

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Gesture and Voice Control of Internet of Things

A thesis presented in partial fulfilment of the requirements for the

degree of

Master of Engineering

in

Electronics and Computer Engineering

at Massey University, Auckland,

New Zealand.

Xiao Han

October 2015

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Mohammad Rashid for his guidance throughout my master study. His enthusiasm, encouragement, and faith in me have been extremely helpful.

Next, I would like to thank Dr. Liqiong Tang who is the one inspired me and helped me find the direction of study. Her support and faith on me have been very meaningful and made me start to believe myself too.

I would like to thank all my friends, my sister and my parents for their companionship and support.

ABSTRACT

Nowadays, people's life has been remarkably changed with various intelligent devices which can provide more and more convenient communication with people and with each other. Gesture and voice control are becoming more and more important and widely used. People feel the control system humanized and individualised using biological control.

In this thesis, an approach of combined voice and gesture control of Internet of Things is proposed. A prototype is built to show the accuracy and practicality of the system. A Cortex-A8 processor (S5PV210) is used and the embedded Linux version 3.0.8 has been cross-compiled. Qt 4.8.5 has been ported as a UI (User Interface) framework and OpenCV 2.4.5 employed as vision processing library. Two ZigBee modules are used to provide wireless communication for device control.

The system is divided into control station and appliance station. The control station includes development board, USB camera, voice recognition module, LCD screen and ZigBee module. This station is responsible for receiving input signal (from camera or microphone), analyzing the signal and sending control signal to appliance station. The appliance station consists of relay, ZigBee module and appliances. The ZigBee module in the appliance station is to receive control signal and send digital signal to connected relay. The appliance station is a modular unit that can be expanded for multiple appliances.

The system can detect and keep tracking user's hand. After recognizing user's gesture, it can control appliances based on certain gestures. Voice control is included as an additional control approach and voice commands can be adjusted for different devices.

TABLE OF CONTENTS

Acknowledgements	I
Abstract	III
List of Abbreviations	IX
List of Tables	XI
List of Figures	XII
CHAPTER 1 Introduction	- 1 -
1.1 Background	- 1 -
1.1.1 The Rise of Internet of Things.....	- 1 -
1.1.2 Gesture and Voice Control on IoT.....	- 4 -
1.2 Objectives.....	- 5 -
1.3 Organization of the thesis.....	- 6 -
CHAPTER 2 An Overview of the Internet of Things	- 7 -
2.1 Introduction	- 7 -
2.2 The History of IoT	- 7 -
2.3 The Landscape of IoT	- 9 -
2.4 IoT Components	- 11 -
2.5 Technologies in IoT	- 12 -
2.5.1 Voice Control in IoT	- 12 -
2.5.2 Gesture Control in IoT	- 13 -
CHAPTER 3 Literature Review and Related Works	- 15 -
3.1 Hand Gesture Recognition	- 15 -
3.1.1 Non-Vision Based Approaches.....	- 15 -
3.1.2 Vision Based Approaches.....	- 24 -

3.2	Voice Control	35 -
3.2.1	Classification of Voice Recognition	35 -
3.2.2	Process of Speech Recognition.....	37 -
3.2.3	Algorithm	38 -
3.3	Related Work.....	40 -
3.3.1	Natural User Interface Using Color markers.....	40 -
3.3.2	Cursor Control Using Haar Classifier	42 -
3.3.3	The Architecture Using Both Statistical and Syntactic Analysis	45 -
3.3.4	Hand Motion Recognition Using Kinect.....	46 -
3.3.5	Gesture Control of Smart Home	48 -
3.3.6	Summary of Related Work.....	49 -
CHAPTER 4 System Design and Implementation.....		51 -
4.1	Gesture Control and Voice Control Algorithms.....	53 -
4.1.1	Gesture Control Algorithms	53 -
4.1.2	Voice Control Algorithm.....	61 -
4.2	The Embedded System Platform	63 -
4.2.1	Profile of Embedded System	63 -
4.2.2	The Development Board System	67 -
4.2.3	Cross Compilation Environment	68 -
4.2.4	Summary to the Embedded System Platform.....	78 -
4.3	Hardware Design	79 -
4.3.1	Wireless Connection Module	80 -
4.3.2	Camera.....	81 -
4.3.3	Voice Recognition Module.....	81 -
4.3.4	Relay and Devices.....	82 -
4.4	Software Design	83 -

4.4.1	Recognize Gesture from Frame	- 83 -
4.4.2	Using ZigBee Module to Control Device.....	- 94 -
4.4.3	Using Voice Recognition to Recognize Voice Commands.....	- 95 -
4.4.4	Using Gesture and Voice Recognition to Control Devices.....	- 96 -
4.4.5	User Interface.....	- 98 -
CHAPTER 5	Tests and Results	- 99 -
5.1	Introduction	- 99 -
5.2	Test for Gesture Control.....	- 99 -
5.3	Test for Voice Control	- 102 -
5.4	Summary of the Tests	- 104 -
CHAPTER 6	Conclusion and Future Direction.....	- 105 -
6.1	Conclusion.....	- 105 -
6.2	Contributions	- 107 -
6.3	Future Direction.....	- 107 -
References.....		- 109 -

LIST OF ABBREVIATIONS

Various specialized abbreviations are used in this thesis as listed below:

API.....	Application Programming Interface
CamShift.....	Continuously Adaptive Mean Shift
CAN.....	Controller Area Network
CMOS.....	Complementary Metal-oxide Semiconductor
COM.....	Common
Cramfs.....	Compressed ROM File System
CPU.....	Central Processing Unit
DNN.....	Deep Neural Networks
DTW.....	Dynamic Time Warping
EEG.....	Electroencephalograms
EN.....	Enable
EPIC.....	Electric Potential Integrated Circuit
GPIO.....	General-Purpose Input/Output
GND.....	Ground
GUI.....	Graphical User Interface
HMM.....	Hidden Markov Model
HSV.....	Hue-Saturation-Value
IoT.....	Internet of Things
IR.....	Infrared
ISO.....	International Organization for Standardization
JFFS.....	Journaling Flash File System
LCD.....	Liquid-Crystal Display
LED.....	Light-Emitting Diode

MCU.....	Microcontroller Unit
MFC.....	Mel Frequency Cepstum
MFCC.....	Mel Frequency Cepstum Coefficients
MSEPF.....	Mean Shift Embedded Partial Filter
NFS.....	Network File System
NC.....	Normally Closed
NO.....	Normally Open
RAM.....	Random-Access Memory
RFID.....	Radio-Frequency Identification
RGB.....	Red-Green-Blue Color Model
ROM.....	Read-Only Memory
SASOM.....	structure adaptive self-organizing map
SCFG.....	Stochastic Context Free Grammar
SIFT.....	Scale Invariant Feature Transform
SPDT.....	single-pole, double-throw
SURF.....	Speeded Up Robust Features
TTL.....	Transistor–transistor Logic
UART.....	Universal Asynchronous Receiver/Transmitter
UI.....	User Interface
URC.....	Universal Remote Console
USB.....	Universal Serial Bus
YAFFS.....	Yet Another Flash File System

LIST OF TABLES

Table 3.1: The accuracy of three algorithms	- 39 -
Table 3.2: Predefined categories	- 49 -
Table 4.1: Hardware information for Mini210s	- 64 -
Table 4.2: Main menus and illustration of Linux kernel source	- 67 -
Table 4.3: Voice commands	- 96 -
Table 5.1: The results of test 1	- 101 -
Table 5.2: The results of test 2	- 102 -

LIST OF FIGURES

Figure 1.1: The Internet of Things	- 3 -
Figure 1.2: An example of gesture control	- 4 -
Figure 2.1: The landscape for IoT	- 10 -
Figure 2.2: Display of Leap Motion	- 13 -
Figure 3.1: The prototype of SoundSense system.....	- 16 -
Figure 3.2: An example of 3 LED system	- 17 -
Figure 3.3: An internal circuit of a EPIC sensor	- 18 -
Figure 3.4: The prototype of AllSee	- 19 -
Figure 3.5: Eight Gestures AllSee can detect	- 20 -
Figure 3.6: The output of the detector when the using making a certain gesture .-	20 -
Figure 3.7: WiiMote controller and the accelerometer axes	- 21 -
Figure 3.8: Gyro sensor	- 22 -
Figure 3.9: Using magnet to control a smart phone	- 23 -
Figure 3.10: A wearable EEG sensor	- 23 -
Figure 3.11: Color segmentation for hand detection	- 25 -
Figure 3.12: Multiple approaches for hand detection	- 26 -
Figure 3.13: The outline of the system	- 27 -
Figure 3.14: SURF vs SIFT	- 28 -
Figure 3.15: The mean shift algorithm	- 29 -
Figure 3.16: Sampled tracking results	- 30 -
Figure 3.17: Recognition results for hand postures	- 31 -
Figure 3.18: Process of 3D hand modelling	- 32 -
Figure 3.19: Example of 3D hand tracking system	- 33 -
Figure 3.20: Retrieval results of sample images	- 34 -
Figure 3.21: An example of word 'No' in Matlab	- 36 -
Figure 3.22: Process of speech recognition	- 37 -
Figure 3.23: Natural user interface system	- 40 -
Figure 3.24: The gesture used to zoom	- 41 -
Figure 3.25: Samples of negative images	- 42 -

Figure 3.26: Samples of positive images	43 -
Figure 3.27: Flow chart of process	44 -
Figure 3.28: Architecture of system	46 -
Figure 3.29: Kinect device	47 -
Figure 3.30: Gestures and their representation in 3D	48 -
Figure 4.1: Overview of the system design	51 -
Figure 4.2: Local area network connected to the Internet	52 -
Figure 4.3: An example of contours of images.....	53 -
Figure 4.4: An example of surroundness and border	55 -
Figure 4.5: Conditions for outer and hole borders	56 -
Figure 4.6: Decision rule for the parent border	56 -
Figure 4.7: The process of border following algorithm	57 -
Figure 4.8: Convex hull of a simple polygon	59 -
Figure 4.9: Algorithm lefthull	60 -
Figure 4.10: An example of using the algorithm	61 -
Figure 4.11: The optimal alignment for the given two sequences	61 -
Figure 4.12: Wrapping path between the two sequences	62 -
Figure 4.13: The development board	65 -
Figure 4.14: Typical flash memory layout	66 -
Figure 4.15: Framework of Qt for Embedded Linux	66 -
Figure 4.16: Linux kernel configuration	68 -
Figure 4.17: Step 1 of minicom setting	70 -
Figure 4.18: Step 2 of minicom setting	70 -
Figure 4.19: Restarting of NFS service	72 -
Figure 4.20: Version check for cross compiler.....	73 -
Figure 4.21: CMake configuration for OpenCV cross compilation – configure step2..	76 -
Figure 4.22: CMake configuration for OpenCV cross compilation – configure step3..	77 -
Figure 4.23: Hardware overview.....	79 -
Figure 4.24: Zigbee module	80 -
Figure 4.25: 10moons V804 camera and Voice recognition module	81 -

Figure 4.26: Pololu basic SPDT relay carrier	- 82 -
Figure 4.27: Lamp and desk fan used in project	- 83 -
Figure 4.28: The process of gesture recognition.....	- 84 -
Figure 4.29: Sampling image.....	- 86 -
Figure 4.30: The process of generating the binary image	- 88 -
Figure 4.31: The process of recognizing gesture from the binary image.....	- 90 -
Figure 4.32: An example of convexity defects of the contours	- 92 -
Figure 4.33: The rule to determine whether the points are relevant	- 93 -
Figure 4.34: The vision of gesture recognition.....	- 93 -
Figure 4.35: The wiring diagram of ZigBee communication	- 94 -
Figure 4.36: The process of voice recognition.....	- 96 -
Figure 4.37: The structure of the prototype	- 97 -
Figure 4.38: The Fan and Lamp panel	- 98 -
Figure 5.1: The process of test 1	- 100 -
Figure 5.2: The process of test 2.....	- 102 -
Figure 5.3: Average times that testers speak for each command.....	- 103 -