

Identifying New Persons in the context of the Robocup@Home competition using KNN

1st David Segalle

Computer Engineering Course

UTFPR-Federal University of Technology - Paraná
Curitiba, PR, Brazil - davidsegalle@alunos.utfpr.edu.br

2nd Gustavo Fardo Armênio

Computer Engineering Course - UTFPR

Curitiba, PR, Brazil - gustavofardoarmenio@alunos.utfpr.edu.br

3rd Gustavo Fontes Lewin

Electronic Engineering Course - UTFPR

Curitiba, PR, Brazil - gustavoflewin@hotmail.com

4th Everson de Souza Silva

PPGCA-Graduate Program on Applied Informatics-UTFPR

Curitiba, PR, Brazil - everson@alunos.utfpr.edu.br

5th João Alberto Fabro

PPGCA-Graduate Program on Applied Informatics

UTFPR - Curitiba, PR, Brazil - fabro@utfpr.edu.br

Abstract—Robocup@Home proposes a challenge related to Person Recognition: after presented, a new 'operator' should become 'immediately' recognizable by the robot. The presentation procedure may require the operator to correctly interact with the robot, following a certain procedure, as instructed by the robot itself (for example, staying in front of the robot, so that the robot can take pictures of this person). In this paper, we propose the use of the KNN (K-Nearest Neighbor) supervised machine learning algorithm to include a new 'operator' in a database of persons recognizable by the robot. This algorithm uses information taken from an image segmentation of the face of the operator. The experiment evaluates how long it takes to include a new operator if the robot has from 1 to 12 current operators, evaluating also how long it takes to include this operator based on 1, 2 or more images of the new operator, taken from slightly different points of view. The results confirm that KNN can be used to 'present to the robot' up to 13 new operators, with up to 15 images for each operator, in less than 60 seconds.

Index Terms—KNN algorithm, person recognition, face recognition, online training

I. INTRODUCTION

RoboCup@Home was founded in 2006, and ever since, competitions were held worldwide, looking to eventually reach a level where robots could be used at home as a helper for the resident. Nowadays the competition influences—and sometimes directs—the course of research in the area of domestic service robotics [1].

While there are many important challenges to solve within the competition, one of particular interest is facial recognition. Within the realm of home-based tasks, the robot must obey only registered operators, and thus the robot is required to be able to identify its operators. This can be achieved with facial recognition algorithms.

Often, time limitations for registering of new persons for facial recognition are relaxed or non-existent. On the Robocup@Home competition, significant restraints for the training phase are present. A portable and usually under-powered computation device is available, as well as a very short time for the training to be done. Thoroughly understanding training phases and their limitations can help with choosing the most effective algorithm for the task at hand.

In addition to its applications within the RoboCup@Home competition, facial recognition technology holds significant promise for a wide range of real-world scenarios beyond the realm of robotics competitions. In today's interconnected world, security and access control have become paramount concerns in various industries, such as banking, healthcare, and transportation. Facial recognition systems can enhance security measures by providing seamless and non-intrusive identity verification, leading to safer transactions and smoother user experiences. Moreover, the potential applications extend to fields like law enforcement, where rapid and accurate identification of individuals can aid in criminal investigations. Beyond security, face recognition can also play a transforming role in personalized marketing and customer service by analyzing customer expressions and emotions. As this technology continues to advance, the integration of facial recognition into various domains underscores its potential to reshape how we interact with technology and each other, bridging the gap between the virtual and physical worlds.

Accordingly, this paper explores the capabilities and limitations of training when using a relatively lightweight machine learning model for facial recognition, the KNN - K-Nearest Neighbor - algorithm. Within the top priorities was to maximize results without overextending the process duration.

This paper is divided in 6 sections, organized as follows: section II presents a brief explanation of facial recognition,

while emphasizing applications related to Robocup@Home competitions. Section III presents the approaches taken by other teams, as per their respective TDPs (Team Description Papers). Section IV explains our approach, and the facial recognition software developed. This is of particular importance, since the focus is set on how well an embedded computer can deal with the KNN algorithm. Section V present some results of the experiments, and, finally, section VI summarizes the discussion and presents final conclusions.

II. FACIAL RECOGNITION, APPLICATIONS, AND CHALLENGES FOR DOMESTIC SERVICE ROBOTS IN THE CONTEXT OF ROBOCUP@HOME

Person identification is useful for innumerable tasks. For starters, controlling the people in and out of countries is a national security issue, which has been improved substantially by the use of digital fingerprints. However, situations arise where such solutions become impractical, as an example, modern computers may not have a fingerprint reader, only a camera and a microphone, and therefore, person identification must be done through them. Utilizing only a RGB camera is often the simplest solution, in terms of both user experience and associated cost.

Nowadays, face recognition from images has a multitude of uses, within them some are of particular interest. Firstly, identifying threats in full or crowded public areas, since it is impossible, impractical or unethical to close them off for security reasons. Secondly, smartphones often offer a unlock option which utilizes the camera, taking advantage of speed, being significantly faster than fingerprint reading or typing a pin. Lastly, searching for missing people can be a daunting task, however, a computer can look through hundreds of videotapes in search for the missing person quickly and cheaply, significantly reducing the amount of human effort needed.

In response to its multiple use cases, person recognition through cameras has been researched for many years. Within the plentiful methods of identification, using cameras to look at faces has been one of the more commonly used and analyzed methods. The first attempt at solve this challenge was done in 1966 by Woodrow W. Bledsoe [2], however he failed due to the complexity of the task.

Eigenface [3] is often described as the first successful face recognition software. Invented in the late 1980s, it uses linear algebra to perform feature analysis on a collection of facial images. While this might be an older algorithm and not as accurate as modern ones, it may still prove to be useful, for it was created with limited resources and was restricted to decreased hardware capabilities.

The first time facial recognition has been an option on Robocup@Home competition was in 2014, which offered additional points if the robot could identify and go to the person they had to serve drinks to. In 2015, a proper face recognition task was added in stage 2, which requires the robot to recognize and describe the operator, as well as some characteristics of people in the crowd.

In 2022, tasks which require person identification became more complex, however, they often are still able to take advantage of facial recognition. The receptionist is a task where the robot must "learn" multiple guests as well as the host, and be able to interact with them based on who they are.

For all the tasks previously described, there are similarities in the challenges, however, they are not the same. The most important limitation, presented in all of them, is 'time'. The robot should be able to, almost 'instantaneously', be able to recognize a person just introduced to it. Among the multiple algorithms, it is important to fully understand how much time they take to include a person in the database of known people, and how they perform under the given restrictions to maximize the chances of success.

III. RELATED WORK

Many algorithms have been utilized in Robocup@Home with the intention of recognizing faces, however, it is difficult to determine which ones are more robust since championship results, as well as algorithms used are not easily accessible information. This paper will attempt to analyse training times for a KNN algorithm to simplify the process of choosing how to better solve tasks.

Throughout the years multiple algorithms were used and frequently they are mentioned in Team Description Papers(TDPs). In 2017, for example, Eigenfaces [4], Deep Neural Networks (DNN) [5] and Support Vector Machines (SVM) [6] were used. In 2018, Eigenfaces [7], DNNs [8], Strands Perception People [9] [10] and Convolutional Neural Network (CNN) [11] were used by the teams. In 2019, Sparse Representations [12], Haar Cascade [13], and CNN [14] [15] were used. In 2020 and 2021 the competition was held only on simulation, due to pandemics, and in 2022 there is no collection of all Team Description Papers available, so collecting data became considerably more difficult.

For this paper not all solutions will be analyzed, the focus will be one relatively simple and lightweight algorithm, KNN, chosen due to its simplicity and speed in inserting new persons, not needing an extensive phase of training before being able to recognize new persons, just introduced in the database.

IV. TECHNOLOGIES USED

The vast majority of teams use ROS (Robot Operating System) for their operation. In response to that, the software developed for this paper is fully compatible and tested with ROS. In order to simplify training phases, as well as allowing storage in non volatile memory, the actual training gets pictures from memory and saves the output file so that a recognizer may use it. Training itself does not make use of ROS, however, gathering new pictures, as well as the recognizing portion of the developed software does.

As mentioned previously on section I, most teams have hardware limitations for processing the algorithms, and therefore, are not able to perform large amounts of processing

www.ros.org: ROS is a framework of tools used to develop software solutions for robotics.

in small amounts of time. With such limitation in mind a, computer with a Intel i5-10210U CPU was used for testing, and an expected maximum duration of around 60 seconds was allowed for the processing, from the acquisition of images until the end of execution of the included person recognition.

The programming language chosen was Python, since it offers many libraries which make designing such software easier, as well as being one of ROS officially supported languages. The ROS package is available in a github public repository.

While KNN may not be the newest or most complex solution for the challenge of recognizing people from images, it was chosen for several reasons. Firstly, it is the most practical / non-parametric approach for facial recognition [16] as well as it is fast to train as shown by this paper and has a decent recognition rate, often being at over 90% [17] when setup properly.

KNN is a supervised learning algorithm, that is, the agent observes some example input-output pairs and learns a function that maps from input to output [18]. The simplest form of KNN utilizes a brute force algorithm: by mapping the training data into a multidimensional space, new inputs will be classified by being put into the same space and compared too “k” nearest points. The decision is then made to classify the input image as what appears most in those nearest points.

Identifying which are the k nearest points in a multidimensional space can be a daunting task for computers when no information other than the data points are present. With this in mind a Ball Tree algorithm was used to recursively map out nearby points into groups which can be traversed quickly until a group of size “k” is found. The image will, of course, be assigned to the most common individual in the given group.

It is important to note that for a knn algorithm to work properly it must be given images of just faces with the background cropped out. This process is achieved by using the face_locations method within the face recognition library, which in turns utilizes dlib libraries to perform the task. Dlib utilizes a HOG system to get the locations of where faces are

In order to speed up the algorithm’s run time due to the larger number of images making up points in space, a ball tree was used. This is comparatively quicker than brute force because it assigns areas of the space occupied by certain inputs into clusters, and then further selects those into smaller and smaller subgroups, with the objective that only a fraction of the necessary points need to be analyzed.

The robot for which the software was built, and where it will be executed, was developed for the Robocup@Home Open Platform League (OPL) competitions. It has many characteristics common with other robots, such as a Kinect V1 visual sensor, and a Intel NUC computer, as shown in “Fig. 1”.

For testing the training times of the algorithm, a total number of 13 people were selected (all the members of our

https://github.com/UtBotsAtHome-UTFPR/utbots_face_recognition.

HOG stands for Histogram of Oriented Gradients, the basis on how HOG algorithm works can be found at <https://towardsdatascience.com/hog-histogram-of-oriented-gradients-67ecd887675f>

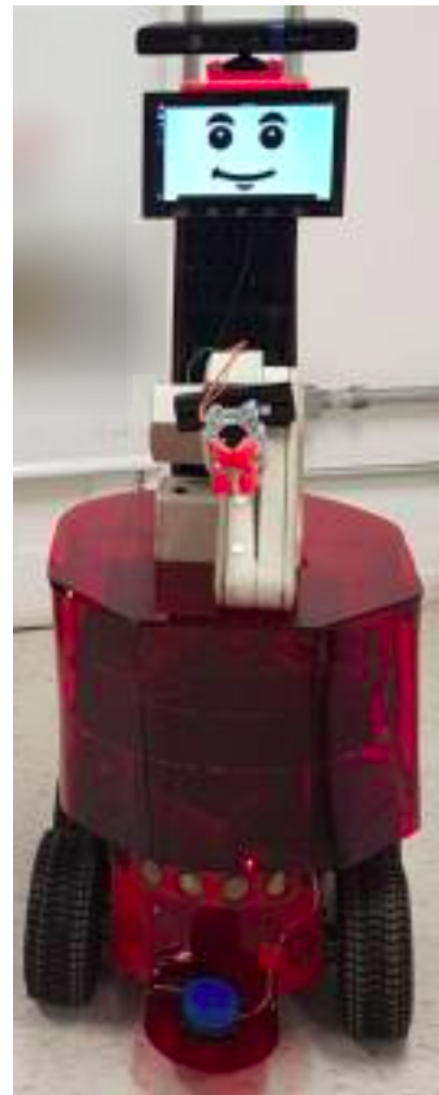


Fig. 1. Image of the robot for which the software was developed.

team). This is expected to be slightly more than what will be found in Robocup@Home competition during any given task. The max number of pictures used to perform the training phase was limited to 15, due to time restraints for picture taking phases. When taking pictures, the objective was to look for a way to get as many angles of face, rather than as many pictures as possible, this ensures that the recognition process will have as much data as possible while not overloading the system and making it too slow.

Pictures were taken by making users follow a predefined routine, however, they were not informed about how the routine works and how to maximize the algorithm’s effectiveness. Similarly, most pictures were taken in the same ambient with less than ideal light sources. While this will not yield the best possible result, they are able to more closely simulate the competition environment, where the participant will not be previously informed about the intricacies of how to take the best possible pictures or be prepared with the best possible

lighting setup.

The routine for taking pictures for training was rather simple: first, participants were asked to look at a screen that is immediately below the RGB camera of the robot. This screen is used as an "Human-Robot Interaction" device, and provides a set of written instructions, which must be followed by any 'operator' being introduced to the robot. Firstly, the routine asks them to look to the camera, and then takes a picture; soon after, it asks them to tilt their head up and takes another picture, then it prompts them to tilt their head down, left and, lastly, right, capturing an image every step of the way. This routine is repeated 3 times, obtaining a total of 15 images from the face of each person, that can then be used by the algorithm.

During the first stages of testing, it was realized that when asked to tilt their head, a subject will most likely hold it at the exact same angle until prompted otherwise, with that in mind it becomes difficult to take partially different pictures quickly. In order to maximize variability without turning the user experience into something complex, the decision was made to repeat instructions three times, taking advantage of nearly all the available time slot (up to 1 minute, by the rules of the competition in 2023). "Fig. 2" presents an example of the pictures taken from a person following this procedure.



Fig. 2. Examples of pictures taken in the training phase.

TABLE I
TIME (IN SECONDS) FOR INCLUSION OF A NEW OPERATOR, USING A NUMBER OF IMAGES FROM HIS/HER FACE. ROWS INDICATE THE NUMBER OF PEOPLE REGISTERED. COLUMNS INDICATE THE NUMBER OF IMAGES USED IN THE INCLUSION OF THE NEW OPERATOR.

images people	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0,44	1,00	3,40	4,88	6,04	7,24	8,39	9,56	10,74	12,02	13,12	13,25	13,78	16,75	19,87
2	1,84	4,01	5,24	5,30	9,09	7,54	6,87	8,27	8,82	9,79	10,93	14,72	14,03	14,02	14,59
3	1,47	3,10	4,43	5,91	7,32	11,81	10,34	11,66	13,13	15,71	20,25	18,30	18,87	24,07	28,50
4	2,96	6,05	6,53	8,49	10,61	13,12	16,65	16,72	18,75	20,85	22,92	24,99	27,05	29,01	36,17
5	2,72	6,59	11,06	13,15	16,06	19,50	23,86	20,27	27,00	25,81	28,43	35,43	28,63	30,13	32,27
6	2,61	5,19	7,78	10,35	12,95	15,55	18,13	20,71	23,29	25,85	28,44	31,03	33,63	36,21	38,82
7	3,04	6,07	9,28	14,84	15,32	22,37	28,21	31,04	36,64	41,64	45,59	51,36	47,70	42,65	45,63
8	3,49	6,98	10,46	13,96	17,44	20,92	24,40	27,85	31,37	34,85	38,30	44,74	53,86	61,89	85,47
9	5,21	10,36	15,29	20,20	26,32	29,64	36,23	40,70	44,33	49,65	45,73	47,05	50,91	54,78	58,86
10	4,40	8,72	13,06	17,40	21,69	26,03	30,34	34,67	39,00	43,34	47,70	52,02	56,37	60,70	64,95
11	4,80	9,58	14,36	19,14	23,90	28,64	33,40	38,18	42,96	47,71	52,50	57,29	62,06	66,74	71,58
12	5,24	10,44	15,66	20,87	26,08	31,27	36,44	41,59	46,81	52,01	57,20	62,37	67,50	72,75	77,88
13	5,66	11,31	16,95	22,59	28,20	33,80	39,44	45,04	50,67	56,31	61,98	67,71	73,30	79,27	84,53

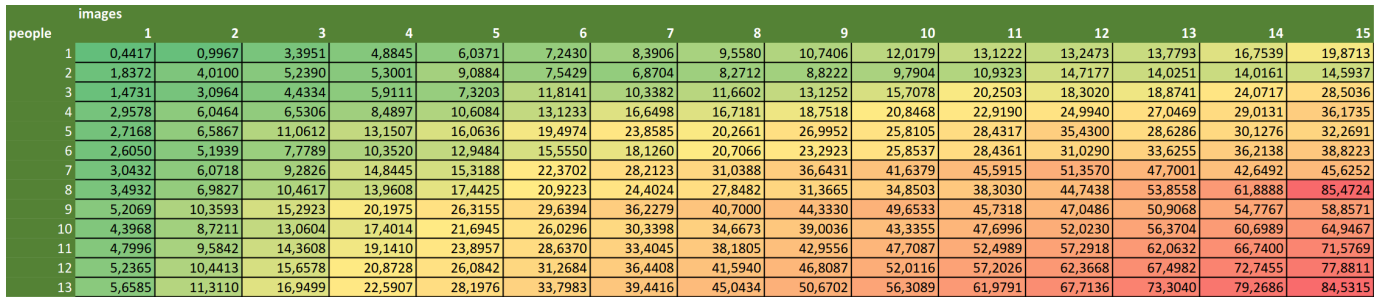


Fig. 3. Heat Map of the inclusion time in seconds, already presented on Table I. Green means less time, Red means more time to include a new operator in the recognition system. Inclusion times of less than 60 seconds are required for the competition.

V. RESULTS

For each available number of people and each available number of images, the training routine was performed a total of 9 times, and Table 1 presents the average value in seconds, to perform the "training". While for most numbers tested the duration was far smaller than the constraints of the competition, the addition of further subjects and images per subject eventually crosses a point where this method is unfit for competition constraints. That limit, around 60 seconds, is only ever reached when there are at least 11 people being trained.

Unfortunately, when a larger number of users are being trained, the picture set needs to be larger in order to decrease misfits, that is, when it misidentifies a person. Similarly, when there is a large number of users who haven't been trained by the algorithm, a larger value of $k_neighbours$ helps preventing overfitting, that is, stop the system from assuming random people to be someone who is a part of the dataset.

Considering the maximum number of people in the competition, this algorithm is still viable. When doing additional tests on the accuracy of recognition, the software seems to be able to accurately recognize the people being trained, provided there are enough pictures being taken for the training phase. For testing the system, we inserted 6 operators, with 15 pictures for each. After this, by using 23 images with 2 people from the training set in each one of them, there were a total of 7

misfits and no overfittings (meaning no one was recognized as an known operator without being one; all errors were related to real operators not being recognized in the image). This resulted in a 15.21% error rate.

Other tests were done, notably, 18 images with 2 people on each, and a 6 people dataset on each. However, only one person was a part of the database. Here, significant overfitting happened, a total of 10 cases. Missfits, on the other hand, were less likely than the previous set, only happening 3 times, totalling a 36% error rate. Besides this, two tests were done, one with 19 images with 2 people in it and one with 12 images and 4 people on each. For both of them, the dataset contained only one subject which was present in all photos. This final test found only one missfit, which identified the operator as "unknown" and 2 overfits, which marked a random subject as the one in the dataset, this achieved best results with a total of 3% errors, showing that the algorithm is best suited for cases where there are less subjects in the dataset.

It is important to understand that, while results were best when using a small dataset of people, they were still fit for use with more subjects, provided there are a couple of seconds to look at the picture and find the desired user.

VI. FUTURE IMPROVEMENTS

While this paper brings new light on training limitations faced by at home service robots, it is quite limited in its scope. Firstly, a proper test database needs to be developed in the context of Robocup@Home competition, which will yield a much different dataset to what is present in most current testing sites.

This is of particular importance due to how in a home scenario it is improbable that hundreds or thousands of different faces will need to be accounted for, scale is not one of the main challenges in this environment. Another point is that not all algorithms require a cropped face for input, therefore this database needs to have images more similar to what might be seen by the camera of a home service robot, that is, sometimes multiple faces per image at different distances, resolutions and lighting environments.

In order to figure out what is best for at home purposes it is of major importance to test more algorithms and how to best utilize them within the given constraints, therefore, this paper can be complemented by additional studies exploring different solutions for facial recognition. More complex algorithms, as well as maybe a two-phase procedure are interesting ideas to be tested. A two-phased solution would work by using KNN or another light solution for the “first introduction and immediate recognition“ of any new operator, followed by a second algorithm, more computing intensive to “better integrate“ the new operator to the set of known people.

Some solutions which are worth comparing to are already being developed, One-Shot Learning [19]

VII. CONCLUSION

In summary, facial recognition software has a considerable amount of applications in the real world, among them, service robots may want to use this technology, which assists the completion of multiple home based interactive tasks. Considering its many uses, many teams have tried implementing their own solutions, mostly based on existing recognition technologies. However, finding data on their results within the competition was very difficult, and, therefore, it is important to document how technologies are being used and how well they work within the context of competition.

While KNN may not be the only simple algorithm which can be used in Robocup, it can be easily tweaked to maximize performance within given restraints. Considering the tasks present in competition, we explored what the limitations are in terms of time, permitting others to best make use of this model within the competition or within similar constraints.

This paper has looked at ways to improve training quality and found out it is possible to train this algorithm even at the limits of what the competition will ask for, provided the programmer knows how to best make use of the resources available, focusing on good quality pictures and keeping the quantity to a minimum.

VIII. ACKNOWLEDGEMENT

The authors would like to thank the support of UTFPR - Federal University of Technology-Parana. The first and second authors would like to thank the support of CNPq for their scholarships in the Scientific and Technological Initiation at UTFPR (Editals 02/2022-PROPPG and 03/2023-PROPPG).

REFERENCES

- [1] Mauricio Matamoros, Viktor Seib, Dietrich Paulus. Trends, Challenges and Adopted Strategies in RoboCup@Home. In 2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 2019
- [2] Woodrow W. Bledsoe. Some Results on Multicategory Pattern Recognition. In Journal of the Association for Computing Machinery, Vol. 13, No. 2, 1966.
- [3] Matthew A. Turk and Alex P. Pentland, Face recognition using eigenfaces. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991.
- [4] Luis A. Pineda, et al. The golem team, Robocup@Home 2017. RoboCup@Home 2017 Team Description Papers, 2017.
- [5] MFB van der Burgh, et al. Tech united Eindhoven @Home 2017 Team Description Paper. University of Technology Eindhoven, 2017.
- [6] Gordon Cheng, et al. Alle@home 2017 team description paper. RoboCup@Home 2017 Team Description Papers, 2017.
- [7] Jesus Savage, et al. Pumas@home 2018 team description paper. RoboCup@Home 2018 Team Description Papers, 2018.
- [8] Kosei Demura, et al. Kit happy robot 2018 team description. RoboCup@Home 2018 Team Description Papers, 2018.
- [9] Sven Wachsmuth, Florian Lier, and Sebastian Meyer zu Borgsen. Tobi - team of bielefeld: A human-robot interaction system for robocup@home 2018. RoboCup@Home 2018 Team Description Papers, 2018.
- [10] M.T. Lázaro, et al. Spqrel 2018 team description paper. RoboCup@Home 2018 Team Description Papers, 2018.
- [11] Jinyoung Choi, et al. Team aupair technical paper for Robocup@Home 2018 social standard platform league. RoboCup@Home 2018 Team Description Papers, 2018.
- [12] Jianghao Huo, et al. Robotics @Home 2019 Team Description Paper. RoboCup@Home 2019 Team Description Papers, 2019.
- [13] Bruno F. V. Perez, RoboFEI@Home Team Description Paper for RoboCup@Home 2019. RoboCup@home 2019 Team Description Papers, 2019.
- [14] Jeffrey Too Chuan Tan, KameRider OPL @Home 2019 Team Description Paper. RoboCup@Home 2019 Team Description Paper, 2019.
- [15] Rémi Fabre, CATIE Robotics @Home 2019 Team Description Paper. RoboCup@Home 2019 Team Description Papers, 2019.
- [16] Naomi S. Altman. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. In The American Statistician. 1992.
- [17] Assin Kortli, Maher Jridi, Ayman Al Falou, Mohamed Atri. Face Recognition Systems: A Survey. in Special Issue, Biometric Systems. 2020.
- [18] Stuart Russel, Peter Norvig. Artificial Intelligence A Modern Approach, 3rd edition. Pearson. In 2009
- [19] J. P. R. Belo, F. P. Sanches and R. A. F. Romero, “Facial Recognition Experiments on a Robotic System Using One-Shot Learning,” 2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE), Rio Grande, Brazil. In 2019