Introduction:

Semantic identification is by one of the most fundamental as well as rudimentary byproducts of human perception. Being able to distinguish a given area A from another potential area B is particularly important when considering human perception in regards to spatial awareness. Comprehension of area and space allows humans the ability to more readily navigate certain locales at any given location. Although these abilities seem rather trivial to a human, imbuing this advanced perception on robots can prove to be quite a challenge as robots cannot inherently detect nor infer nuance or subtleties in given environments.

For example, consider the task of finding “Room 10” in a hallway that possesses ten doors (where each door hangs a badge identifying itself as a room from 1 to 10). Again to humans, this concept is rather elementary; in brief, the person in question would merely have to navigate to each door until finding the room with the appropriate badge that reads “Room 10”. A robot, on the other hand, would not be able to distinguish “Room 1” from nor “Room 10”. In fact, the robot wouldn’t be able to even distinguish a wall from a door.

Herein lies our proposition: what if we were able to create a robot that navigated through a given locale and would be able to then label certain locations with salient information? Thus if a human wanted a robot to find “Room 10”, the robot would simply need to place itself at the location, navigate around that site, and find the appropriate doorway (not that dissimilar from how a human would apply themself in the same scenario).

Accomplishing this particular project would open a number of more complex operations that could be executed as a result. For instance, a fully autonomous cocktail waitress robot could detect which door in Halligan it would need to approach without needing exact coordinate locations; rather, the robot would be able to simply semantically apply labels on certain locations and independently determine its goals.

Outside of the realm of Halligan, this project is essentially derivative of being able to generate an elementary form of autonomous movement and perception. Meaning, this robot is a foundational step in ultimately understanding how exactly can humans ingrain a semblance of human perception into robots and or other robotic systems.
Related Work:

This project will be based on many open source and existing libraries. One approach from the Queensland University of Technology of vision based semantic mapping has done similar work on recognizing certain objects and building a semantic map on top of another existing 2D floor plan.


More information about the general concept of semantic mapping has been done by Kostavelis and Gasteratos at the Democritus University of Thrace, Greece.

https://ac.els-cdn.com/S0921889014003030/1-s2.0-S0921889014003030-main.pdf?_tid=06c56d18-1ef7-4693-9cf4-dd0e837e9daa&acdnat=1521235268_080cd78cb208a74cd98371d1337a

In terms of tools used, one of the main tools used will be OpenCV, for which there exist a variety of ROS compatible packages that will be used in this project. More detailed information about this package can be found at:

http://wiki.ros.org/vision_opencv

One similar project is traffic sign detection, and this project will rely on the work done in that field to accomplish the desired results. One interesting approach that could be used as a starting point is described by Yuille et al at:

https://pdfs.semanticscholar.org/32d8/bf9cc01215c648ffef9d53d46d3cf2dc0.pdf
**Technical Approach:**

The main objective of the project is to develop an autonomous way for a TurtleBot robot to navigate a building, read the door signs of the rooms and offices in the building, and generate a semantic map based on this data. In the specific case of this project, the building that will be mapped is Halligan Hall, the Tufts EECS department office. There already exists a 2D map of Halligan Hall, but in order to navigate this map, the user has to specify the coordinates where it wants the robot to go, or give the robot a direction and a speed. This is not very intuitive for a person to use, so the goal of this project is to add an additional semantic layer on top of the existing map. This allows the interaction between the human user and the robot to be more intuitive.

There are certain rooms of interest that are important for the robot to detect, while others are not too useful. In general, the robot should add to the existing map of the building rooms that are either classrooms, laboratories, or offices to the map. One feature that all of these rooms share is that they are numbered rooms. This separates them from other rooms that are not important for the robot to know, such as elevator control rooms and restrooms. When talking about these rooms of interest, they will just be mentioned as rooms from now on, ignoring the ones that are not going to be part of the semantic map.

This project contains several parts. These are: adding a camera to a TurtleBot at the height of the door signs in Halligan Hall, creating of a complete map of Halligan Hall, navigating all the rooms in an existing map autonomously, identifying the doors of the building and the door signs using computer vision, deciding which of the rooms are of interest, and adding a semantic layer on top of the original 2D map of Halligan Hall that includes the rooms of interest.

A camera needs to be attached to the TurtleBot at approximately 1.5 meters in order for it to see the door signs and identify doors. In order to connect this camera, an extension pole needs to be added to the top of the TurtleBot robot. This extension would then house a webcam connected to the computer mounted on the robot. The setup for this camera should be straightforward.

In order to create the map of Halligan Hall, the TurtleBot existing libraries will be used to create a map based on SLAM, which stand for Simultaneous Localization And Mapping. Ideally, all of the rooms of interest will be mapped, including professor offices. If it were not possible to obtain permission to do this for some professors, the approach could be modified to only identify where the doors are that lead to the rooms of interest.

The robot needs to use computer vision algorithms to perform two tasks. The first task is to identify doors and door signs. A number of approaches to this problem can be applied in order to
test which one works best. One approach would be to use OpenCV to identify certain types of features in each frame of the video. One feature that could be very useful is that all door signs share the same green background color and they all have a rectangular shape. The ratio of the sides of these signs is also similar, although there are certain exceptions, such as the signs that for rooms that contain multiple offices. The doors could then be identified using the characteristic wooden color, or by extracting certain common features from the door handles. It remains to be seen how much processing power this approach needs and whether it can be performed in real time or not.

Another approach to this same problem would be to use YOLOv2, which is a visual classifier that works well in real time. This approach uses a model called Darknet that can be retrained to identify certain objects. In this case, the model could be retrained to identify the door signs and door handles. This, however, requires a lot of computation power, which could be hard to achieve in real time. Therefore, one approach using this method could be to record the video of the robot navigating the 2D map while also recording the x and y coordinates of the robot on the 2D map, and then running the identification of the door signs on a more powerful machine after the whole map has been traversed.

The second computer vision problem of this project is to transform the text found in the picture, in this case the room number and the name of the room, into text that can be added to the map. There are two approaches to this problem. The first one is to use OCR using existing Pythonic implementations with OpenCV. OCR stands for Optical Character Recognition, and it is widely used, which means that there exist multiple open source implementations that could be implemented or modified for the TurtleBot. The second approach to this problem could be to use deep learning in order to implement OCR. It is unclear, however, how computationally expensive this would be, and whether it would require a GPU to do the processing or not.

The last part of the project would be to build the semantic layer on top of the map of Halligan Hall. There are a variety of open source semantic mapping repositories on GitHub that could be either implemented or modified in order to generate the semantic map.

To summarize, the project is mainly computational, as it does not require the creation of any mechanical or electrical components apart from the webcam extension. The main input to this system would be the visual feed from the webcam as well as the odometry information for the position of the robot. The only output to this system is the semantic layer on top of the 2D map of Halligan Hall.
**Evaluation:**

There would be a number of milestones by which each leg of the project can be evaluated. First and foremost, is the robot capable of autonomous movement in a given area? This project is designed around a “wander” movement option that enables the robot to navigate a certain locale for an indefinite amount of time. Each semantic ID operation would be gradually built upon this infrastructure as the robot certainly cannot apply IDs without first being able to maneuver towards. This first leg of the project will be judged on speed, efficacy, reaction to collision, and finally where the “wander” movement can be consistently reproduced.

The second significant milestone would be whether the robot could then be able to perform this “wander” movement while also properly identifying certain salient locations and details. Essentially, this would first be tested in a controlled environment where we would allow the robot to approach a certain door, independently identify that this is pertinent information, store that data, and repeat. This leg of the project would again be judged on speed, efficacy, reaction to collision, and reproduction.

Once these two primary features have been developed, the third and most important leg of the project is to combine these two functions in a realistic setting: Halligan. If the robot is able to “wander” from one starting point to a goal point and correctly identify each location as a relevant piece of data, then the robot is deemed to have successfully completed autonomous semantic mapping.

From here on, we would then manually place semantics onto these independently identified locations such that the robot would know that the location is saved in memory now has a more meaningful name such as “H104”. Now, the robot could potentially go to room H104 upon command without the need of more exact coordinate planning.

The ultimate goal (assuming all these other pre-steps have been achieved) would be to eventually allow the robot to independently apply meaningful names on each salient data point in real time. Consequently, the robot would be able to immediately identify a certain door as H104 and not simply something to remember for later.