method of implementation delayed 30 sec to recognize a letter, and is also implemented one at a time in separate files

Most robots are usually controlled by human operators and demand autonomy, mainly anthropomorphic robots. Hence, they are not confused with expensive toys that walk and talk without any functionals applications. In this current *NAO* application, researchers need specialized knowledge to explore the *Choreographe* resources to contribute in the inclusive context. It is still necessary more effort towards finding the best techniques for the development of modules for the interactive purpose, thus improving the strategies of interaction and communication of the robot, as well as to evaluate the robot performance.

In future works, word recognition using the alphabet in LIBRAS (Datylology), the name entered is represented letter by letter, then joined to identify the word, this will also be possible with machine learning algorithms. The idea is to explore robot vision capabilities to identify letters and words and also recognize gestures, phrases made in LIBRAS.

In general, robots are expensive machines for accessibility and usage in social rehabilitation therapies, but reports have shown excellent results, especially those with Artificial Intelligence in applications of vision for facial recognition and gesture recognition.

7. ACKNOWLEDGMENT

The authors are grateful to the University of Brasilia who provided the *NAO* robot to perform the work in the Computer Vision and Robotics laboratory of the Automation and Control Group at the Faculty of Technology.

8. REFERENCES

Adade F. A., 1992. “Fundamentos de Robótica - Cinemática, Dinâmica e Controle de Manipuladores Robóticos”. São José dos Campos: ITA.

Aldebaran, 2019. “NAO H25”. Jun. 2019. < http://doc.aldebaran.com/2-5/family/nao_h25/index_h25.html#nao-h25 >.

AllVisionRecognition, 2019. “ALVisionRecognition”. Jun. 2019. <http://doc.aldebaran.com/2-5/naoqi/vision/alvisionrecognition.html#alvisionrecognition>.

AlMemory, 2019. “AlMemory”. Jun. 2019. <http://doc.aldebaran.com/2-5/naoqi/core/almemory.html#almemory>.

AlPhotoCapture, 2019. "ALPhotoCapture". Jun. 2019. <http://doc.aldebaran.com/2-1/naoqi/vision/alphotocapture.html #alphotocapture>.

Brasil, 2009. "Subsecretaria Nacional de Promoção dos Direitos da Pessoa com Deficiência - Comitê de Ajudas Técnicas". Tecnologia Assistiva. Brasília: CORDE, 138 pp.

Csala E.; Németh G. and Zaink Cs., 2013. "Application of the NAO humanoid robot in the treatment of bone marrow transplanted children". Department of Telecommunications and Media Informatics Budapest University of Technology and Economics, Budapest, Hungary. Interspeech: Show & Tell Contribution.

Duchi J.; Hazan E. and Singer Y., 2011. "Adaptive subgradient methods for online learning and stochastic optimization. The Journal of Machine Learning Research", 2121-2159.

Feil-Seifer D. and Mataric M., 2011. "Automated detection and classification of positive versus negative robot interactions with children with autism using distance based features". Proc.6thACM/IEEEInt.Conf.Hum. Robot Interact. (HRI 11), March 6–9, Lausanne, Switz, pp. 323–30. New York: ACM.

Han JingGuang; Campbell N.; Jokinen K. and Wilcock G., 2012. "Investigating the Use of Non-verbal Cues in Human-Robot Interaction with a Nao Robot". Trinity College Dublin/Speech Communications Laboratory, Dublin. University of Helsinki. Finland hanf, Kristiina Jokinen.

Miskam, M. A.; Shamsuddin S., Yussof H.; Omar, A. R., Muda, M. Z., 2014. "Programming Platform for NAO Robot in Cognitive Interaction Applications". lCOE for Humanoid Robots and Bio-sensing (HuRoBs). Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka Hang Tuah Jaya, 76100 Durian Tunggal, Malaysia MARA. IEEE International Symposium on Robotics and Manufacturing Automation, Malaysia.

OPAS, 2019. "Organização Pan-Americana da Saúde/Organização Mundical da Saúde (OPAS/OMS)". Jun. 2019. <https://www.paho.org/bra/index.php?option=com_content&view=article&id=5881:oms-lanca-o-aplicativo-para-dispositivos-moveis-que-ajuda-a-detectar-perda-auditiva&Itemid=839>.

Pinheiro, M. L., 2011. Língua de Sinais Brasileira. São Paulo: Know How.

Robotlab, 2019. "NAO Power V.6 Educator Pack". Jun. 2019. <https://www.robotlab.com/store/nao-power-v6-educator-pack>.

Robot Video, 2019. "Robot Video". Jun. 2019. < http://doc.aldebaran.com/2-5/family/robots/video_robot.html#robot-video>

Shamsuddin S.; Yussof H., Ismail L., Fazah A. Hanapiah, Mohamed S.; Ali Piah, H. and Zahari N. I., 2012. "Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO". *IEEE 8th International Colloquium on Signal Processing and its Applications*, Melaka, pp. 188-193. doi: 10.1109/CSPA.2012.6194716.

Softbankrobotics, 2019. "SoftBank Robotics - Education and Research". Abr. 2019 <https://www.softbankrobotics.com /emea/en/nao>.

Tozadore, Daniel C., 2016. "Aplicação de um robô humanoide autônomo por meio de reconhecimento de imagem e voz em sessões pedagógicas interativas". *Dissertação de Mestrado*. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

Wade Eric, Avinash Parnandi; Ross Mead and Maja Mataric, 2012. "Socially assistive robotics for guiding motor task practice". Paladyn 2: 218. https://doi.org/10.2478/s13230-012-0017-0. Online ISSN 2081-4836.

9. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.