

# Development of an Intelligent Vision Enhanced Multimodal Human-Robot Interaction for Service Robots in Domestic Environment

Pilippu Hewa Don Arjuna Shalitha Srimal

(168052E)

Thesis submitted in partial fulfillment of the requirements for the degree  
Masters in Electrical Engineering

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

August 2017

## DECLARATION

---

I declare that this is my own work and this dissertation does not incorporate without acknowledgment any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:

The above candidate has carried out research for the M.Sc. thesis under my supervision.

Signature of the Supervisor:

Date:

(Dr. A.G.B.P. Jayasekara)

**Abstract** – In the recent past, domestic service robots have come under close scrutiny among researchers. When collaborating with humans, robots should be able to clearly understand the instructions conveyed by the human users. Voice interfaces are frequently used as a mean of interaction interface between users and robots, as it requires minimum amount of work overhead from the users. However, the information conveyed through the voice instructions are often ambiguous and cumbersome due to the inclusion of imprecise information. The voice instructions are often accompanied with gestures especially when referring objects, locations, directions etc. in the environment. However, the information conveyed solely through gestures is also imprecise. Therefore, it is more effective to consider a multimodal interface rather than a unimodal interface in order to understand the user instructions. Moreover, the information conveyed through the gestures can be used to improve the understanding of the user instructions related to object placements.

This research proposes a method to enhance the interpretation of user instructions related to the object placements by interpreting the information conveyed through voice and gestures. The main objective of this system is to enhance the correlation between the user expectation and the placement of the object by interpreting uncertain information included in user commands. Furthermore, several human studies have been carried out in order to understand the factors that may influence and their level of influence on the object placement. The proposed system is capable of adapting and understanding, according to the spatial arrangement of the workspace of the robot as well as the position and the orientation of the human user. Fuzzy logic system is proposed in order to evaluate the information conveyed through these two modalities while considering the arrangement, size and shape of the workspace. Experiments have been carried out in order to evaluate the performance of the proposed system. The experimental results validate the performance gain of the proposed multimodal system over the unimodal systems.

**Keywords-** Human Robot Interaction, Natural Language Understanding, Deictic Gestures, Fuzzy Spatial Terms

## ACKNOWLEDGMENTS

---

It is with great pleasure that I acknowledge the assistance and contribution of all the people who helped me to successfully finish my Master's thesis.

First, I extend my sincere gratitude to my research supervisor Dr. Buddhika Jayasekara who provided me with his continuous support and assistance throughout the course of this thesis. Without his advices and encouragement, this thesis would never have been accomplished. I also would like to thank my progress review committee members, Dr. Chandima Pathirana and Dr. Ruwan Gopura for delivering their guidance and valuable comments for successfully continuing my thesis work throughout the past year.

I specially acknowledge the efforts put into reviewing this thesis by Prof. Nalin Wickramarachchi and Prof. Chandimal Jayawardena and thankful for the comments and suggestions.

I would not have been able to complete this thesis successfully, if it was not for the overwhelming assistance given by the individuals in my research group. I sincerely thank and appreciate Viraj Muthugala, Chapa Sirithunge and Sajila Wickramarathne for their support and suggestions in pursuing my Masters thesis.

Also, my appreciation goes to the staff and all of my friends in the Department of Electrical Engineering for their invaluable support, specially in taking part in the conducted experiments and surveys.

Finally, I would like to extend my deepest gratitude to my family. The blessings of my mother, father and sister undoubtedly helped me in making this endeavor a success.

This work was supported by University of Moratuwa Senate Research Grant Number SRC/CAP/16/03.

# TABLE OF CONTENTS

---

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgments</b>	<b>iii</b>
<b>Table of Contents</b>	<b>viii</b>
<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem Statement . . . . .	3
1.2 Thesis Overview . . . . .	3
<b>2 Literature Review</b>	<b>5</b>
2.1 Human-like Robotic Assistants . . . . .	5
2.2 Human Robot Interaction Methods . . . . .	7
2.3 Understanding Uncertain Information . . . . .	8
2.3.1 Uncertainty in Spatial Terms . . . . .	9

2.4	Summary of Literature Review . . . . .	10
<b>3</b>	<b>Human Studies on Understanding Behavioral Concepts</b>	<b>13</b>
3.1	Understanding of Spatial Terminology . . . . .	15
3.1.1	Arrangement . . . . .	17
3.1.2	Procedure and Metrics . . . . .	17
3.1.3	Analyzing of Data . . . . .	18
3.1.4	Results and Discussion . . . . .	18
3.2	Effect of the User Location and Orientation . . . . .	19
3.2.1	Arrangement . . . . .	19
3.2.2	Procedure and Metrics . . . . .	19
3.2.3	Analyzing of Data . . . . .	23
3.2.4	Results and Discussion . . . . .	23
3.3	Effect of the Surface Area . . . . .	24
3.3.1	Arrangement . . . . .	24
3.3.2	Procedure and Metrics . . . . .	25
3.3.3	Analyzing of Data . . . . .	26
3.3.4	Results and Discussion . . . . .	26
3.4	Effect of the Table Shape . . . . .	29
3.4.1	Arrangement . . . . .	29
3.4.2	Procedure and Metrics . . . . .	29
3.4.3	Analyzing of Data . . . . .	29

3.4.4	Results and Discussion . . . . .	32
3.5	Effect of the Objects on the Table . . . . .	32
3.5.1	Arrangement . . . . .	33
3.5.2	Procedure and Metrics . . . . .	33
3.5.3	Analyzing of Data . . . . .	33
3.5.4	Results and Discussion . . . . .	35
3.6	Effects of the Restrictions for Reachability . . . . .	35
3.6.1	Arrangement . . . . .	35
3.6.2	Procedure and Metrics . . . . .	36
3.6.3	Analyzing of Data . . . . .	36
3.6.4	Results and Discussion . . . . .	36
3.7	Summary of Human Studies . . . . .	37
<b>4</b>	<b>System Design</b>	<b>40</b>
4.1	System Overview . . . . .	40
4.2	Understanding Vocal Commands . . . . .	41
4.3	Understanding Pointing Hand Gesture Location . . . . .	41
4.3.1	Kinect . . . . .	42
4.3.2	Tracking of User . . . . .	43
4.3.3	Obtaining Pointed Location . . . . .	43
<b>5</b>	<b>Understanding Uncertain Information in User Commands</b>	<b>46</b>
5.0.1	Uncertain Information Understanding Module . . . . .	46

5.0.2	Module 1 - For Voice Based User Commands . . . . .	47
5.0.3	Module 2 - For Hand-gesture Based User Commands . . . . .	47
5.0.4	Module 3 - Combined User Commands . . . . .	49
<b>6</b>	<b>Spatial Concerns</b>	<b>51</b>
6.1	Concerns for Space Properties . . . . .	51
6.1.1	Table Size . . . . .	51
6.1.2	Table Shape . . . . .	51
6.2	Effect of Dynamic Space Constraints . . . . .	53
6.2.1	Objects on the Table . . . . .	53
6.2.2	Reachability of the Robot . . . . .	55
6.3	Special Attention . . . . .	57
6.3.1	Concerns for User Location and Orientation . . . . .	57
6.3.2	Calculating the Placement Position . . . . .	58
6.3.3	Safety Distance . . . . .	58
6.4	Example Scenario . . . . .	58
<b>7</b>	<b>Results and Discussion</b>	<b>65</b>
7.1	Hardware Implementation . . . . .	65
7.2	Experimental Setups . . . . .	65
7.2.1	Experimental Setup 1 . . . . .	65
7.2.2	Experimental Setup 2 . . . . .	69
7.3	System Evaluation . . . . .	72



7.3.1	Basic Multimodal System . . . . .	72
7.3.2	Effect of Dynamic Space Constraints . . . . .	73
7.3.3	Special Attention . . . . .	74
7.3.4	System Limitations . . . . .	75
<b>8</b>	<b>Conclusions</b>	<b>78</b>
	<b>List of Publications</b>	<b>81</b>
	<b>References</b>	<b>82</b>

## LIST OF FIGURES

---

1.1	Number of elderly population in the world estimated till 2050. . .	1
1.2	The sales of service robots up till 2010 and estimated values till 2050. . . . .	2
2.1	Zora robots . . . . .	6
2.2	ASIMO by Honda . . . . .	6
2.3	Motoman by Yaskawa Electric . . . . .	7
2.4	Summary of the literature review. . . . .	12
3.1	Example scenarios for human studies. . . . .	14
3.2	Arrangement for the study inorder to understanding of spatial terminology . . . . .	15
3.3	Effect of the user rating. . . . .	15
3.4	Color coded cards that were used for the experiment . . . . .	16
3.5	Results of the study perfomed to understand the spatial terminology	20
3.6	Summery of the table area allocation . . . . .	21
3.7	The effect of the user's orientation. . . . .	22
3.8	Effect of the reference frame. . . . .	23
3.9	Results of the study conducted to understand the user's attention.	25

3.10	Results of the study for effect of table area . . . . .	27
3.11	The box plot diagram of X coordinate for spatial term “Back Edge” for left handed participants . . . . .	28
3.12	Different shapes of tables that were used for the study . . . . .	30
3.13	Summery of the results of thr study effect of the table shapes . . .	31
3.14	The table setup that was used to the study the effects of the objects on the table. . . . .	32
3.15	Results for table setting (a) . . . . .	33
3.16	Results for table setting (b) . . . . .	34
3.17	Results for table setting (c) . . . . .	34
3.18	Effect of the restricted reachability. . . . .	36
4.1	System overview. . . . .	40
4.2	The tracked joints of the body by the Kinect . . . . .	42
4.3	Reference axis of Kinect . . . . .	43
4.4	Top view of the Kinect with respect to the table. . . . .	44
4.5	Skeleton tracking of Kinect . . . . .	44
5.1	Command types that are used by the user. . . . .	46
5.2	Input and output fuzzy inference functions for UIUM submodules 1,2 and 3. . . . .	48
5.3	An example scenario for hand gesture based user command. . . .	49
6.1	Recalculating of pointed hand gesture location for oval and circular shaped tables. . . . .	52

6.2	Shift in the fuzzy output curved due to the objects on the table. .	54
6.3	An example for extracting occupied area by objects on the table. .	55
6.4	Effect of the User's orientation . . . . .	56
6.5	Safety concerns when placing the object. . . . .	59
6.6	Example scenario using both voice and hand gesture based com- mands. . . . .	60
6.7	Robots point of view of the user. . . . .	60
6.8	Tracked skeleton of the user's body. . . . .	61
6.9	Y and X axis transformations from oval shaped table to a rectan- gular shaped table. . . . .	61
6.10	The robots point of view of the table. . . . .	63
6.11	The robot moves forward to place object. . . . .	63
6.12	The robot completes the placement of the object. . . . .	63
6.13	The robot is returning to the starting position . . . . .	64
7.1	MIROB - The hardware that was used to implement the system. .	66
7.2	Experimental setup 1 table settings . . . . .	67
7.3	Placement of objects on the table for setup 1. . . . .	68
7.4	User satisfaction for setup 1. . . . .	68
7.5	Shows the room setup that was used for experimental setup 2. . .	70
7.6	The map of the room that was used for experiments in setup 2 . .	71
7.7	Different table shapes that were used during the experimental setup 2 . . . . .	71

8.1	User satisfaction in unimodal and multimodal systems. . . . .	80
-----	---	----

## LIST OF TABLES

---

3.1	Usage of Area Terms . . . . .	18
3.2	Area Terms Used in Study 2 . . . . .	19
3.3	Results for Effects of Table Shape . . . . .	32
3.4	Summary of Human Studies . . . . .	39
5.1	Rule Base For Fuzzy Modules 1 And 2 . . . . .	47
5.2	The Rule Base For Fuzzy Module 3 . . . . .	50
7.1	Experiment Results for Setup 1 . . . . .	76
7.2	Experiment Results for Setup 2 . . . . .	77