**Gesture Recognition for Enabling Control of Electrical Devices**

by David T. Tung

B.S. in Mechanical Engineering, National Sun Yat-Sen University

Master of Computer Science, University of Southern California

A Praxis submitted to

The Faculty of

The School of Engineering and Applied Science

of The George Washington University

in partial fulfillment of the requirements

for the degree of Doctor of Engineering

August xx, 2025

Praxis directed by

Shahryar Sarkani  
 Adjunct Professor of Engineering Management and Systems Engineering

John Fossaceca  
 Professorial Lecturer of Engineering Management and Systems Engineering

The School of Engineering and Applied Science of The George Washington University certifies that David T. Tung has passed the Final Examination for the degree of Doctor of Engineering as of August xx, 2025. This is the final and approved form of the Praxis.

**Gesture Recognition for Enabling Control of Electrical Devices**

David T. Tung

Praxis Research Committee:

Shahryar Sarkani, Adjunct Professor of Engineering Management and Systems Engineering, Praxis Co-Director

John Fossaceca, Professorial Lecturer of Engineering Management and Systems Engineering, Praxis Co-Director

Amir Etemadi, Associate Professor of Engineering and Applied Science, Committee Member

© Copyright 2025 by David T. Tung

All rights reserved

# Dedication

I would like to dedicate this research to my family [todo]

# Acknowledgements

I wish to thank Dr. Shahryar Sarkani and Dr. John Fossaceca for their guidance and support throughout the studying, research, and writing processes. [todo]

# Abstract of Praxis

**Gesture Recognition for Enabling Control of Electrical Devices**

Individuals with mobility impairments, particularly wheelchair users, often face challenges in accessing and controlling electrical devices. A gesture-based control system can greatly enhance accessibility for this group. This research praxis aims to develop and evaluate a user-friendly, cost-effective machine-learning approach that enables users to control devices through pointing gestures. Specifically, it focuses on extending DeePoint, a 3D pointing direction prediction model (Nakamura et al., 2023), into a unified machine learning system capable of identifying the electrical devices a user is pointing at. The goal is to lay the groundwork for future applications that empower individuals with mobility impairments to intuitively and efficiently control household devices using simple, natural gestures.

**Table of Contents**

[Dedication iv](#_Toc179138418)

[Acknowledgments v](#_Toc179138419)

[Abstract of Praxis vi](#_Toc179138420)

[List of Figures ix](#_Toc179138421)

[List of Tables x](#_Toc179138422)

[List of Symbols xi](#_Toc179138423)

[List of Acronyms xii](#_Toc179138424)

[Chapter 1—Introduction 1](#_Toc179138425)

[1.1 Background 1](#_Toc179138426)

[1.2 Research Motivation 3](#_Toc179138427)

[1.4 Problem Statement 4](#_Toc179138428)

[1.4 Thesis Statement 4](#_Toc179138429)

[1.5 Research Objectives 4](#_Toc179138430)

[1.6 Research Questions and Hypotheses 5](#_Toc179138431)

[1.7 Scope of Research 6](#_Toc179138432)

[1.8 Research Limitations 7](#_Toc179138433)

[1.9 Organization of Praxis 7](#_Toc179138434)

[Chapter 2—Literature Review 9](#_Toc179138435)

[2.1 Introduction 9](#_Toc179138436)

[2.2 Another Section 9](#_Toc179138437)

[2.3 Another Section 9](#_Toc179138438)

[Chapter 3—Methodology 10](#_Toc179138439)

[3.1 Introduction 10](#_Toc179138440)

[3.2 Another Section 10](#_Toc179138441)

[Chapter 4—Results 11](#_Toc179138442)

[4.1 Introduction 11](#_Toc179138443)

[4.2 Another Section 11](#_Toc179138444)

[Chapter 5—Discussion and Conclusions 12](#_Toc179138445)

[5.1 Discussion 12](#_Toc179138446)

[5.2 Conclusions 12](#_Toc179138447)

[5.3 Contributions to Body of Knowledge 12](#_Toc179138448)

[5.4 Recommendations for Future Research 12](#_Toc179138449)

[References 13](#_Toc179138450)

[Appendix A 17](#_Toc179138451)

# List of Figures[todo]

[Figure 1. The architecture of the device classification system. 3](#_Toc179180953)

[Figure 4-1. XYZ. 8](#_Toc179180954)

[Figure A-1. Histogram of XYZ. 14](#_Toc179180955)

# List of Tables[todo]

[Table 4-1. Pearson Correlations Between W and T 9](#_Toc519272309)

[Table A-1. Parametric Correlations of X and Y 14](#_Toc519272310)

# List of Symbols

State of the system

Output of the system

Noise

# List of Acronyms

ADA Americans with Disabilities Act

YOLO You only look once is a state-of-the-art, real-time object detection system

CNN Convolutional Neural Network

DL Deep Learning

ML Machine Learning

U.S. United States

# Chapter 1—Introduction

## 1.1 Background

As the number of older adult households continues to grow in the decades ahead, more homeowners are expected to undertake expensive accessibility renovations (JCHS, Harvard, 2023). One of these projects' most critical and costly aspects is improving access to electrical devices, particularly for individuals with mobility impairments who rely on wheelchairs for movement. Accessible housing design features can significantly enhance the independence of people with mobility disabilities within their homes (JCHS, Harvard, 2023). Despite advances in assistive devices, laws, and technology aimed at improving safety and aiding daily activities, people with mobility disabilities continue to face barriers in accessing electrical devices, reducing their capacity to live independently.

Over the years, various solutions have been developed to assist individuals with mobility impairments control household devices, including voice-activated systems and augmented reality or mixed reality smart glasses (Zhou, K. *et al*., 2023). Despite their potential, these technologies are not yet ready to be widespread within this group due to several challenges. Voice-activated systems, for instance, often underperform in noisy environments, making them unreliable in busy or public settings. Additionally, they are unsuitable for environments where quiet is essential, such as a baby’s room or a shared living space. While smart glasses provide a hands-free solution, they come with drawbacks, such as high costs and discomfort when worn for extended periods, which limits their practicality for long-term, daily use.

Moreover, many existing gesture-based control systems rely on specialized hardware, like RGB-D sensors, which can be effective but are often expensive, bulky, and cumbersome to use and maintain. These limitations prevent widespread adoption, particularly for individuals seeking affordable and convenient solutions to improve their independence. As a result, there is a growing need for more user-friendly, intuitive, and cost-effective approaches to overcome these barriers and provide greater accessibility in daily device control (Chang, V., *et al*., 2022).

Our solution uses finger-pointing gestures, a natural and efficient way to indicate objects or devices of interest. Implementing an intuitive system that allows device control through pointing gestures could significantly benefit individuals with mobility impairments. For example, a person in a wheelchair could navigate a room and point at devices like lights or appliances to control them, with the scene being captured by cameras mounted in the corners, which may already be available or simple to install.

In this work, illustrated in Fig. 1, We propose a multi-stage machine learning system that extends the 3D pointing direction prediction model, DeePoint (Nakamura et al., 2023). This system incorporates a device classification transformer, which utilizes the predicted pointing direction and images from DeePoint's outputs to form a unified device classification system capable of identifying which devices the user is pointing at. This classification system can process video input that monitors the wheelchair user, detecting their intent to control devices by pointing at the device and generating a command for the device to act[[1]](#footnote-2).

A diagram of a transformer

Description automatically generated

##### Figure 1. The architecture of the device classification system.

With the proposed classification system, our contribution establishes a foundation for future touchless device control applications extending beyond household devices to include elements like elevator buttons. This system aims to empower individuals with mobility impairments to intuitively and efficiently control devices through natural pointing gestures

## 1.2 Research Motivation

The motivation behind this research arises from the persistent challenges faced by individuals with mobility impairments, particularly wheelchair users, in controlling household devices. While offering a degree of autonomy, existing solutions like voice-activated systems and smart glasses have notable limitations, such as reduced effectiveness in noisy environments, physical discomfort, and high costs. Additionally, hardware-dependent systems, such as those utilizing RGB-D sensors, are often expensive and cumbersome, further restricting accessibility.

More natural, intuitive, and affordable solutions that enable individuals with mobility impairments to interact seamlessly with devices are needed. Gesture recognition, especially through pointing direction, offers a promising alternative. However, research on the visual interpretation of 3D pointing gestures is still limited (Nakamura et al., 2023). To bridge this gap, this study seeks to develop a machine-learning model that leverages standard RGB cameras for 3D pointing recognition, to provide a practical and cost-effective solution for device control.

## 1.4 Problem Statement

## *Individuals using wheelchairs often face challenges accessing electrical devices, so approximately 35% of U.S. housing units may need to be modified to meet the accessibility requirements of the devices (U.S. Department of Housing and Urban Development, n.d., 2015).*

## Even in places that comply with the Americans with Disabilities Act (ADA), individuals using wheelchairs are often required to stretch to reach electrical switches.

## Furthermore, Assistive devices, such as canes, robotic arms, intelligent eyewear, mobile phones, or AI-enabled wearables, are expensive.

## 1.4 Thesis Statement

*A two-stage classification system is needed to identify the electrical devices a wheelchair user points at, enabling touchless device control and enhancing accessibility.*

## The primary research product is a device classification system developed in Python, designed to enhance device accessibility for wheelchair users. This technology enables integrators and developers to improve accessibility. The research presents a new contribution by developing a classification system that identifies electrical devices pointed at by a wheelchair user, leveraging DeePoint and object detection as its core machine learning technologies. The methodology incorporates machine learning, transformers, deep learning, computer vision, and object detection. Input data will consist of videos of a seated individual pointing at electrical devices, while the system’s output will classify the device and provide a probability score.

## 1.5 Research Objectives

The primary objective of this research is to develop a machine-learning model that can classify devices a wheelchair user is pointing at. The key research focuses on exploring and evaluating the classification system’s accuracy and performance across various network architectures and scenarios. Detailed objectives are as follows:

Evaluate the feasibility of a machine learning device classification system that integrates pointing direction prediction with object detection capabilities. Assess the system’s performance using test data to explore its effectiveness and accuracy.

Investigate the impact of using different pointing direction prediction models as components within the proposed classification system, comparing their accuracy and efficiency, specifically DeePoint vs. OpenPose.

Analyze the impact of using different object detection models as components within the proposed pointing device classification system, comparing their accuracy and efficiency, specifically YOLO vs. Fast R-CNN.

## 1.6 Research Questions and Hypotheses

This study aims to clarify and explain the following three research questions (RQs):

**RQ1:** Does tracking gaze direction enhance the accuracy of pointing direction prediction in the first stages of the two-stage classification system?

**RQ2:** Can a two-stage classification system be developed to identify the electrical devices a wheelchair user points at, enabling touchless device control and improving accessibility?

**RQ3:** Which ML model works best for the second stage of the two-stage classification system for identification of electrical devices, as pointed by wheelchair user?

**H1:** Tracking gaze direction can improve the accuracy of pointing direction prediction by approximately 5% in the first stages of a two-stage classification system.

**H2:** The proposed two-stage classification system can reach 70% accuracy in identifying the electrical devices a wheelchair user points at.

**H3:** In the second stage, the proposed device classification system incorporating the YOLO is expected to outperform the model using the Fast R-CNN as an object detection component.

## 1.7 Scope of Research

The scope of this praxis is to evaluate the feasibility of developing a device classification system capable of interpreting gesture-based interactions to improve device accessibility for wheelchair users. The proposed solution extends a pointing direction prediction model and integrates a device classification transformer. Furthermore, the actual electrical device control part is not in the scope.

## 1.8 Research Limitations

The following factors limit this research:

While incorporating a confirmation step for pointing gestures would greatly improve reliability and user experience in practical applications, it falls outside the scope of this study. Instead, the focus is on developing a device classification model.

This research intentionally limits the device categories to three common household items: a laptop, a Fan, and a Lamp—to streamline model development and evaluation. Future work could expand this range to provide a more comprehensive solution for individuals with mobility impairments.

Additionally, the experiments will be conducted exclusively in indoor environments, and the test data is derived from videos of seated users rather than actual wheelchair users.

## 1.9 Organization of Praxis

This Praxis consists of five chapters, as follows:

Chapter 1 begins with the background and the research motivation, then continues with the research objectives, questions, and hypotheses. It ends with the scope of the research and the limitations of this research.

Chapter 2 presents a review of relevant literature, beginning with the DeepPoint paper (Nakamura *et al*., 2023), which serves as the foundation for this praxis. Additional literature provides context on accessibility challenges faced by wheelchair users (JCHS, Harvard, 2023; U.S. Department of Housing and Urban Development, n.d., 2015). The WorldPoint paper from CMU offers technical insights into the implementation of the ray-casting algorithm for object intersection (Kim D. et al., 2023). Furthermore, the MultiNet framework demonstrates how multiple models can be effectively combined into one (Teichmann M. et al., 2018). The end-to-end human-object interaction detection paper presents a solution based on the HOI pattern. (Zou, C. *et al*., 2018).

Chapter 3 presents the three research questions used in this praxis and testing of the hypotheses.

Chapter 4 covers the results and analyses of the statistical methods presented in Chapter 3.

Chapter 5 closes the praxis with a discussion of results and a conclusion. It also includes discussions of the contributions to the body of knowledge and recommendations for future research in the area.

# Chapter 2—Literature Review

## 2.1 Introduction

## 2.2 Another Section

## 2.3 Another Section

**2.4 Summary and Conclusion**

# Chapter 3—Methodology

## 3.1 Introduction

## 3.2 Another Section

# Chapter 4—Results

## 4.1 Introduction

## 4.2 Another Section

# Chapter 5—Discussion and Conclusions

## 5.1 Discussion

## 5.2 Conclusions

## 5.3 Contributions to Body of Knowledge

1. [todo]

## 5.4 Recommendations for Future Research

# References

U.S. Department of Housing and Urban Development. (n.d.). Accessibility of America’s housing stock: Analysis of the 2011 American Housing Survey. Office of Policy Development and Research.

Joint Center for Housing Studies of Harvard University. (2023). Improving America’s housing 2023 report. Harvard University.

Nakamura, S., Kawanishi, Y., Nobuhara, S., & Nishino, K. (2023). DeePoint: Visual Pointing Recognition and Direction Estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 20577-20587).

Cheng, T., Song, L., Ge, Y., Liu, W., Wang, X., & Shan, Y. (2024). Yolo-world: Real-time open-vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16901-16911)

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440-1448).

Constantin, S., Eyiokur, F. I., Yaman, D., Bärmann, L., & Waibel, A. (2023). Multimodal Error Correction with Natural Language and Pointing Gestures. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1976-1986).

Zou, C., Wang, B., Hu, Y., Liu, J., Wu, Q., Zhao, Y., ... & Sun, J. (2021). End-to-end human object interaction detection with hoi transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11825-11834).

Kim, D., Mollyn, V., & Harrison, C. (**2023**). **WorldPoint**: Finger Pointing as a Rapid and Natural Trigger for In-the-Wild Mobile Interactions. *Proceedings of the ACM on Human-Computer Interaction*, *7*(ISS), 357-375.

Zhou, Kanglei & Chen, Chen & Ma, Yue & Leng, Zhiying & Shum, Hubert & Li, Fred & Liang, Xiaohui. (2023). A Mixed Reality Training System for Hand-Object Interaction in Simulated Microgravity Environments. 167-176. 10.1109/ISMAR59233.2023.00031.

Xu, J., Wang, H., Zhang, J., & Cai, L. (2022). Robust hand gesture recognition based on RGB-D Data for natural human–computer interaction. *IEEE Access*, *10*, 54549-54562.

Chang, V., Eniola, R. O., Golightly, L., & Xu, Q. A. (2023). An Exploration into Human–Computer Interaction: Hand Gesture Recognition Management in a Challenging Environment. *SN Computer Science*, *4*(5), 441.

Pelgrim, M. H., He, I. X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024). Find it like a dog: Using Gesture to Improve Object Search. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).

Jirak, D., Biertimpel, D., Kerzel, M., & Wermter, S. (2021). Solving visual object ambiguities when pointing: an unsupervised learning approach. *Neural Computing and Applications*, *33*, 2297-2319.

Etoh, M., Tomono, A., & Kobayashi, Y. (1989). Direct Finger Pointing as a Man-machine Interface. *IFAC Proceedings Volumes*, *22*(12), 125-130.

Lee, K. (2024). *Point it Out: Using Gesture to Improve Object Search* (Doctoral dissertation, Brown University Providence, Rhode Island).

Wong, N., & Gutwin, C. (2010, April). Where are you pointing? The accuracy of deictic pointing in CVEs. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1029-1038).

Rümelin, S., Marouane, C., & Butz, A. (2013, October). Free-hand pointing for identification and interaction with distant objects. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 40-47).

Aftab, A. R., von der Beeck, M., & Feld, M. (2020, October). You have a point there: object selection inside an automobile using gaze, head pose and finger pointing. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 595-603).

Harika, M. (2016). *Finger-pointing gesture analysis for slide presentation* (Doctoral dissertation, 부경대학교 대학원).

Erden, F., & Cetin, A. E. (2014). Hand gesture based remote control system using infrared sensors and a camera. *IEEE Transactions on Consumer Electronics*, *60*(4), 675-680.

Erden, F., Velipasalar, S., Alkar, A. Z., & Cetin, A. E. (2016). Sensors in assisted living: A survey of signal and image processing methods. *IEEE Signal Processing Magazine*, *33*(2), 36-44.

Sudhakar, S., Liu, R., Van Hoorick, B., Vondrick, C., & Zemel, R. (2024). Controlling the World by Sleight of Hand. *arXiv preprint arXiv:2408.07147*.

# Appendix A

#### Table A-1. Parametric Correlations of X and Y

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| Pearson Correlation | 1 | .627\*\* |
| Sig. (2-tailed) |  | .000 |
| Sum of Squares and Cross-products | 960.119 | 607.382 |
| Covariance | .559 | .354 |
| N | 1719 | 1719 |
| Pearson Correlation | .627\*\* | 1 |



##### Figure A-1. Histogram of XYZ.

1. The actual device control research and implementation are not in the scope of this praxis. [↑](#footnote-ref-2)