# The Improvement of the Domestic Robot DoRIS

Lucas S. Avila[1] João F. de S. S. Lemos[1] Jardel dos S. Dyonisio[1] Laura de A. Dalmolin[1] Andrew M. Alaniz[1] Beatriz C. da Costa[1] Igor P. Maurell[1] Renan T. Fonseca[1] Ricardo B. Grando[1] Paulo L. J. Drews-Jr.[1] Cedenir B. da Costa[1] Rodrigo da S. Guerra[2] David R. Richards[2] Cleber D. Werlang[2] Douglas S. Brum[2] Eduardo R. Santana[2] Gabriel N. Niederauer[2] Julio P. C. R. Jardim[2] Junior C. Jesus[2] Luciano A. Cardona Jr.[2] Victor A. Kich[2]

***Abstract*— After one year of work and research, the *Brazilian United Team for Intelligent Automation - BUTIA* has updated and changed many aspects of the robot *DoRIS*. In this paper some choices made by the team are described in detail, such as how and why they were executed. Likewise, the updates are compared with the last version of the robot.**

## I. INTRODUCTION

Since everything is in constant change and evolution, the same applies to the *Domestic Robotic Intelligent System - DoRIS*. In order to develop a robot with the best technologies on the rising and be able to build a better performing and more robust domestic robot, changes and updates were made necessary.

In comparison with the last version of the robot [1], a lot of work was done, accomplishing enhancements from its outer shell design to its internal software algorithms. Thus, those improvements aimed solving some flaws of its older architecture and producing a more elegant solution. On this paper, the updates are documented and explained meticulously as how and why they were executed.

This paper is structured as follows: Section II describes the problems encountered in the last version. Section III describes the modifications and updates made to the robot, detailing the changes made in the robot design and structure (Subsection III-A), manipulation arm (Subsection III-B), vision system (Subsection III-C) and robot-human interaction (Subsection III-D). At the end, Section IV discusses about future directions and closes this paper.

## II. ISSUES IN THE PREVIOUS DORIS

When DoRIS was first created, the team used the resources and tools that were available at the time. The choices were made in order to build a robot from scratch and compete at the RoboCup@Home League as fast as possible. Developing a robot that way worked, but brought some consequences to its appearance and performance.

### A. Hardware Problems

DoRIS' absence of a manipulation robotic arm prevented it to execute tasks from the competition that involve object manipulation. The robot's outer shell made of wood was insecure, unstable and not designed to last long after the competition. Its internal organization of the electrical cables and the connectivity of the internal components were needed to be restructured to be more secure and be able to fit into a new shelter.

[1]Universidade Federal do Rio Grande (FURG), Rio Grande, RS, Brazil.
[2]Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brazil.

### B. Software Problems

In order to have a good interaction between human and robot, it is necessary that the robot is able to hear and talk to the human. In its last version, *DoRIS* [1] was dependent of online APIs to do it, causing the robot to be incommunicable when facing connection issues. We needed a fast training of the YOLO v3 [2] to obtain a more precise model of the objects presented in the competition. The older face recognition algorithm was bad in sorting faces that looked alike. The use of matching of SIFT features [3] was a good choice to know if the person is the same at each frame captured, but it was not able of keep track of the person after it comes out of the frame.

## III. NEW DORIS

### A. Robot Appearance And Structure

*1) Torso's Frame:* The robot aesthetic and structure is composed of three columns of structural aluminium, in a triangle formation, which gives the support to all the other parts. Inside there are four transparent acrylic shelves, each one prop up the components that the robot must have. Finishing up, DoRIS was covered with 2mm black acrylic in a pentagon form. Its outside structure can been seen at Figure 1.

*2) Animatronic Face:* The DoRIS's face was designed to be a friendly robotic face with human's characteristics, improving the human-robot interaction. Its facial parts were manufactured in a 3D printer using ABS, wich is stronger, resistant and have a good cost-benefit. Figure 2 shows a 3D rendering of the designed face.

The movements of the face are reproduced by the 12 hobby servo-motors which are connected to a Arduino micro-controller powered by 5V DC. The microcontroller receives data from the Intel NUC, connected by a RS485 conversor, interpret it and send to the micro servos. Table I shows its specifications.

*3) Internal Organization:* Inside of the robot are a few essential components, among them are: a NUC and Nvidia Jetson that works in the control part of the system, two batteries (each one has 12V and 18A) and, to convert its energy, is used three stepdown with differents output voltage.

Fig. 1. DoRIS in CBR/LARC 2019.

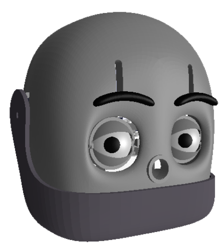

Fig. 2. Robot's Face

TABLE I
SPECIFICATIONS OF THE ANIMATRONIC FACE

| Face components | |
|---|---|
| Micro Servo SG90 | 10 |
| Micro Servo SG92R | 2 |
| LMS8UU Linear Bearing | 2 |
| Aluminum Bar - 8 mm x 70 mm | 2 |
| Arduino Uno | 1 |
| LM2596 DC 5v | 1 |
| Conversor TTL - RS485 | 1 |
| Arduino Sensor Shield v5.0 | 1 |
| Physical Specifications | |
| Height | 20 cm |
| Width | 18 cm |
| Depth | 17 cm |
| Weight | 0.9 kg |
| Material | ABS |

Fig. 3. Robot's Arm

This way the eletronic components and the exact power that it needs could be provided. It also have some USB and HDMI hubs to connect the peripheral electronics, like the microphone, the screen etc.

### B. Manipulation Arm

Our arm project started in 2019 (Figure 3), and it was an adaptation of a already built leg [4]. It was part of another robot, made of eight Dynamixel MX-106R robot actuators from Robotis, compliant joints and carbon fiber parallel links. It uses the RS-485 bus circuit with the Dynamixel protocol to be networked together by a Intel NUC, specified at [1].

The most important problem faced was to transfer the strength of rotation from the horizontal to the vertical axis, because of the construction geometry. The solution found was to fix it on the robot torso and provide a higher current discharge. The results achieved with this method are: free movement on all 3 axes, low overload concerns and almost 1kg of payload.

### C. Vision System

The complete system as explained in [1] continues to work in the same way and with the same packages, communicating by the same ROS topics. Updates were made in the method to train the object detector and at the algorithms of face recognition and people tracking.

*1) Better Object Detection:* Into some tasks of the competition, the team receives six random objects where the robot have to be prepared to recognize them. It's a few days between the delivery of the items to the teams and the start of the tasks.

The older process was to take as much pictures of the objects as possible and in so many position as possible, capturing, at least, one hundred photos from each object. After a YOLO v3 [2] was trained with that pictures and deployed into the darknet_ros package [5], as presented here [1].

In order to achieve a better detection model, we started taking pictures of the objects with a green background.


Fig. 4. Object detection using a deep learning network YOLO v3 [2].

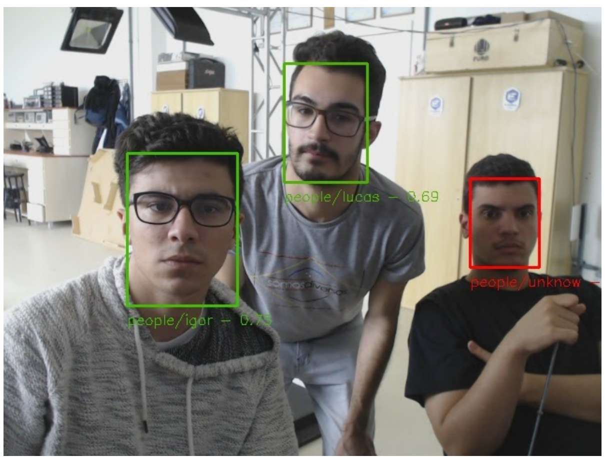


Fig. 5. The new method of the Face Recognition module.

Then, we process the images adding noisy backgrounds, flipping the images in both axes and rotating them. Thanks to this step of data augmentation, we could achieve a more precise detection model in time to participate in the objection detection tasks. Figure 4 shows a demonstration.

*2) Face Recognition Updated:* The face recognition stills use the same process as used in [1]. Its continues being based on FaceNet [6] approach and implemented with the Openface [7] framework. On its last step the classification was updated in a form that optimizes the recognition of faces which look alike and added a feature to tell if the person tested where a face know to the robot.

This step consists in the use of two classifiers available at the scikit-learn [8] library, a multi-class SVM and a one-class SVM. We first train a multi-class SVM with all the data set, then we train a one-class SVM for each person of the data set. By doing this, when we test a face into the recognizer, the multi-class classifier dictate which person that face look the most and the one-class classifier gives the verdict telling if the face is the same person indicated by the multi-class classifier. The Figure 5 shows the result.

*3) A New People Tracker:* The inputs and outputs of the people_tracking package stills the same as presented at [1]. Its receives the image and the detection of the YOLO v3 [2], after detects a person, and returns the recognition of the tracked person.

Deep SORT[1] framework is being used, based on the implementation of [9]. It takes the detection of each person and process into a neural network that returns a vector representing that person, then passes the detections into a kalman filter [10] that predicts where the localization of the person in the next frame. When in the next frame, it compares each detection with the predictions and the encoded vectors, tracking all the persons in the image. Figure 6 demonstrate its working.

### D. Robot-Human Interaction

On previous competitions, the team has struggled with internet connection problems. This scenario damaged DoRIS capability of understanding and speaking, since the libraries and APIs used were internet dependent. Therefore, a change to offline text-to-speech (TTS) and speech-to-text (STT) APIs and libraries was the logic solution for enhancing DoRIS skills related to robot-human interaction.

*1) Speech-to-Text:* This module is responsible to hear the human which the robot is interacting with. Inside the competition scope, *DoRIS* has to be able to convert into text everything that the human operator says to it, ignoring other speakers and noisy background sounds of the competition. By using the DeepSpeech framework[2], an open source implementation of speech recognition based on Baidu's Deep Speech research paper [11]. It is composed of two main subsystems: an acoustic model and a decoder. The first one makes the live capture of the voice and the second is responsible to transcribe it into words. Inside the system, it publishes all the text that it can recognize to be used by the rest of the nodes.

[1]Nicolai Wojke and Alex Bewley, Deep SORT, https://github.com/nwojke/deep_sort

[2]Mozilla, Mozilla DeepSpeech, https://github.com/mozilla/DeepSpeech

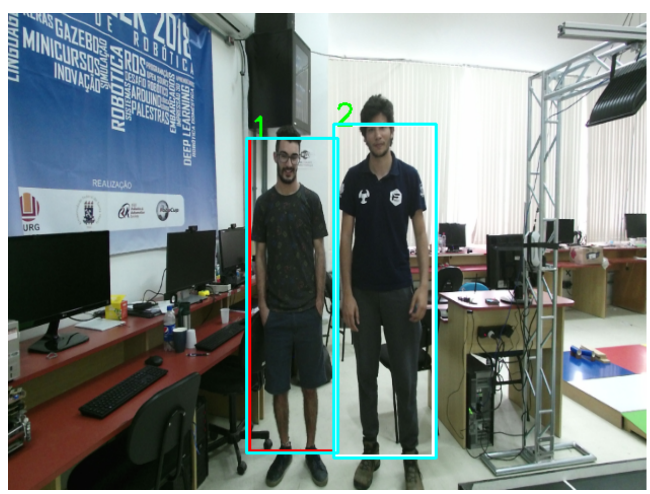


Fig. 6. Deep SORT algorithm.

*2) Text-to-Speech:* In regard to text-to-speech offline, Mozilla TTS[3] fits like a perfect solution to be the voice of the *DoRIS*, due to its fluid and clear voice. This project is an open source implementation of the Tacotron 2 [12], a neural network architecture that synthesizes speech directly from text. Tacotron 2 was developed as an entirely neural approach to speech synthesis that combines the best previous approaches, such as the sequence-to-sequence Tacotron-style model [13] and a modified WaveNet vocoder [14]. The Tacotron 2 model learns to synthesize voice that sounds as natural as human speech, through training directly on normalized character sequences and corresponding speech waveforms. Mozilla TTS was implemented in a ROS node on DoRIS, subscribed on the text-to-speech topic. Whereas, when a string is published at the mentioned topic, DoRIS' speech is synthesized.

## IV. CONCLUSIONS

All of the remaining work of the *BUTIA* robotic team can be found at the GitHub Page of the team [15] where the algorithms explained here can be found. The instructions are being added and updated to execute the algorithms in order to make it easy and available for everyone.

The team has some future directions in mind, like working in the navigation system to make it able to navigate in unknown places where there is no map of the environment. Also, the team wants to add a laser in the back of the robot to give more mobility to the robot.

The team is planning to build a new robotic manipulator based at [16] to be smaller, efficient and agile than the actual. To continue the plans to the robot-human interaction, the team is researching in how to insert a natural language module to improve the conversation skills of *DoRIS*.

## REFERENCES


[1] Renan T. Fonseca Ricardo B. Grando Miguel M. Santos Rafael R. Teixeira Paulo L. J. Drews-Jr. Cedenir B. da Costa Igor Maurell, Vinicius S. Marques. General aspects and perception of the domestic robot doris. *Proceedings of the II Brazilian HumanoidRobot Workshop (BRAHUR) and III BrazilianWorkshop on Service Robotics (BRASERO)*, 2019.

[2] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

[3] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004.

[4] Christopher Tatsch, Ahmadreza Ahmadi, Fabrício Bottega, Jun Tani, and Rodrigo da Silva Guerra. Dimitri: an open-source humanoid robot with compliant joint. *Journal of Intelligent & Robotic Systems*, 91(2):291–300, 2018.

[5] Marko Bjelonic. Darknet ros package. `http://wiki.ros.org/darknet_ros`. [Online; accessed 6-July-2018].

[6] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, June 2015.

[7] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[9] Nicolai Wojke and Alex Bewley. Deep cosine metric learning for person re-identification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 748–756. IEEE, 2018.

[10] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[11] Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, and Andrew Y. Ng. Deep speech: Scaling up end-to-end speech recognition. *arXiv*, 2014.

[12] Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, RJ Skerry-Ryan, Rif A. Saurous, Yannis Agiomyrgiannakis, and Yonghui Wu. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. *arXiv*, 2017.

[13] Daisy Stanton Yonghui Wu Ron J. Weiss Navdeep Jaitly Zongheng Yang Ying Xiao Zhifeng Chen Samy Bengio Quoc Le Yannis Agiomyrgiannakis Rob Clark Yuxuan Wang, RJ Skerry-Ryan and Rif A. Saurous. Tacotron: Towards end-to-end speech synthesis. *arXiv*, 2017.

[14] Kazuhiro Kobayashi Kazuya Takeda Tomoki Toda Akira Tamamori, Tomoki Hayashi. Speaker-dependent wavenet vocoder. *Interspeech*, 2017.

[15] Renan T. Fonseca Ricardo B. Grando Miguel M. Santos Rafael R. Teixeira Lucas S. Avila João F. de S. S. Lemos Jardel dos S. Dyonisio Laura de A. Dalmolin Andrew M. Alaniz Igor Maurell, Vinicius S. Marques and Beatriz C. da Costa. Página do time butia bots na plataforma github. `https://github.com/butia-bots`, 2018. Visited on 29/02/2020.

[16] Pierre Herman, Alejandro Mosteo, and Ana Murillo. A low-cost custom-built robotic arm on the turtlebot platform. *ROBOT 2013: First Iberian Robotics Conference*, 11 2013.


[3] Mozilla, Mozilla TTS, https://github.com/mozilla/TTS