CHAPTER I
INTRODUCTION

This chapter will discuss the background of this thesis, in this case, the method of memorizing an extracted feature of a detected object using visual input and then using that extracted feature to map the object according to the area. This chapter will describe this thesis's objective and list the problem this research aims to solve. This chapter will also outline the previous research that has delved into this kind of topic to support the development of the proposed method.

1.1 Background

In the current era, a program's ability to use vision as a manageable input to the system is very common. We see lots of use for this kind of thing in industrial, security, and even social settings. Unfortunately, most of the use of this method only achieves the detection status and is not used any further, even though it can achieve so much more [2] [3]. The only commercial technology that goes beyond detection status is the face recognition algorithm, which can detect a person's face but also know whose face it is. The meaning of this is that this technology is not only able to detect something but also able to memorize the thing that it detects [4].

The use of memorization in this way the human brain works is not that common, even though memorization is a must in any program [5]. Most program memorization techniques involve matching the extracted feature to the database of the program and then memorizing the match based on the database. That is enough for some applications, but the important note is that a database first needs to be built to use this method, and without it, the program can only detect the object, not recognize the object. Meanwhile, human brain memorization can easily recognize the object, even if the real object has a different feature from the memorized object [6]. The other problem that can benefit from memorization is re-detection, in which a program detects the same object multiple times and analyzes its data multiple times in one session, thus wasting time and computation power.

Other projects that involve feature extraction focus on clearing the extracted feature using Principal Component Analysis (PCA) [7] or Local Ternary Pattern (LTP) [8]. Some of the projects of mapping using cameras have been known to focus on improving the accuracy of the mapping by detecting and discarding potential

dynamic objects [9] [10], increasing its effectiveness with a dense mapping framework [11], or trying to localize the camera view position based on geo-referenced images [12]. These two methods of feature extraction and mapping using a camera have been done with the following method that tries to mimic how humans localize themselves [13]. In this method, the program extracts the ”gist” [14], which is a statistical signature of the image, to obtain a rough location and enhance the rough location using salient landmarks, which reduces the time to identify the landmark by using a measure of interest in the image. It can achieve localization more efficiently, compared to previous research, by reducing the number of matches in a single image by making it into a region. Unfortunately, it cannot change its location because the camera that is used only takes an image from the same perspective.

This thesis focuses on applying a new method of memorization of an extracted feature and making it a module so it can be used by any machine learning project. This research uses CNN as a means to detect the object. This selection is based on research by Srivastava et al., who concluded that CNN has a precision of 0.716 while other algorithms such as YOLO have a precision of 0.337 and SSD has a precision of 0.247 [15].

The main purpose of this method is to apply area mapping using a robot, which not only makes the robot map an area using less time and computation power but also prevents the robot from lingering in an area that has already been mapped. This research also tries to improve the quality of a map based on visual mapping in that case, most of the time, only the outline of an object that the visuals can detect is mapped, not the entire object but only the key point of said detected object. This thesis proposes a new method that first extracts the feature of a detected object, memorizes it in the form of a table on a file used at a later time, and finally makes a map according to the table of the extracted feature that will be compared to the robot encoder for accuracy testing.

1.2 Problem Identification

The current object recognition or feature extraction algorithm does not remember what object it has detected [14], and the study in this field is a bit low. Because of this, the program spends more time and resources when it detects the same object multiple times, which isn't efficient. The other note is that visual mapping only outputs point-cloud data, restricting it to mapping only and not other problems [9]. Thus, this research propose adding a memorization algorithm to a known object recognition algorithm to improve the effectiveness of the program by mimicking

the way the human brain processes.

1.3 Objective and Contributions

The objective of this thesis is this following:

1. This thesis create a module that successfully extract an object feature and able to memorize it.
2. This thesis able to create an algorithm that can handle a view conversion from front view to top view.
3. This thesis improve the mapping of object from vision mapping by using the extracted feature.

1.4 Hypothesis

In this research, the goal of the proposed program is to become an independent module that can be used in any type of feature memorization or object detection research. It also aims to make a program able to remember things like how the human brain remembers things by making it remember certain unique aspects of an object so that the program doesn't need to compute again if the object it sees is the same object it had seen. By doing this, this research hopes that the program can reduce the time and resources needed to compute if it encounters an already memorized object. This research also aims to produce a map based on the memorized data.

1.5 Scope of Work

To keep the experiment from being too long, this thesis limits the work as follows:

1. For research purposes, the object recognition method will be using CNN.
2. The robot that will be used is a landmobile robot.
3. The field of view of the robot will be static to the robot's movement.

1.6 Research Methodology

This thesis will be divided into several work packages (WP). On this list is the following WP:

- WP 1: Literature study and review.
- WP 2: Algorithm planning for feature extraction, memorization, and mapping.
- WP 3: Algorithm trial using image and then video.

1.7 Research Plan and Action Point

This thesis is planning to do the following:

1. Make an algorithm that, after detecting an object, extracts its color, position, and shape features.
2. Make an algorithm that memorizes the extracted feature and saves it to a file.
3. Make an algorithm that is able to use the data in the file to simulate a map based on the robot view.