Learning Names through Facial Recognition

Ailyn Aguirre, Anjuli Goring, Matthew Webb

April 2, 2017

Contents

1 Abstract

The goal of this project is to implement facial recognition software that will allow the BWIBots to identify people by having them stand in front of the camera. In order to do this, we will have a database that matches names to faces so the robot can recall names of people it has seen before. If the robot encounters an individual whose face is not in the database, it will learn the strangers name by asking them and using natural speech interaction to store the new information. Eventually, this can be expanded upon to be used for groups of people which may include multiple strangers amongst familiar faces. The end of the paper discusses future implementations and some of the many potential applications of this development, such as attendance systems or personalized human-robot interaction.

2 Introduction

Thanks to Hollywood, most of us have been exposed to an idealized version of human-robot interaction (HRI) since a young age. In order to implement our plan there are multiple problems that we need to address. This includes the concept of recognizing and distinguishing people through facial recognition software, creating a database that maps faces to names, learning new faces through natural language communication, and dealing with people in groups of mixed strangers and friendly faces.

The problem of facial recognition seemed unsurmountable not too long ago, but today there are many publicly available facial recognition softwares that have high accuracy and incredibly large connected databases [3]. We plan on combining this with spoken-language dialogue to create more natural communication between the human and robot. This next level of HRI implements a personalized approach to the relationship, since the robot would be talking to the person as if they knew each other [2]. Noun and verb-phrasing and separation algorithms will allow the robot to select important words from responses spoken by individuals, such as when they are prompted for their name. This would become very useful in the future, if the robots were to further interact with humans and take verbal commands or instructions. Currently there is not a large enough divide between human-robot relationships and human-machine relationships, which are non-personal and highly predictable. In order to be marketable or more commonplace in our environments, robots must be able to bring an emotional connection to a previously physical interaction.

3 Related Work

As far as facial recognition and natural language processing (NLP) research projects go, they are generally in separate fields of development. Our project plans to build a bridge between these very different subjects. As humans, facial recognition is something that comes naturally to us - even one to three day old babies have proven the ability to differentiate between known faces. However for computers or robots, this ability must be implemented. Luckily, facial recognition software has made incredible strides in advances in just the past few years. At this point there are many different algorithms for identifying facial features, for example distinguishing people by their features versus looking at a basic template. One of the most intuitive approaches to face recognition is based on the geometric features of a face. Marker points are based on noteworthy features and the feature vectors are then built between these points.

Calculating distances between these vectors and a reference image allows the computer to recognize a face. Since this method is not yet perfect - because geometrical features alone may not suffice for specific differentiation between similar faces - other methods, such as utilizing eigenfaces, which incorporates a much more holistic approach, are being experimented with [1].

On the other hand, NLP is concerned with computers and robots being able to process human speech. For a while, most NLP systems were based simply on hand-written rules, but more recent research is more concerned with creating machine learning algorithms for language processing. These algorithms include categories such as lexical and morphological analysis, noun phrase generation, and word segmentation. In the context of this project, noun phrasing is an important technique because it is key in information retrieval. When the robot asks the stranger for their name, we want to be able to separate my name is from John Doe otherwise the robot may map the picture of the person with my name is John Doe which is incorrect. This simple noun-phrase-based system has been found to be very accurate, as it performs just as well as a key phrase extractor [2]. However, as we see in programs such as Siri or Amazon Echo, voice instructions can still be misheard or misinterpreted, which is why NLP is still a quite heavily researched topic in current projects.

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 emotional aspects of HRI.

4 Technical Approach

The most direct approach to giving a robot a behavior is through a logic tree, in which each branch is a decision or observation the robot has to make, and each node is an action the robot will take. At the top of the logic tree is the entry point, which is an event that allows the robot to enter the behavior. Until the entry point event is fired, the robot will enter a null state and do nothing but observe and wait for the event. Ideally, a logic tree will span every single possible action the robot can take (i.e. move, turn, localize, detect, speak), but our initial logic tree will only deal with basic actions like saying hi to a friend and identifying a face. However, this project will not simply define a deterministic behavior; the goal is to utilize a learning action at one node of the logic tree so that the robot will learn only when it needs to. Using this form of

constrained learning, the software implementation becomes a lot simpler yet the project still holds merit. After the logic tree is defined, the robot should be able to autonomously learn to identify and address people using its speech and vision packages, an ability that none of the BWIBots currently have.

LT1 - logic tree version 1 - will be the logic tree implemented at the beginning of the project to test basic functionalities of our implementation (Figure 1). The robot will enter LT1 when it sees a face and branch based on whether the face is a friend - someone the robot already knows - or a stranger. If the face is a friend, LT1 will reach an end node where the robot will say hi to the friend and exit. If the face is a stranger, LT1 will direct the robot to prompt the stranger for a name and learn that persons face for the future. This behavior is not complex, but it satisfies all the requirements for a human-robot interaction and it learns based on previous encounters.

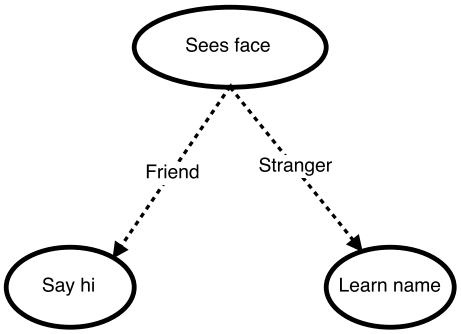

Figure 1: LT1 representation

Although LT1 is a satisfactory implementation, it does not take full advantage of the NLP, learning, facial recognition, and speech systems at our disposal. LT2 is based on LT1, but takes into account visual edge cases like a turned or partial face, and how to learn from that image. In order to improve HRI, LT2 also has more conversational end nodes, such as saying goodbye when a friend leaves frame and asking people for their names if they are wrong or misspelled. Lastly, LT2 implements a roaming behavior while it is not in any stage of the logic tree. Essentially, the goal of LT2 is to extend the technical success of LT1 into tangible HRI results by making the robot more accessible.

Logic trees are a very high level approach to determining what the robot will be doing at any given time, but success of this project relies on the implementation of each action node. The most important algorithm is the function that converts a picture to a name, based on some map of faces to names. This function cannot simply return the closest match; it has to utilize feature and

color matching to determine if the face is a friend or stranger. If the face is a friend, the function will return the name, otherwise it will not return a name. This leads to another implementation choice: how the faces are stored. The straightforward method is to save images of the friend into an array and map that to a name. Alternatively, the robot can save average images to save space and increase processing speed, but that may distort the faces over time. Both will work to different degrees, so the correct method must be determined through extensive testing. A lot of the implementation choices in LT2 will be made when the code is written depending on test results.

LT1 and LT2 are not perfect behavior models, of course, and one can predict problems with their implementations. Firstly, the picture-to-name function needs to know the definition of a friend and a stranger. This comes down to setting some threshold, which will determine how similar a picture must be to the average image in order to be classified as a friend. LT2s implementation utilizes feature and color mapping, so it will have less of a problem with this than LT1s. Another problem arises when determining when to say a face enters or leaves frame. OpenCVs facial recognition does not work well with partial faces for obvious reasons, but our robot should be able to handle partial faces in a real world situation. Ideally, an LT3 model will be developed that integrates movement into the robots learning action nodes so that if it does encounter any visual edge cases, it can overcome them (i.e. turn to see a full face). For LT2 and especially LT1, ideal test cases will be run so that implementation details can be worked out. The last major problem lies in detecting names of people. Assuming LT1 asks for just a name, the only subproblem would be ensuring that the name is spelled/pronounced correctly. However, to satisfy the improved HRI goals of LT2 and LT3, the robot should be able to comprehend standard English introductions exemplified in the previous section (Hi I am John, My name is William, Im Jane Smith) without templates/patterns. LT2 and LT3 then would have to deal with another subproblem: relearning incorrect names. If a face has been incorrectly mapped to a name, LT2 has no mechanism to fix that problem and the robots confidence in the incorrect name will only grow. To fix this, LT3 will need to have a relearning end node that remaps a face to a new name and transfers any learned information.

As LT1-3 are created and the implementation is developed, there are a set of technical goals that we wish to reach. Firstly, the picture to name function must be extremely reliable and efficient. LT1-3 will simply result in erroneous behavior if this function does not work. Next, the goal is to get a robot to learn a strangers name from a constrained set of faces. In practice, that means the robot will come preset with multiple face-name mappings, and

will have to learn new ones. Success at this stage occurs when the robot is able to differentiate between friends and strangers. After that, the goal will be to have the robot interact with random strangers successfully. If this step becomes reliable, it will be thoroughly tested in controlled sets of visual and learning edge cases. Lastly, the robot will interact with a group of people and communicate with friends and strangers within the group. And finally, if this is successful, LT3 will implement a roaming behavior allowing the robot to explore the GDC and learn in a real world environment. At this stage, success means the BWIBots are able to interact and get to know people in the GDC. From here, extensions become much simpler thanks to the framework laid out by these technical goals.

5 Evaluation and Expected Results

To check whether the implementation is correct, 3 stages of testing will be performed: single stranger, small groups, and full scale. The first stage will be evaluated by introducing a single stranger and determining if the robot correctly mapped their face to their name. Next, the robot will interact with small groups of friends and strangers. In this stage, the testing will be considered successful if the robot can differentiate faces within the group. Lastly, the robot will be subjected to full scale random tests around the GDC, at which point success will be determined by the robots capability to accurately identify and get to know people.

6 Conclusion

The purpose of this project is to improve HRI between BWIBots and people within the GDC, by learning their names and faces. Using facial recognition and voice packages, BWIBots will have the ability to create more personalized interactions for tasks that are purely mechanical (i.e. getting a coffee). By having the robot greet the people that it has met, we hope that people will be able to feel connected to the robot through the personalized interaction and see it as more than just a computer with wheels.

7 Possible Extensions

Our project allows BWIBots to create more personalizable experiences. For instance, a future project could associate people with locations they are com-

monly found in. This would allow the robot to more easily find a person based on a probability map. Secondly, another project could involve learning details about a person, such as age, gender, or position. Either way, any extension of this project will lead to improved human-robot interaction and a better experience for the consumer.

References

[1] Kanade, T. *Picture processing system by computer complex and recognition of human faces.* PhD thesis, Kyoto University, November 1973.

[2] 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).*

[3] Pentland, A., & Choudhury, T. (2000). Face recognition for smart environments. *Computer; 33(2)*, 50-55.