Roombot Project Proposal
Developmental Robotics
10/28/17
Sonal Chatter & Chris Hylwa

I. ABSTRACT

Roombot—This paper describes a next generation robot, Roombot, that classifies rooms using visual inputs and machine learning. We extend what it means for artificial intelligence to understand what the function of a room is by moving beyond just serial object recognition and recall. Our robot will intelligently determine what the purpose of a large, but enclosed, space is based on the components found in the room. Over time, it will become more accurate by getting feedback on its classifications, and continue to learn and make modifications.

II. INTRODUCTION

Recognizing objects is one of the first things we are taught as infants. Distinct associations such as “Mom”, “Dad”, “toys”, “food”, and “bottles” are created and reinforced from day one, and we learn to distinctly distinguish those objects quickly. As the brain develops, we reach the point where we can observe a collection of objects that we can identify and infer information about our surroundings. We see open water in a setting and we think of swimming.

One of our fundamental skills as humans is the ability to walk into a space and immediately determine its function based on visual input and object recognition in our surroundings. At the same time, one of the primary rules of interior design and architecture is to provide form and details or objects that fulfil and represent a room’s function. Finding a chemical fume hood in a kitchen, for example, is highly unlikely, save for the most intense of molecular gastronomy classes. In this way, the constant back and forth between categorizing rooms and building rooms to be categorized allows humans to more easily adapt to and utilize their surroundings.

On the other hand, computer technology has been unable to execute this simple human task with accuracy and versatility until recently. Due to great variations within classes and the multi functional affordances for many objects, it’s really hard for an algorithm to actually figure out what an object is. The computing power necessary to process images in volume has only been developed in the past decade or so, and the algorithms to recognize a large swath of different objects were fairly inaccurate until breakthroughs around 2012. Since then, we have come a long way, but there is still great work to be done.

Technology today still cannot provide holistic recognition when surveying a scene. It can enumerate all the objects in a picture, but it still has limited accuracy for pinpointing the functionality of a room when considered as a whole. Roombot is our attempt to address this shortcoming, which is a pressing need in many areas. The most obvious is providing accessibility to the visually impaired. Roombot can survey a room, and rather then just list a bunch of objects to the user, who then has to determine the room type, Roombot will just state the room. Another accessibility concern would be for people who are movement-impaired. Being able to discern the general use for each room would also give the robot a head start to not only navigating the space, but also a general primer for being able to find a given object. On top of

that, if that person asks to be taken to a specific type of room in an unfamiliar space, we want the robot to be able to stop when it recognizes that it is in the correct room. Another use case of this technology would be for investigation, and or security, especially if this project is extended into other spaces. For example, being able to survey hostile spaces for reconnaissance, or investigating spaces such as archeological sites, where it may be unsafe for humans to enter.

III. RELATED WORK

There has been a great deal of work done in academia and industry. Here, we will note a few papers from academia that are relevant and provide inspiration for our work as well as a few commercial products that exist in the industry today.

3.1 INDUSTRY

There are many different products in the market that recognize objects, and implement various aspects of human vision. We take most of our inspiration from Microsoft Seeing AI, Google Vision, and Clarafai. We will be using Clarafai's API as a tool in our project as well, and possibly Google Vision's API too.

3.2 ACADEMIA

Places [9]

MIT's CSAIL has the cutting edge in scene recognition in terms of both dataset and accuracy. We will use this as a benchmark and guide as we proceed with our work.

How Robots Learn to Classify New Objects Trained from Small Data Sets [7]

This paper is interesting to us as a sort of case study, since our project, by its very nature does not have a lot of data to train. The paper is an interesting analogue and useful for experiment implementation.

Indoor scene recognition [8]

Since Roombot is only for indoor areas, this paper is very useful as it discusses the specifics of dealing with defined and cluttered spaces.

Scene recognition through audio input [10]

While we are focused on visual input, this paper provides a great analogue to our goals. It also points out that there are way too many inputs in a given scene, and many are irrelevant so greedy algorithms are very useful. This technique is something we will strongly consider in our work.

IV. PROBLEM FORMULATION

Given a set of clearly defined indoor rooms, a set of possible purposes or identifiers for that room, and the ability to navigate each room, can we train a robot to be able to recognize the function of that room to a reasonable degree of accuracy?

V. TECHNICAL APPROACH

Below we breakdown our approach into three main areas; the hardware, which are the physical objects we are using, the software, which is all the code, APIs, and languages we plan on using, and the integration of all the components into a comprehensive project.

5.1 HARDWARE

For our project, we would like to take advantage of the TurtleBot units provided to us in this course. These robots are an open source robotics project developed by Willow Garage, and are small, lightweight, and generally useful for these kinds of experiments. While we do not currently know the exact camera that will be mounted onto each of the robots, we assume that it is a reasonably high-quality camera that will be able to take pictures. We also know that the turtlebot netbooks are not only compatible with ROS, but also have wireless network cards. Since a lot of our method relies on being able to send data to more powerful image processing APIs, this is essential for our project.

5.2 SOFTWARE

There are several main pieces of software that we wish to focus on for the sake of our methods. The first is ROS, which is the open-source robotics framework that the TurtleBot runs on. We expect to be doing a substantial amount of coding on top of this software in order to interface with the rest of our experiment. This setup also means that we're going to have to do a lot of data transfer programmatically, as well as possibly adding some sort of simple machine learning classifier on top of that in order to learn the different associations between room types and the labeled objects found within.

We also will be using an object recognition API in order to determine what objects are in a room, and use them as vectors in our classification of which room. We have several possible platforms for this, however, we are most likely going to focus on using Clarifai. Clarifai is an API that analyzes images and returns predictions about what is in the image based on the model that you send the data to. So for example, if we send a photo to the "food" model, we will get classifications like "apple", "soup", or "peanut butter." For the sake of our experiment, we are going to try using the standard model for object classification, but fully expect that we will have to train a custom model on room and scene data specifically, as well as any objects we might consider useful for classification.

In addition to the aforementioned software, we also will be using APIs for movement, especially those for simple autonomous exploratory movement, and preventing collisions with obstacles and objects.

5.3 INTEGRATION

From a general, overhead perspective, most of the interfacing between our robot and the API, as well as the guessing mechanism will be done within the ROS system. This system will not only control the camera and move the robot, but also send data to whichever APIs we use. This also means that the obstacle avoider API and autonomous movement API will also be sending signals to our robot through ROS. While we have not yet devised the mechanism that will provide positive and negative feedback to the robot for the purposes of reinforcement learning, we expect that this interaction will most likely be through the physical computer

connected to the given Turtlebot. In general though, while we have multiple options in terms of exactly which APIs to choose and other software questions, the basic details of integration remain the same.

VI. EXPECTED RESULTS/EXPERIMENTAL VALIDATION

Our goal for the project is that the TurtleBot should be able to recognize a set of rooms in Halligan, defined as single-purpose types based on labels that it gets from the environment. For every room, both training and testing, we will have the robot take a series of pictures and try to gauge information about its surroundings. This will be achieved by sending these pictures to our object recognition APIs, which will then provide the robot with details and labels for what exactly is in the room. It will then analyse that data and either make an attempt at guessing the room, or be provided with the correct answer for what the room is. Note here that since we defined all rooms as “single-purpose” for the purpose of this experiment, this means that there is only one thing a room can definitively be.

During the experiment itself, we will start by leading the robot around various rooms in Halligan that are already labeled, and let the robot obtain data about how to classify each given type of room. This will form our baseline training / supervised learning session, which will allow the robot to begin to formulate a structure for how to identify a room.

Once this initial training is complete, we will continue to the second phase of training where we will take the robot into the same rooms as during the initial training session, and test whether or not it can reproduce the correct label after exploring. This also means that the robot may be positioned in a new place within a familiar room, and will therefore have to make more inferences about the room or type of room that it is currently inhabiting. Once the robot comes up with a guess for the type/purpose of the room it is in, we will provide positive or negative feedback based on whether or not the robot’s guess was correct. We will repeat this process for the robot as the first part of our reinforcement learning procedure.

Finally, we will take the robot to rooms that were not a part of the initial training data, and ask it to explore said rooms and classify them based on its current model. As before, we will provide the robot with positive or negative feedback as warranted by its guess. This will form the final part of our training, and will also be a good final standing for the experiment. If at all possible, it would also be good to test in rooms outside of Halligan at this point in experimentation.

VII. TIMELINE

We expect to begin coding as soon as possible and plan on sticking to the schedule outlined below.

1. Code (2 weeks) - November 12th
2. Progress Report - November 16th
3. Run Experiments - November 13th-21st
4. Analysis (concurrent with running experiments) - November 13th-21st
5. Additional Experiments + Analysis - November 27th - 30th
6. Write paper (1 week) - December 1st - 7th

We believe this will give us sufficient time to complete all major aspects of this project in a meaningful way. While it is a tight schedule, we fully consider this timeline doable.

VIII. FUTURE WORK

In creating this experiment, we notice that our methods do make several assumptions about rooms and how “room types” are assigned to them. First, every room in our experiment is assumed to be single-use, which is not analogous to the design of the world we live in. Rooms such as the “collaboration room” in Halligan Hall for example could have multiple assigned uses depending on the people within and the setup. This would seem to function more as a tagging system rather than a binary classification system, which would complicate our methods severely. However, this also would be fruitful ground for future experimentation with more sophisticated forms of classification.

We also specify rooms to be spaces that are clearly delineated by walls or partial walls on all four sides, and/or spaces demarcated by changes in flooring. This doesn’t account for a more “open floor plan” residential design such as studio apartments, where rooms such as the kitchen, dining room, or living room may not have distinct divisions between them, but none the less have a defined area associated with them.

In expanding our research and in continuing study, we would like to explore spaces that are not exclusively inside Halligan Hall, not only in terms of more and varied academic spaces, but also spaces that are designed for other needs. This would include residential, corporate, and communal spaces, as well as spaces that have been destroyed or dilapidated. This would not only increase the number of possible rooms, but would also contribute to some of the human-centric problem space as defined in Section IV.

We would also like to explore being able to classify rooms based on multiple types of sensory input. We would like to work with the audio input sensory recognition algorithm in conjunction with the visual input that we have. An idea that we thought may be interesting would be for a robot to listen to how people in the room are speaking, or the objects in the room they are referring to, and be able to classify the room based on that data as well. So, for example, if a robot is eavesdropping on a conversation in a room where two people are discussing laboratory equipment, we want it to be able to categorize the room, tentatively, as a laboratory. Obviously, this would be used in combination with other classifiers as a way to make the robot’s inferred understanding of a room more certain and/or correct. This does not have anything to do with humans speaking directly to the robot themselves, but rather, simply what the audio sensors happen to pick up on.

If we can find extra time within the already cramped schedule, we would like to at least begin to explore one of these possible future avenues. Our first avenue of expansion would be to take the TurtleBots outside of Halligan in order to classify more room types. Specifically, we would like to test our algorithm in other science and technology buildings as well as residential spaces. The recent SciTech wing would be a great example, and would give us several new types of rooms to work with, while using more residential data would allow us to start exploring the potential practical uses of our research.

References

1. https://www.microsoft.com/en-us/seeing-ai/
2. https://www.clarifai.com/
3. https://cloud.google.com/vision/
4. http://aipoly.com/
5. https://www.producthunt.com/alternatives/seeing-ai
6. https://medium.com/@nikasa1889/the-modern-history-of-object-recognition-infographic-aea18517c318
7. Tick Son Wang, Zoltan-Csaba Marton, Manuel Brucker, Rudolph Triebel: http://proceedings.mlr.press/v78/wang17a/wang17a.pdf
8. https://dl.acm.org/citation.cfm?id=2506925 Indoor scene recognition by a mobile robot through adaptive object detection (2013): P.Espinacea, T.Kollarb, N.Royb, A.Sotoa
9. http://places2.csail.mit.edu/PAMI_places.pdf
10. https://www.researchgate.net/profile/Maja_Mataric/publication/221263267_Where_am_I_Scene_Recognition_for_Mobile_Robots_using_Audio_Features/links/00b49527481161c422000000.pdf
11.