Constructing A Dataset For Gesture Recognition Using Ambient Light

Femi Akadiri
Supervisors: Ming Kun Yang, Ran Zhu
EEMCS, Delft University of Technology, The Netherlands
22-6-2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Preface

This paper is part of the CSE3000 Research Project course at TU Delft, which is the final part of the Computer Science and Engineering Bachelor's programme. This project consisted of 5 students working on different sub-projects of a complete system. Due to the close link between all sub-projects, they are referenced frequently throughout the paper.

I would like to express my deepest gratitude to my responsible professor, Qing Wang, whose guidance throughout the project helped me stay on track and find new approaches when I was stuck. I also could not have completed this project without the constant feedback and support from my supervisors Mingkun Yang and Ran Zhu.

I am also grateful for my group members: Dimitar Barantiev, Matthew Lipski, William Narchi and Stijn van de Water. The project would not be possible without their commitment and hard work.

Finally, I am deeply indebted to my friends and family whose moral support kept my motivation high during this process.

Abstract

Public technology has been shown to have a strong dependence on physical touch, which increases the transmission of diseases. Gesture recognition helps to reduce this transmission, as the dependence on physical touch is removed. Furthermore, the use of visible light for gesture recognition would reduce the power consumption of public technology, as less power would have to be supplied for a light source. In this paper, a dataset for gesture recognition using ambient light is presented, alongside the design process and challenges faced. The gestures were collected using three photodiodes in order for a machine learning algorithm to identify the patterns made by the shadows cast when the gestures are performed. The dataset consists of 10 gestures performed by a total of 50 people 5 times on each hand. This was collected under 5 different light intensity ranges. This dataset was then passed through a machine learning model to be trained and tested, resulting in a 86.8% (3.s.f) validation accuracy. There are many factors related to the light source that caused the accuracy of the algorithm not to be as high as expected, however, the highest accuracy was found in environments with light intensities of 100-1000 lux; a well-lit indoor room.

1 Introduction

The COVID-19 pandemic has highlighted how often interactions with public technology require physical touch, which increases the transmission of diseases [1]. An example of this technology is automated teller machines (ATMs), which require users to touch the screen and buttons in order to store or withdraw money [2]. To prevent this spread, there has been an urgent need for contact-less public technology.

Hand gesture recognition removes the need for this physical touch, as people can simply perform a gesture for a system to understand the intended action. These gestures have to be simple enough for a user to perform them with ease, but also complex enough for the same actions that were done with physical touch to still be possible with gesture recognition. Simple public technology can make use of this, as they are designed to have simple and intuitive interfaces [3]. This can be seen in elevators, where a user can swipe to the floor they would like to go to and tap to select it as opposed to pushing the button with their finger.

Gesture recognition as a solution to the transmission of diseases is viable, however, there needs to be a medium for it to be detected. One common medium is the use of cameras to perform feature extraction [4]. Nevertheless, cameras require extra power to be run consistently. They also raise privacy concerns, as a data breach could lead to the camera feed being used maliciously.

The use of a camera is a form of modulated light, which is when the system provides the light source that will be used and selects its frequency or wavelength. Any use of modulated light will also face the issue of power consumption on top of requiring pre-deployment and modification to the existing environment in order to work effectively.

An alternative to modulated light is the use of unmodulated light (ambient light), which is light that is not changed by the system. Examples of these are indoor lamps, sunlight or candles. Ambient light is more energy efficient than modulated light and it reduces the privacy risks.

For the reasons presented above, this study is going to investigate the use of ambient light and embedded AI for gesture recognition in order to identify gestures in an energy efficient way. The system comprises of an Arduino 33BLE [5] and 3 OPT101 photodiodes [6] to detect the shadow patterns when gestures are performed.

1.1 Limitations of Existing Work

There are numerous existing papers related to gesture recognition using ambient light which have shown to be highly accurate, however, there are areas in which they could all be improved. The main issue is related to the number of sensors used, as systems such as GestureLite [7] and LightDigit [8] use a 3x3 grid of sensors, allowing them to recognize a range of gestures. In comparison, SolarGest [9] utilizes only one sensor, thus is only able to recognize very simple gestures. Furthermore, existing research does not focus on varying the environments and the candidates performing the gestures, causing them to not accurately reflect results in real-life scenarios.

1.2 Research Question

In order to create a machine learning (ML) algorithm that recognizes different gestures effectively, an extensive dataset is required to allow for proper training. Therefore, the research question discussed in this paper is the following:

"What kind of gesture dataset should be constructed for the purpose of training a machine learning model?"

Building the dataset is a necessity for this project, as there is no existing dataset for the scope of this research. A large majority of datasets for gesture recognition are collected using cameras, which cannot be applied to this project, as it utilizes photodiodes. Furthermore, the gestures used in other similar datasets also cannot be applied to this project due them using a different number of sensors, thus being able to recognize more complex gestures or oversimplified gestures. As a result, there are three main challenges incurred during the building of the dataset.

Selection of Hand Gestures

The first challenge is selecting the hand gestures that can be distinguished by an ML algorithm based on what the photodiodes see. As a result of there being no final ML algorithm during the data collection phase, in order to argue that the hand gestures can be distinguished by the algorithm, there has to be a clear visual distinction between the graphs of the gestures.

Diversification of Data

The diversification of data is an applicable challenge to any project requiring data collection for machine learning. Diverse data improves the algorithm's ability to adapt to different situations [10]. Diversifying the data involves data being collected in different scenarios, such as environments with varying light intensities and people with varying hand sizes. On top of this, it is also important to know when the dataset is diverse enough.

Controlling the Experiment

In order to have reliable results, certain elements of the experiment need to be controlled, such as the orientation of the system or the way in which certain gestures are performed. This has to be decided beforehand with evidence to argue why it is done that way.

1.3 Contributions

This paper makes multiple contributions to the research of gesture recognition using ambient light.

A large dataset is provided for recognizing gestures using 3 photodiodes which can be used and built upon by future work related to this research.

Knowledge about how the environments with varying light intensities affect the gesture detection is also provided. This allows conclusions to be drawn on the environment the system performs best in.

Initial insight into what gestures can be defined with three photodiodes is also a result of this project. This increases knowledge of whether the gestures collected are too complex or too simple and ways in which they can be extended for future work.

An evaluation of the intuitiveness of the gestures is presented based on the initial instincts of the candidates, which can increase understanding on the relation between the gestures and user experience.

2 Background

This section aims to give insight into the necessary background information required to understand the rest of this paper. It provides definitions for frequently used terms and explains how they are related to this project.

2.1 Gesture Recognition

Gesture recognition can be defined as "recognizing meaningful human movements from image sequences" [11]. The human movements focused on in this project are hand gestures. The gestures in this project are all dynamic, which means the hand has to be moving in order for the gesture to be understood.

2.2 Ambient Light

Ambient light can be defined as unmodulated light from the environment. Some examples of this are natural sunlight, indoor lamps and candles. The wavelengths and intensities of these light sources are not controlled by the system, which is why it is not as easy to work with as modulated light. However, it is more energy efficient, as power does not need to be supplied by the system for the light source. Working with ambient light requires an adaptable system to different environments, therefore, the environments are varied during the data collection process.

2.3 Photodiode

A photodiode is "a photoelectric semiconductor device for detecting and often measuring radiant energy" [12]. It returns a voltage value depending on the ambient light intensity. A higher intensity would result in a higher value and a lower intensity would result in a lower value. The photodiode used in this project is the OPT101 photodiode [6]. The photodiodes are used to recognize the shadow pattern of the hand performing the gesture.

3 System Overview

Section 3 provides an overview of the system and the main technologies used during this project followed by a more detailed overview of the sub-project.

3.1 Overview of entire project

The entire project can be split into different parts of the pipeline. These sections will briefly be discussed below in order to provide more context.

Hardware

The first part of the pipeline is the hardware, which consists of an Arduino 33BLE, three OPT101 photodiodes and 3 sets of multiple resistors with resistance values ranging from 1k Ohms to 2M Ohms [6]. The range of resistors allow the photodiodes to be calibrated based on the intensity of the ambient light. The system also includes three 10 micro Farad capacitors that act to smooth the values from the photodiodes. This setup can be seen in Figure 8 [13].

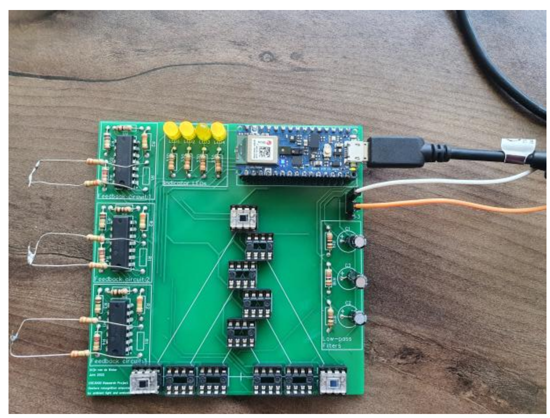

Figure 1: The hardware setup of the final system. This consists of three OPT101 photodiodes an Arduino 33BLE and three 10 micro Farad capacitors

Processing pipeline

An important part of the system is the edge detection, which detects when a live gesture starts and ends. This allows the noise around the gesture to be cut off, therefore only the important data is passed through to the ML model. This data is then processed through normalization in order to overcome the difference in durations of the same gesture and the variations of the depth of the troughs as a result of changing light intensities [14].

Data Collection

The data collection section of the entire project covers the selection of which gestures can be detected by the system and the actual collection of the data. Shadow analysis has to be performed in order to know which gestures are distinguishable and a detailed experiment has to be carried out to build the dataset. This is what this paper will focus on.

Machine Learning

Once the data is collected, it will have to be passed through to the ML algorithm, where it will be trained. From then on, the live data will be run on the model in order to classify it [15, 16].

3.2 Overview of sub project

Gesture Detection

The way in which the gesture recognition works is utilising the fact that photodiodes sense light intensity. As a result, when one moves their hand over the photodiodes, the shadow cast causes a reduction of values emitted by the photodiodes due to a lower light intensity being detected. The gestures are able to be distinguished based on the patterns of these values, for example, swiping from right to left would cause the values of the right-most photodiode to decrease first and the left-most photodiode to decrease last. This is evident in Figure 2.

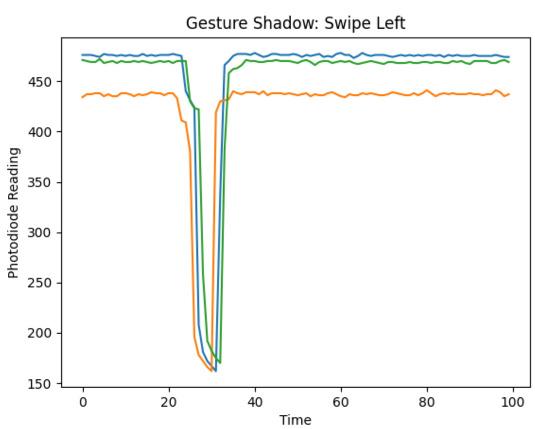

Figure 2: Graph displaying the photodiode readings of a swipe left gesture based on its shadow. The right-most photodiode (orange line) dips and rises, as the shadow of the hand passes over it first. Whereas, the left-most photodiode (green line) dips and rises last.

3.3 Research Question Sub-Questions

The research question was introduced in section 1.2 and broken down into general challenges that can be incurred from it. These challenges can further be broken down in order to have concrete sub-questions to answer. The sub-questions are presented below.

What gestures would be identified using three photodiodes

As most other research related to gesture recognition using ambient light made use of more than 3 photodiodes, a more diverse set of hand gestures is able to be detected. However, as this research only uses 3 photodiodes, it is important to initially figure out which hand gestures could be identified with this setup and see how they differ from the hand gestures selected in existing works.

What are the different habits people have when performing the selected gestures?

Diversity is a very important attribute of datasets, as it allows the machine learning algorithm to be able to adapt to different environments, such as different users having different habits when performing a gesture. In order to make the dataset more diverse, it is crucial to know what habits to expect and to make sure they are represented in the dataset.

What are the major differences between the habits from the photodiodes' point of view?

Once the habits are identified, they would have to be analyzed to see how they differ and whether the photodiodes detect a similar pattern or a completely different pattern.

How does the intensity of the ambient light affect the dataset?

When collecting the data for the dataset, there are certain conditions of the environment that need to be controlled in order to improve the quality of the dataset. Therefore, it is important to find out whether varying intensities of ambient light have a drastic effect on the patterns seen by the photodiodes in order to know how to control them during the data collection phase.

How can the proportions of the person's hand affect gesture recognition?

It is also essential to investigate how the proportions of a person's hand affect gesture recognition, as different hand sizes may produce different patterns. For example, the shape when a large hand that covers all the photodiodes during a tap may differ from when a small hand also performs a tap but is not able to cover all photodiodes.

How to decide when the dataset is diverse enough?

The question of whether a dataset is diverse enough has frequently been debated, as there is no one size fits all solution. For this situation, deciding whether the dataset is diverse enough would involve reflection of the data collection methods and training a machine learning algorithm on the data, examining how it handles unseen data.

4 Design

The design of the experiment is described and explained in section 4. Alongside this, the reasoning behind the way the implementing is carried out is also discussed.

4.1 Variables That Affect the System

There are many variables related to the system that can affect the final dataset negatively. It is important to highlight these variables and propose solutions to minimize their effects and subsequently increase the quality of the dataset.

Impact of physical attributes of people

In order to allow for the final model to adapt to different people using the system, the dataset it is trained with must contain people with varying physical attributes.

The main physical attribute of a person that is likely to affect the adaptability of the system is the proportions of the person's hand. A very small hand relative to the system may not cover all photodiodes when a gesture is performed, whereas a larger hand may cover all photodiodes for the same gesture. These inconsistencies would make it difficult to recognize gestures across people. Although this problem can be solved by adapting the placement and spacing of the photodiodes, it is also important to collect data for these varying scenarios.

The ages and genders of the candidates are also able to affect the results. Younger candidates may be more used to doing the gestures, as they are used to performing gestures on technology, whereas, older candidates may perform the gestures incorrectly [17]. The gender of the candidates also affect the hand sizes as males tend to have larger hands than females.

To account for these changes based on physical attributes, the candidates selected varied in hand sizes and ages and consisted of a balanced split between male and female.

Impact of ambient light

The ambient light affects the shape of the gestures based on its intensity and position. This can lead to inconsistent results across environments of varying light intensities.

Environment with high light intensities, such as outside in peak sunlight, causes the shape of the gestures to have steep drops. Contrastingly, lower intensity environments cause smoother, shallower curves [Image for proof]. This effect can be reduced by selecting the appropriate resistors for the environment.

The position of the light sources also heavily influences the gestures and can lead to cases where the gestures are not recognized. This is evident when the light source is at a low angle to the system. Despite the candidates hand being directly above the photodiodes during the gesture, the position of the light source causes the shadow to be cast further away from the photodiodes. This has a massive impact, as the detection of the gesture is heavily reliant on the shadow of the hand passing over the photodiodes. Positioning the light source directly above the system would mitigate this problem, however, as the ambient light is not controlled, this situation cannot be guaranteed. Additionally, in the situation when there are two light sources at different positions, two shadows may be cast onto the photodiodes from one hand, which also affects the shape of the gestures.

Impact of the way the gesture is performed

In order for the gestures to be correctly classified across different people, it is important to analyze the different ways in which people perform them. Good gestures should not have much variation across people, as it would be intuitive for everyone. On top of this, there may be differences between right-handed and left-handed people which will also cause the shapes of the gestures to not be similar. For the purpose of avoiding skewed data that only represents right-handed people, all candidates were asked to perform the gestures with both their right and left hand.

Impact of placement of photodiodes

The placement of the photodiodes also has an effect on the dataset, as it would need to allow for different gestures to be recognized. The spacing of the photodiodes is also important, as it should be close enough that the shadow of a hand can cover all 3 photodiodes, yet not too close where there is no lag between the dips of the photodiodes.

4.2 Selection of the hand gestures

The process of selecting the hand gestures requires careful consideration into what gestures are possible to distinguish between by three photodiodes. This task is not very simple due to the fact that many existing datasets use cameras for gesture recognition, which allows for the features of the hand to be identified. Drastically reducing the resolution to only 3 photodiodes greatly limits the number of gestures that can be identified.

The main factor that would allow to discern between gestures would be the time at which the values of a photodiode fall and rise relative to the other photodiodes. To cause this effect, the hand would have to be moving during gestures, therefore all gestures for this dataset have to be dynamic.

Swipe Gestures

As a result of this, the first four gestures are swipes in all four directions (left, right, up and down). These can be distinguished due to the lag between the troughs of the photodiode readings. For example, a swipe from right to left would cause the values of the right-most photodiode to decrease first and the left-most photodiode to decrease last. Not only are these gestures easy to distinguish between, they are also practical in everyday use and can be adapted to be used in different technologies.

Tap Gestures

The values of the photodiodes don't always have to have a lag, but can also fall and rise at the same time. This occurs when all photodiodes are covered and released simultaneously. The gestures that can cause this pattern are a tap and a double tap. These are also simple to visually distinguish between and can also be applied to different settings.

Rotation Gestures

With the hardware setup, rotation gestures should also be able to be distinguished. This pattern involves values of photodiodes falling and rising at different times and at least one of the photodiodes having two dips.

Zoom Gestures

The zoom gestures are also able to be distinguished, however the exact shape varies slightly. The ideal shape would be to

Light Intensity Range (lux)	Number of Candidates
0-100	10
100-300	10
300-1,000	10
1,000-10,000	10
10,000-100,000	10

Table 1: The light intensity ranges in which the data was collected and the number of candidates collected under each light intensity.

have one photodiode be dipped the whole time due to the bottom of the hand covering it while the other two photodiodes only dip briefly as the fingers pass over them.

4.3 Methodology

After considering the different factors that can affect the data collection process and selecting the gestures that can be collected, a detailed methodology can be constructed.

There are 50 candidates that data is collected from and each candidate performs each hand gesture for a total of 5 times on each hand. The number of candidates was chosen to have a diverse group of people and the candidates perform the gestures using both hands to account for both left-handed and right-handed people.

There were five environments chosen with ranging lux values as shown in Table 1. The first range was 0-100 lux, which covers dark indoor rooms, followed by 100-300 lux, which is a poorly lit indoor room. Normal indoor light settings were also covered with the range from 300-1000 lux and then 1000-10,000 lux accounts for bright, indoor areas next to windows. Finally, the range of 10,000-100,000 covers outdoor environments ranging from cloudy days to peak sunlight. There were 10 candidates for every lux range in order to have an even split between the settings.

This methodology assures diversity in the data, as when it is compared to other studies related to gesture recognition with ambient light, it uses more candidates and a larger variety of environments.

When collecting data from people, the candidates were first asked to perform the gesture naturally without instructions and if it strays far from what is intended then they will be told how to do it. This allows for more insight into how people will intuitively do the gestures.

All gestures have to be performed with the palm facing down and fingers close together. Certain gestures require more instructions, for example, the swipe up and swipe down gesture need the hand to be perpendicular. On top of that, the rotational gestures require one and a half rotations to be done. This was to allow the general patterns to be retained across different people performing the gestures. It would not work otherwise.

The resistors are also to be adjusted manually to be at a base value of 400-800 [13]. They were changed manually, as during the data collection stage, the final setup was not ready.

5 Implementation

This section aims to highlight the key aspects about the data collection process. One system was used during the entire data collection in order to allow for consistent results. The data collection mostly happened on the campus on the TU Delft in areas of varying light intensities. Figures 1-3 show three areas where a large portion of data collection occurred. The age, gender and hand sizes of the candidates were collected in order to keep track of the diversity of people. The candidates were also asked if they were left or right handed.

When collecting data, candidates were given minimal instructions at the start in order to see what the intuitive way to do the gestures were. During the rotational gestures, candidates were also asked to start at whatever position is most natural for them and aim to make one and a half circles.

The sampling frequency of the Arduino was changed around halfway through the data collection process from 20Hz to 100Hz. This was due to the latter working best for the edge detection [14]. It can be argued that the effect of this should not be too great, as the shape of the gestures were still the same between the two sampling frequencies. This is evident in figure 4, where the

Once the data for all 50 candidates was collected, it was cleaned by manually going through every instance to ensure that the correct number of gestures were done and none were missing. The final dataset contained 17 females and 33 with hand widths ranging from 8.5cm to 12cm and hand lengths ranging from 15.6cm to 21.4cm. The age range was also from 19 to 38 years and it had an overwhelming majority of right handed people with only 5 left handed candidates.

6 Evaluation of Results

In this section, the results of the final dataset will be presented, analysed and evaluated. The section will reflect on the overall performance when the dataset is used to train the algorithm and consider how different factors may have led to the results.

6.1 Overall Performance

The final dataset consists of 10 gestures performed by 50 candidates 5 times on each hand, therefore it contains 5000 instances of gestures. The results of this dataset being passed through the ML algorithm are as follows: When passing the dataset through the processing pipeline [14] prior to training the algorithm, the model had an 81.0% (3 s.f) validation accuracy, whereas when the raw data was used directly to train the algorithm, it achieved an 86.8% (3.s.f) validation accuracy [16].

Existing research showed to have accuracies of over 95% [8, 9, 18]. The accuracy for the raw data being lower than expected can partially attributed to the large buffers before and after the gestures. This was meant to be solved by the processing pipeline, which cuts off the part of the data that does not contain the gesture so only the important information is trained on, however, it does not perform well with the data collected in poorly lit areas, thus causing its accuracy to be lower than that of the raw data.

The lower than expected accuracy of the raw data is also a cause of the variations in which people perform the gestures. This highlights a very important trade-off between data that would produce high accuracy and data that would

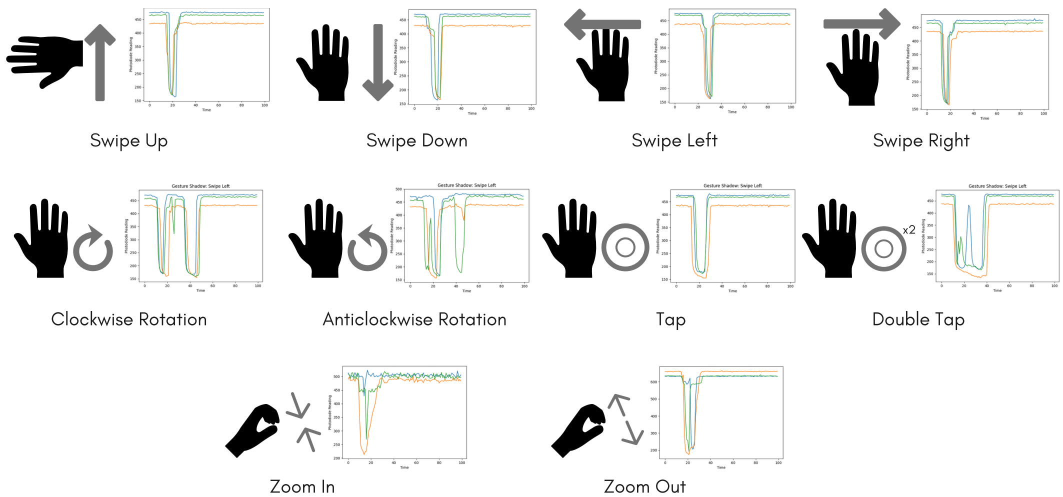

Figure 3: The gestures that the dataset consists of along side example graphs of what each gesture looks like (larger version in appendix). The gestures can be distinguished based on the lag between the dips in the graphs and the number of dips.

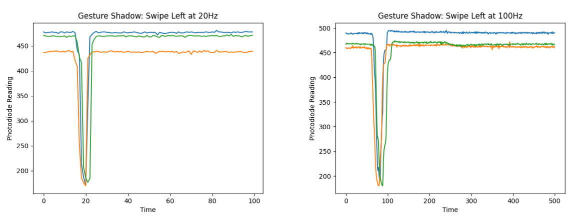

Figure 4: Comparison between the shape of the gesture recorded at 20Hz (left) and at 100Hz (right) to prove that the shape of the gesture remains the same across the two sampling frequencies.

reflect real life scenarios. Due to the varying environments and the varying ways people perform the gestures, the accuracy through the algorithm is reduced, as it is not able to generalize the gestures being performed under different conditions well enough. However, the goal was to build a diverse enough dataset that reflects real life data. Furthermore, the system is meant to adapt to different environments, which is why there were 5 different environment scenarios collected. If they were all collected in one environment, the accuracy would have been higher.

6.2 Ambient light

When analysing the dataset, some of the concerns raised in section 4 regarding how ambient light would affect the data were present.

When the light source was at an angle (mainly in the environment near a window), the shadow of the gesture was offset by a certain amount and sometimes caused the shadow to not cover the photodiodes despite the gesture being done above the photodiodes. This required the candidates to perform the gestures vertically closer to the system with a horizontal offset in order to allow the shadow to move over the photodiodes and avoid no gesture being recognized. This, however, did not affect the accuracy, as it can simply be solved by shifting the hand to counteract the offset caused by the angled light source.

When there were two light sources (commonly in the environment near a window with an indoor light source as well), the shape of the gestures became less similar to what was expected, as the two shadows cast over the photodiodes created many fluctuations. This is evident in figure 5, where a swipe left gesture is shown. It is clear that it looks very dissimilar to the swipe left gesture in figure 2/ This does affect the accuracy, as the shape of the gesture is changed drastically. Furthermore, this cannot be fixed with processing, as it is not possible to salvage the intended shape from the noise.

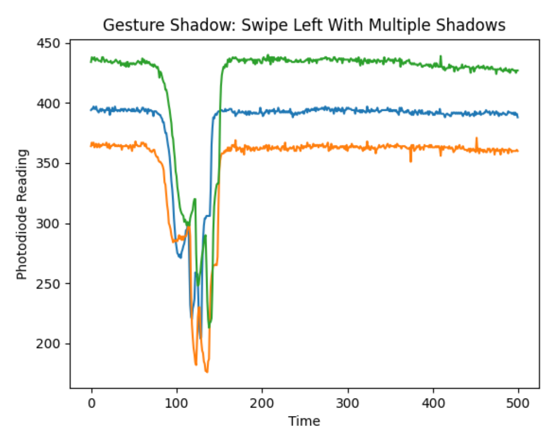

Figure 5: Example of the shape of a swipe left gesture when there are two light sources (sunlight and indoor light) causing two shadows of the hand when a gesture is performed.

Contradictory to the hypothesis that environments with high light intensities would reduce the accuracy due to gestures being too steep, this did not cause any drastic problems with the accuracy when passed through the processing pipeline due to the normalization. The main issue with outdoor environments was the fact that the light intensity can change drastically over a matter of seconds as a result of the clouds blocking the sunlight.

6.3 Performing Gestures

The way in which the candidates performed the gestures did vary, which may have led to a lower accuracy and caused certain gestures to be confused with others.

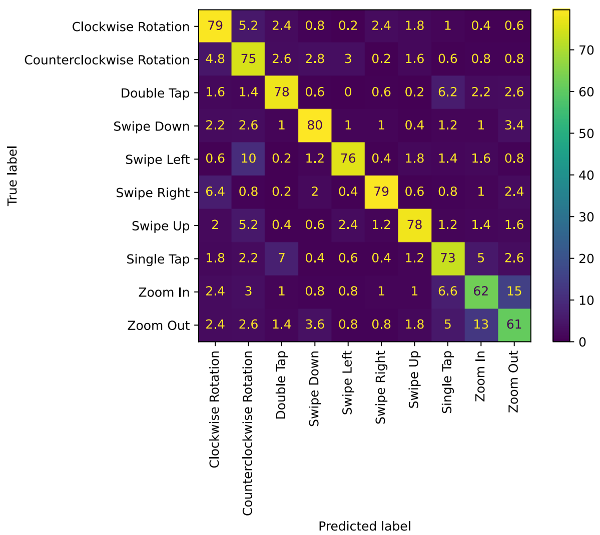

Figure 6: Confusion matrix when the raw dataset is run with a 10-fold cross-validation.

Swipe left and swipe right

The swipe left and swipe right gestures were the most consistent and intuitive across candidates. As these were the first gestures that candidates were asked to perform, some candidates performed them using one finger instead of their whole palm before being instructed on how to perform them properly. These were shown to be accurately distinguished from the other gestures in the confusion matrix in figure 6.

Swipe up and swipe down

When candidates were asked to perform gestures naturally prior to the data being collected, the nomenclature of the "swipe up" and "swipe down" gestures caused confusion, as candidates ended up lifting their hands vertically up as opposed to a forward movement that was intended. This can easily be solved by changing the name of the gestures to "swipe forward" and "swipe backward" respectively.

Regardless of the name change, the way in which these two gestures are meant to be performed are generally not intuitive. For example, for a swipe up, most candidates moved their hands forward fingers-first, thus ending the gesture with their forearm still over the photodiodes. This can be seen in figure 7. This would not be identifiable, as the shadow never finished passing over the photodiodes, which is why the gesture needs to be done with the side of the palm crossing first and ending with the other side of the palm. This motion tends to cause discomfort for candidates, however, there is no other way to perform these gestures for them to be identifiable.

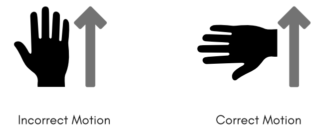

Figure 7: A comparison of the incorrect (left) and correct (right) motion for swiping up. A vast majority of the candidates had an initial instinct to perform the gesture the incorrect way. The hand needs to be horizontal so the entire shadow passes over all of the photodiodes. The incorrect motion causes the forearm of the hand to still cover the photodiodes once the gesture is done, thus making the gesture unrecognizable.

Clockwise and anticlockwise rotations

The clockwise and anticlockwise rotations were intuitive enough for most candidates. As the candidates had freedom as to where to begin the gesture, it did cause some variance within the final dataset, however, the general shape was still very similar and the confusion matrix in figure 6 still showed the algorithm to consistently identify it.

Zoom in and zoom out

The zoom in and zoom out gestures caused the most confusion and variance between candidates. The initial way candidates expected to perform the gesture was over the photodiodes, whereas the hand is meant to be rotated so as to not cover all photodiodes all the time. This gesture had the most confusion as can be seen in figure 6.

6.4 Adaptability to new environments

The reason for changing the light intensities and the hand proportions of the candidates was to allow the algorithm trained on the dataset to adapt to a range of environments. This ultimately reduced the overall accuracy, however, it gave more insight into the tolerance of the system under real-world scenarios. The changing of the hand sizes did not affect the dataset in the end, as the placement of the photodiodes were close enough for all hands to cover them at once. It is important to note that the hand widths ranged from 8.5 to 12cm, therefore the effects of younger users using the system is not known.

7 Related Work

There have been other works which also use ambient light for gesture recognition that differ from this project in different ways. These projects can be compared in terms of scope, number of gestures and size of dataset.

7.1 LightDigit

LightDigit is a system that recognizes digits from 0-9 written in the air. It also uses ambient light by having a 3x3 grid of photodiodes to recognize the gestures. The scope is very similar to this project and mainly differs due to the number of photodiodes and the gestures being detected.

The dataset consisting of the same number of gestures but with 20880 instances, which is much larger than the dataset presented in this project. However, the data was collected by a single person and the different environments it was performed in (if any) were not recorded. As a result, the research mainly focused on consistency of the data as opposed to this research, which aims to present data that is representative of real-life scenarios.

7.2 SolarGest

SolarGest also implemented gesture recognition using ambient light, however, this was done by utilizing a single solar cell. Having only one sensor reduces the number of possible gestures that can be recognized, therefore the dataset created by SolarGest only consists of 6 gestures.

The dataset they created contains 6960 instances of gestures, which is also slightly larger than the dataset for this research. Their research also varied the environment by collecting data under 5 different light intensities, as well as collecting data with and without human interference in the background, such as people walking around. The former is similar to this research, but the latter differs, as this research assumed minimal interference during the data collection. 3 subjects were used to collect the data, therefore the main focus of this project was varying the environment and not the people.

7.3 GestureLite

GestureLite is another gesture recognition system using ambient light which utilizes a 3x3 grid of photodiodes to detect the gestures. The scope mainly differs to that of this project by the number of photodiodes, which makes the system for this project more energy and cost efficient.

The dataset has 3300 instances of gestures, which is lower than the number used in the dataset of this project. The environment was varied by collecting data in "light" and "dark" environments and in two locations (a dorm room and a classroom). Only one candidate performed all of the gestures, which reduces the adaptability of the dataset to other people, however, the candidate "did not perform more than 10 samples in one sitting", which may have slightly improved this issue. The dataset assumes all future users will perform the gestures the same way as how the candidate performed it, which may not be the case. The gestures contained in the dataset are very similar to the gestures in the dataset of this project, as it has all four swipe gestures and the two rotational gestures. It also has "flick open", "flick open twice", "rise" and "fall".

8 Responsible Research

It is important to analyse the ethics of this research and to prove reproducibility of the results.

The experiments carried out were done with candidates who were talked through the entire process and gave consent. With regards to the safety of the experiment, it was minimal risk, as the candidates never touched the system, and even if they did, the amount of voltage required by the system does not affect humans. The ages, hand sizes gender and whether the candidates were left handed or right handed were collected in order to have general statistics about the dataset and to prove diversity. This data is not published and cannot be traced back to the candidates, therefore this is also minimal risk.

The results are reproducible, as the detailed explanation of the system used during data collection and the steps taken are outlined in section 4.3. The dataset is also made publicly available, along with all the code used to collect it [19].

9 Conclusion

In this paper, a dataset for gesture recognition using ambient light was designed and built and the results of passing the dataset through a machine learning algorithm were analysed and evaluated. The overall accuracy of the algorithm when trained on the dataset was 86.8%. The evaluations showed that environments with light intensities ranging from 100-1000 lux performed better than those less than 100 lux. The different factors that could have led to the lower accuracy were described, such as having multiple light sources being present that cause two shadows for one gesture to be cast and also the different ways the candidates performed the gestures.

Throughout the paper, the sub-questions of the research question have been answered. The gestures that could be identified using three photodiodes were shown and the habits and initial instincts people have when performing those gestures were explained, such as for the swipe up gesture. The intensity of the ambient light was shown to have an effect on the dataset, as certain light intensities performed worse than others, however, this effect was meant to be minimized by the processing pipeline.

Overall, the dataset was set up to provide a diverse set of data that reflects the real-world in which a machine learning algorithm can be trained on. Extending the data to work with the processing pipeline could result in a more accurate system that can adapt to the different environments better. However, the results thus far have shown the limitations of using ambient light for gesture recognition and has given insight into the possible improvements to public technology to mitigate the transmission of diseases through touch.

10 Future Work

There can be many conclusions drawn from this research, which can then be built upon by future work. The different ways in which the dataset can be extended and the whole system can be built upon are as follows:

Include data with noise

One way to extend the dataset would be to have data with noise in the background, such as people walking around and slightly obstructing the light source. This would test the tolerance of the system against an even more realistic scenario, as now the environment is expected to be unchanged during a gesture.

Select environment

The system was built with no single preconceived environment in mind in which it will be deployed, therefore multiple environments were tested. If one environment was selected, it would allow for more data to be collected for that environment, potentially providing a higher accuracy.

Orientation of system

The system is also assumed to be laid horizontally on a flat surface. The effects of changing this orientation to be more vertical would produce different results, as it may cause more issues with the obstruction of the light source.

Fix light offset

As mentioned in section 4, when the light source is at an angle, it causes the shadow of the gesture to be shifted further away from the system. Research into a way to mitigate this effect would be beneficial in order to avoid shadows not passing over the photodiodes.

References

[1] M. A. Marinella, "Covid-19 pandemic and the stethoscope: Do not forget to sanitize," *Heart & Lung Journal*, vol. 49, no. 4, p. 350, 2020.

[2] A. A. Hossein M. Elbadawy, Amin Khattab, "The detection of sars-cov-2 in outpatient clinics and public facilities during the covid-19 pandemic," *Journal of Medical Virology*, vol. 93, no. 5, pp. 2955–2961, 2021.

[3] R. Rousi, "The experience of no experience elevator ux and the role of unconscious experience," *The Computer Journal*, no. 4, pp. 289–292, 2013.

[4] F. A. et al., "A proposed touchless hand-gesture and voice based model eliminating covid-19 infection risk while using atm machines," *International Journal of Computer Science and Network Security*, vol. 22, no. 3, pp. 419–494, 2022.

[5] Arduino, "Arduino nano 33 ble sense - product reference manual."

[6] T. Instruments, "Opt101 monolithic photodiode and single-supply transimpedance amplifier."

[7] M.-D. Kaholokula, "Reusing ambient light to recognize hand gestures,"

[8] E. Group, *Technical Report: LightDigit Dataset*, 2021.

[9] D. Ma, G. Lan, M. Hassan, W. Hu, B. M. Upama, A. Uddin, and M. Youssef, "Solargest: Ubiquitous and battery-free gesture recognition using solar cells," *Proceedings of the 25th ACM Annual International Conference on Mobile Computing and Networking, Los Cabos, Mexico*, vol. 21-25, pp. 1–15, 2019.

[10] Z. GONG, *Diversity in Machine Learning*, 2019.

[11] V. A. Matthew Turk, "Gesture recognition," 2020.

[12] M.-W. Dictionary, "Photodiode," 1999.

[13] S. van de Water, "Designing an adaptable and low-cost system for gesture recognition using visible light," 2022.

[14] D. Barantiev, "Designing a software receiver for gesture recognition with ambient light," 2022.

[15] M. Lipski, "Gesture recognition on arduino using recurrent neural networks and ambient light," 2022.

[16] W. Narchi, "Recognising gestures using ambient light and convolutional neural networks," 2022.

[17] A. T. Cabreira, "Lan investigation of mid-air gesture interaction for older adults," *PhD thesis, University of Reading*, 2019.

[18] R. H. VENKATNARAYAN, *Gesture Recognition Using Ambient Light*, 2018.

[19] "Cse3100-gesture-recognition repository." Available at: https://github.com/StijnW66/CSE3000-Gesture-Recognition.

A Figure 3 Enlarged

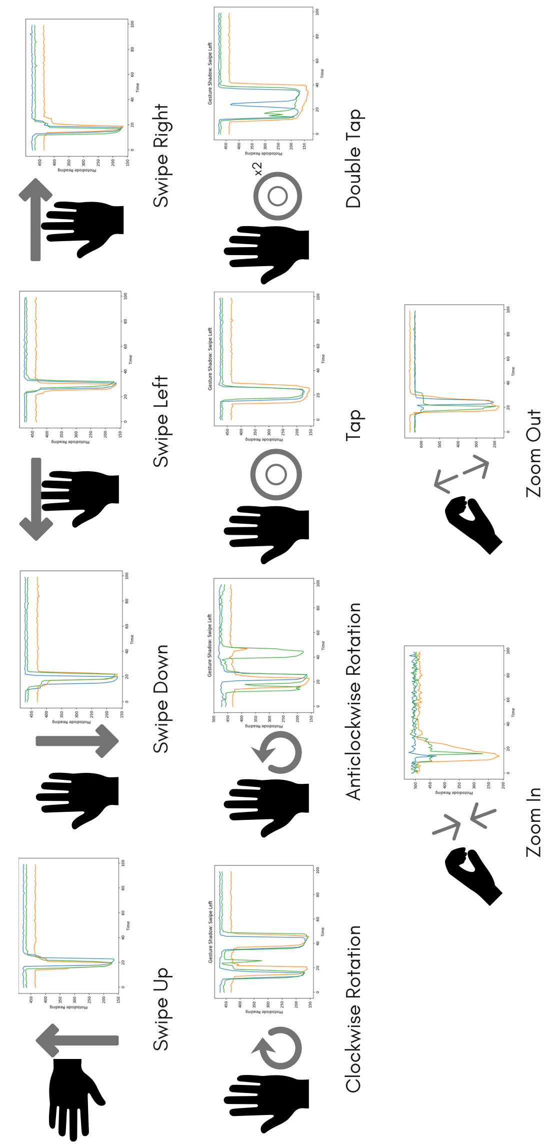

Figure 8: The gestures that the dataset consists of along side example graphs of what each gesture looks like. The gestures can be distinguished based on the lag between the dips in the graphs and the number of dips.