Learning Names through Facial Recognition

Ailyn Aguirre, Anjuli Goring, Matthew Webb

May 12, 2017

Contents

1 Abstract

In order to be marketable and more commonplace in our environments, robots must be able to bring an emotional connection to the human-robot interaction. This project implements facial recognition software that allows the BWIBots to identify people and interact with them by saying hello or goodbye along with their name. In order to do this, we created a database that matches names to faces so the robot can recall the names of people it has seen before. If the robot sees a new face, it will learn their name by asking the individual to enter it into the console. We expanded this project so it even works with groups, including mixed strangers and friendly faces. Possible future implementations of this development include attendance systems and personalized human-robot interactions.

2 Introduction

Our project implements facial recognition software, creates a database that maps faces to names, learns new faces by interacting with the people, communicates with them through natural language, and deals with people in groups of mixed strangers and friends. This project utilizes the Haar cascades facial recognition technique because it has fairly high accuracy and allows flexibility of the faces positions. This method works by imposing rectangles onto images classifying them as positive or negative to identify if the image is a face. We also used the *sound_play* library to create the spoken-language dialogue between the human and robot, which is a big component of incorporating HRI. This library provides a ROS node that translates commands of a ROS topic into synthesized speech. The node receives the name as a string when it is typed in by the user, and converts it to audio. This ability to talk to robots would become very useful in the future, to do things such as report information and create friendly relationships.

3 Background

Our project plans to build a bridge between facial recognition and natural language processing. Unlike humans, for computers and robots, the ability to recognize faces and use speech must be implemented. The approach for facial recognition that we took utilizes the Haar cascades technique. OpenCV offers a method that utilizes a grayscale form of template matching to find eyes and a mouth. This method has many edge cases: slightly turned faces, covered faces, partial faces. Wright et al. proposes a sparse algorithm, which implements a NFL (nearest feature line) algorithm but considers all possible supports and adaptively chooses the minimal number of training samples needed to represent each test sample. Instead of processing pixel by pixel and building patterns on the pixel level, the algorithm can split the image up into regions of interest - rectangles, for example - and use a single-search algorithm in that region. This speeds up processing time and removes the need to process images globally multiple times, as Haar cascades currently does [3].

When the robot prompts the user to enter their name, the person has to go around to the other side of the robot, type in their name, then go back around to the other side of the robot so it can say Hello John. It will then put together this name with the chosen phrase from the the greetings and say it using the *sound_play* node. However, as we see in programs such as Siri or Amazon Echo, words and especially names can still be mispronounced, which is why NLP is still a quite heavily researched topic in current projects [4].

This project would combine existing facial recognition software with NLP algorithms to improve human-robot interaction by building a more natural feeling relationship. Many HRI projects are currently focusing on having physical interactions with humans in such a way that would allow robots to be integrated into daily life more easily. Our project aims to improve the more personal aspects of HRI.

4 Technical Approach

Until the robot detects a face, it executes an idle state in which the robot aimlessly drives, waiting to interact with a human. Once the robot sees a human, it will have to determine which path of action to take. The first option: the person will turn out to be a stranger, in which case the robot asks the person for their name and stores that along with an image of their face in the database. Alternatively: the person is a known face - friend - and the robot will simply greet them.

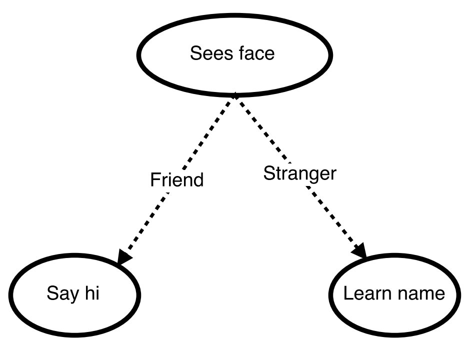

Figure 1: LT1 representation

The first component of this involves implementing the Haar cascades technique to give the robot the ability to recognize and distinguish faces. We chose to use OpenCVs implementation because it is more accurate, so it will detect when a face is in frame in constant time. It also has a tilted and rotational feature that allows the face to be at a slight angle or turned to the side and still be recognized. This technique imposes simple rectangular features onto grayscale images which represent the difference of the sum of pixels in that area. Each feature will indicate whether a certain characteristic is present or not in an image. It is common that the eye and mouth regions are darker than the cheeks and chin, so rectangles are placed on these areas to detect if a face is present in the image.

The second component is to be able to classify the identified face as a friend or a stranger. In order to be able to do this, we have created a database that stores an average image of a persons face and maps it to their name. We chose to use dynamic representations of peoples faces for the images because it allows for slight changes in the image, such as if a person who usually wears glasses chose to wear contacts for the day. In order to account for this slight variation, we also implemented a threshold function. This function determines how different an image is allowed to be from all the stored images in order to still be categorized as a friendly face. If there is no match to the current face or it exceeds the threshold, that means the robot has never seen this person before so it will have to prompt the user for their name. It will then store an image of their face in the database and map it to the newly learned name. We also have a method that marks faces as present if they are currently in the screen, so if another face that is similar to the first face appears in the screen, the robot will realize that name is already taken so it must be another person.

The last step is to say hi or bye now along with the new name. We have given the robot a set of greetings and farewells that the robot can cycle through when a face comes in and out of frame. This way the robot will not always say the exact same phrase every time, adding an additional element to HRI. It simply randomly selects an expression from the hello or the goodbye category, goes through the database to find the matching name, and then says it using the ROS *sound_play* library. This library includes a ROS node that translates commands on the *robotsound* ROS topic into sounds by using speech synthesis via the festival node. This way the text entered by the user can be stored as a string and then converted into sounds.

After all these steps have been carried out, the robot enters into its idle behavior, where it randomly chooses between four doors and roams between them to mock a wandering state. The robot will continue doing this until it detects the presence of at least one face, in which case another *SimpleActionClient* is fired. This second *SimpleActionClient* sends a navigation goal of 0, 0, 0, 0 so the robot will stop immediately, interrupting its idle behavior to interact with the face. From here, the robot will try to classify the face as a stranger or friend, restarting the process.

5 Evaluation

Before the facial detection and interaction portion of the program could be tested, we had to ensure that the idle behavior was functioning properly. In order to evaluate this, we let the robot roam freely and made sure its roaming behavior was interrupted when it detected a face. To check whether our implementation was correct, we conducted four stages of testing: single stranger, single friend, small groups, and full scale. The first stage was evaluated by introducing a single stranger and determining if the robot correctly identified the individual as a stranger. Next, we had a friend enter the frame and determined if the the robot correctly mapped their face to their name. For the third stage, the robot interacted with small groups of friends and strangers, and was successful if the robot could differentiate faces within the group. Lastly, the robot was subjected to full scale random tests around the BWI lab, at which point success was determined by the robots capability to accurately identify and get to know people.

Throughout the testing, the robot was successful in most scenarios. In specific edge case scenarios, we discovered multiple issues, some due to our code, others due to OpenCV source code. One issue within our control is the implementation of the idle behavior. Currently, two *SimpleActionClient*s are used to start and stop idle behavior, resulting in twitching mid-conversation. One problem out of our control occurs if people look too similar: both will satisfy the threshold and the robot will confuse them with each other. However, if the threshold is lowered, the robot cannot correctly identify people if they happen to look slightly different from their stored image.

6 Conclusion and Future Work

This project improves interaction between BWIBots and people within the GDC by utilizing facial recognition and text to speech processing. The robots creates personalized interactions for the users by learning something unique to each user. By having the robot physically talk, we hope people will feel more connected to the robot and see it as more than just a computer with wheels, advancing HRI. One possible future expansion would be improving the facial recognition. The existing software cannot recognize a face if it moves too fast, is even slightly covered, or is rotated past a certain point. Another idea would be to associate people with locations they are commonly found at so the robot could easily find people based on a probability map. Both of these extensions would lead to improved HRI and a better experience for the consumer.

References

[1] Coeckelbergh M. (2011) Talking to Robots: On the Linguistic Construction of Personal Human-Robot Relations. In: Lamers M.H., Verbeek F.J. (eds) Human-Robot Personal Relationships. HRPR 2010.

[2] Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots." *Robotics and autonomous systems* 42.3 (2003): 143-166.

[3] Ganesh, S. S. Sastry and Y. Ma, "Robust Face Recognition via Sparse Representation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, Feb. 2009. doi: 10.1109/TPAMI.2008.79

[4] Lopes, L., & Teixeira, A. (n.d.). Human-robot interaction through spoken language dialogue. *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No.00CH37113).*

[5] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," *Object recognition supported by user interaction for service robots*, 2002, pp. 511-514 vol.2.

[6] N, Rathan and A., Venkatesh. Design and Implementation of Soft-I-Robot for Face Detection and Obstacle Avoidance Using Neural Networks and Genetic Algorithm (February 2017). *Asian Journal of Applied Science and Technology (AJAST)* Volume 1, Issue 2, Pages 120-126, March 2017.

[7] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.