Framework for Adaptive Human-Robot Interaction Initiation for Domestic Environments

Hewa Pelendage Chapa Sirithunge

(168060C)

Degree of Doctor of Philosophy

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

December 2020

Framework for Adaptive Human-Robot Interaction Initiation for Domestic Environments

Hewa Pelendage Chapa Sirithunge

(168060C)

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

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

December 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate 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 conducted research for the PhD thesis under my supervision.

Signature of the Supervisor(s): Date:

Prof. A.G. Buddhika P. Jayasekara

DEDICATION

To my beloved family and all the people who strive hard to make the world better than today

Abstract

Intelligent robot companions contribute significantly in improving living standards in the modern society. Therefore human-like decision making skills and perception are sought after during the design of such robots. On the one hand, such features enable a robot to be easily handled by a non-expert human user. On the other hand, the robot will have the capability of dealing with humans without causing any disturbance by the its behavior. Mimicking human emotional intelligence is one of the best and reasonable ways of laying the foundation for such an emotional intelligence in robots. As robots are widely deployed in social environments today, perception of the situation or the intentions of a user prior to an interaction is required in order to be proactive. Proactive robots are required to understand what is communicated by the human body language prior to approaching a human. Social constraints in an interaction could be demolished by this assessment in this regard.

This thesis addresses the problem of perceiving nonverbals in human behavior and fusing human-environment semantic representations with a robot's cognition before interacting with humans. The novelty lies in laying the background to relate nonverbal human behavior and the features of the environment to generate context-aware interactive responses during robot-initiated interaction. That informs the robot about its environment. Toward this end, we introduce novel methods of perceiving human nonverbals and spatial factors in the environment which make up a context in which we integrated that knowledge to determine appropriate responses from a robot to assist its user. Visual information acquired by a vision sensor was analyzed, and the level of emotional engagement demanded by the user's nonverbals was evaluated, before a robot initiates an interaction. After such an analysis, a robot's conversational and proxemic behavior was adjusted to maintain an empathetic relationship between the user and the robot. Our algorithms efficiently sustained the empathy between user and robot so that the interaction resembles human-human interaction to a larger extent. To assist the main problem, we formulated novel methods to recognize human nonverbals such as postures, gestures, hand poses, psychophysiological behavior of humans and human activities, and decode the emotional hints displayed to the outside world. In support of this work, we conducted a series of human studies to explore the patterns in human behavior which could be perceived by a proactive robot using its cognitive capabilities.

We introduce separate systems which can decode the sentiments of humans using observable cues based on accepted social norms. We detail the meanings of human nonverbals by observing human behavior over time and evaluating the context for any patterns in behavior. Ambiguities in human behavior and random, meaningless behaviors could be omitted through this approach. This approach further omits the negative effect of human responses that can be faked, such as facial expressions and words. Finally we introduce an adaptive approach towards robot-initiated human-robot interaction by letting a robot observe a context and generate responses while changing its responses continuously as human behavior changes. We first developed algorithms based on a limited number of observable human cues and decoded their sentiments based on modern psycho-physiological interpretations of human behavior. Next, we expanded such approaches towards multiple observable human cues. Finally we integrated observations from the human and the environment which create the context during HRI (Human-Robot Interaction). Hence we integrated all the recognition

approaches to perceive a complete scenario which comprises the user, robot and the environment.

Upon unimodal systems to identify these features independently, we propose a multi modal approach to evaluate above features together to understand a scenario. Through this approach, we took an effort to make proactive behavior of a social robot more instinctive, ethical and socially acceptable or simply, humanlike. We evaluate this approach by means of physical experiments in simulated social and domestic environments and demonstrate its performance in appropriate occasions as determined by a robot according to the formulated criteria of perceiving a context.

***Keywords*- Nonverbal behavior, Interaction initiation, Proactive robots, Social Human-robot Interaction**

ACKNOWLEDGMENT

It is with great pleasure that I acknowledge the support and contribution of all the mentors and colleagues who assisted me in successfully completing my PhD thesis.

First I would like to express my sincere gratitude to my advisor Dr. Buddhika Jayasekara. It was a privilege to work under his guidance and I thank him for his support, encouragement and patience. Without his tireless efforts, this thesis would never have been accomplished. Also I would like to extend my gratitude to my progress review panel, Prof. Ruwan Gopura and Dr. Chandima Pathirana for their insightful comments and helpful suggestions. I specially acknowledge the efforts put by the thesis examination panel.

Furthermore I would like to thank the following people who helped me undertake this research.

- Arjuna Srimal, Sajila Wickramaratne and the members of Intelligent Service Robotics Group (ISRG), especially Gayashan Porawagamage, Viraj Muthugala, Sahan Kodikara, and Ravindu Bandara, Department of Electrical Engineering, University of Moratuwa.
- My parents and Nikolas Dahn for their kind support and patience through out the journey.
- Dear teachers and my dear friends who set me off on the road to this PhD a long time ago should accept my sincere gratitude.
- The staff of the Robotics and Automation Laboratory in the Department of Electrical Engineering

-People who were generous with their time and effort in participating in experiments related to research.

- The staff and my friends in the department of Electrical Engineering for their immense support.
- My family, my friends, numerous to mention, and everybody who were with me in making this endeavor a success.

This work was partially supported by the University of Moratuwa Senate Research Grants SRC/CAP/17/03 and SRC/LT/2018/20. Furthermore, some of the publications originated from this thesis were partially supported by travel grants awarded by the National Science Foundation of Sri Lanka, IEEE Robotics and Automation Society and European Association for Artificial Intelligence.

TABLE OF CONTENTS

Declaration i

Dedication ii

Abstract iii

Acknowledgments v

Table of Contents xiii

List of Figures xx

List of Tables 1

1 Introduction 1

1.1 Development of assistive robots 1

1.2 Building Attentive Robots 5

1.2.1 Study of human behavior 5

1.2.2 "Meet, Perceive and Act" 7

1.2.3 Multimodal perception 7

1.3 Problem Statement 8

1.4 Thesis Contributions 8

1.4.1 Identifying observable human behavior 8

1.4.2 Recognizing components embodied in a context 9

1.4.3 Decoding nonverbals to interpret user situation 10

1.4.4 Adapting to a situation 11

1.5 Thesis Overview 11

2 Literature Review **13**

2.1 Nonverbal Interaction during HRI 13

2.1.1 Features to observe 14

2.1.2 Earlier Approaches 15

2.1.3 Modern Developments 17

2.1.4 Systems with Adaptive Perception of Situation based on Human Cues 19

2.2 Outlook 26

3 System Design **28**

3.1 Extracting information 28

3.2 Robot Platform 28

3.3 Overview of the system 29

4 Study of Human Behavior **33**

4.1 Study 1: Human Interest towards Robot Initiated Human-Robot Interaction . . . 33

4.1.1 Monitoring Human Behavior . . . 33

4.1.2 Results of the experiment . . . 35

4.1.3 Observations and Discussion . . . 38

4.2 Study 2: Conversational Preferences in Social Human-Robot Interaction . . . 41

4.2.1 Setting . . . 42

4.2.2 Results and discussion . . . 46

4.2.3 Observations and Discussion . . . 49

4.2.4 Conclusions and Implications . . . 52

4.3 Study 3: Human Gaze Behavior . . . 54

4.3.1 Understanding Human Behavior: Evaluate Attention . . . 54

4.3.2 Time scales of human actions . . . 54

4.3.3 Monitoring Gaze Behavior . . . 55

4.3.4 Results of the experiments and Discussion . . . 57

4.3.5 Finding Answers to previous questions . . . 64

5 Recognition of nonverbals **66**

5.1 Psychophysiological Aspects of Human Behavior . . . 66

5.1.1 Simple Human Activities 66
5.1.2 Criteria used to analyze activity space 67
5.1.3 Analyzing Extracted Information 69
5.1.4 Decision Evaluation 70
5.2 Hand pose 75
5.2.1 Recognizing hand pose 75
5.2.2 Experiment and results 79
5.3 Pose recognition and robot-initiated interaction 81
5.3.1 Interpretation of Fuzzy Information 85
5.4 Interpreting movements of the body 90
5.4.1 Measuring Behavioral Responses 90
5.4.2 Observable Nonverbal Human Behavior 91
5.4.3 Decision Making Criteria 93
5.4.4 Monitoring Human Behavior 95
5.4.5 Experiment 1 98
5.4.6 Experiment 2 98
5.4.7 User Responses Towards Robot 99
5.5 Recognition of Hand and Body Pose 103
5.5.1 Changes in Behavior 106
5.5.2 Mechanism used to differentiate postures 108

5.5.3 Behavior Evaluation Model 113
5.5.4 Posture based Interaction Decision Making 114
5.5.5 Valence of a situation 115
5.5.6 Decision-making in dynamic user behavior 117
5.5.7 Experiment and Results 118
5.5.8 Results and Discussion 122
5.6 Recognition of complex human activities 127
5.6.1 Robot, user and the context 127
5.6.2 Features of an activity 131
5.6.3 Recognition of activities 132
5.6.4 Learning new activities 135
5.6.5 Experiment and Results 137
5.6.6 Setting 137
5.6.7 Results 139
5.6.8 Conclusion 144

6 Decoding nonverbals and Modelling User Behavior 145

6.1 Identification of Gaze 145
6.1.1 Determination of Gaze direction and Gaze time 146
6.1.2 Return of Gaze 146

6.2 Identification of Friendly Gestures 147
6.3 Identification of Changes in Pose 148
6.3.1 Evaluation of attention 150
6.3.2 Interpretation of Fuzzy Parameters 152
6.3.3 Decision Making 152
6.3.4 Experiment and Results 153
6.3.5 Conclusions 156

7 Determining Robot's Proactive Behaviors: Proxemics 157

7.1 Probabilistic analysis 158
7.2 Decision making criteria 159
7.3 Results and Discussion 162
7.3.1 Observations and discussion 163
7.3.2 Conclusions 167

8 Modeling a context: Autoregressive Model 168

8.1 Level of Interest 169
8.2 AR model to calculate LOI 170
8.3 Decision making criteria 172
8.4 Results and Discussion 173
8.5 Observations and Discussion 176

8.6 Conclusions 180

9 A Social Robot with Experience and Learning 181

9.0.1 Semantic Mapping of the features 183

9.0.2 Robot behavior 184

9.0.3 Selection of participants and creating training data 184

9.1 Conclusions 187

10 Conclusions 189

10.1 Implications for Theory 193

10.2 Implications for Design 195

10.3 Considerations for future 196

References 201

LIST OF FIGURES

1.1 An example of a cooking robot 2

1.2 An airport cleaning robot 2

1.3 Example scenarios where the required level of situation-awareness differs. 4

1.4 An example domestic environment is shown. 10

1.5 Different types of settings and components associated with HRI . 11

3.1 Hardware setup used in experiments. 29

3.2 Overall system 31

4.1 Poses used as behavioral changes 36

4.2 An example scenario during the experiment 36

4.3 A comparison of the types of behavioral changes observed during a relaxing task 38

4.4 A comparison of the types of behavioral changes observed while user was engaged 38

4.5 A typical conversation between the robot and its user 44

4.6 An example scenario during the experiment 47

4.7 A situation in which the user was having a drink 47

4.8 A stacked graph drawn for the comparison of conversational preferences with the two conditions; when the user is alone and when surrounded by few people 51

4.9 conversational preferences during the selected tasks while the domestic area and people in the surrounding were kept constant . 53

4.10 Phases in a typical encounter prior to an interaction 55

4.11 Two scenarios during the experiment 57

4.12 Existence or nonexistence of gaze during HHI is plotted against each user 58

4.13 The percentage (%) of encounters with any kind of gaze plotted againt each occasion 59

4.14 The average duration of existed gaze was plotted against the delay 59

4.15 The average delay of gaze was plotted against the occasion 60

4.16 The delays of encounters where an averted gaze was present . . . 60

4.17 The delays of encounters where a continued gaze was present . . . 61

4.18 The delays of encounters where an averted gaze was present . . . 62

4.19 Responses observed from users during HHI are plotted against each individual 63

5.1 Arrangement of activity zones near a person 67

5.2 Dimensions and joints used to separate zones 68

5.3 Approach of QASA 70

5.4 Various occasions tested with the 2 systems:ASA and OASA . . . 72

5.5 Boxplot from the results obtained for ASA and QASA 74

5.6 Criteria used to determine length, width and height of the mesh . 76

5.7 The 3D mesh created out of the 'activity zones' established around a human 77

5.8 Utilization of zones during a certain gesture 77

5.9 The set of gestures used to recognize using the proposed approach 78

5.10 Depth images corresponding to the gestures from a-k 78

5.11 Types of Standing poses used in the experiment 82

5.12 Types of Sitting poses used in the experiment 82

5.13 Routes followed by robot during various situations. 85

5.14 Input and output membership functions used to find the interaction demanding 86

5.15 An occasion with 'high' interaction demanding 88

5.16 An occasion with 'low' interaction demanding 89

5.17 Skeletals extracted from a standing and a seated person 92

5.18 Nonverbal behavioral responses observed during the experiment . 93

5.19 The decision grid 95

5.20 Two occasions from the experiment 99

5.21 Two occasions in which adaptive approach behavior was implemented 100

5.22 A comparison between the average feedback scores received for direct approach and adaptive behavior 100

5.23 Two occasions from the experiment 102

5.24 Overview of the system 105

5.25 A set of body postures encountered daily 107

5.26 A set of arm postures encountered daily 107

5.27 Body angles derived from the limbic vectors of a person with stretched hands 109

5.28 Limbic vectors of a seated person 109

5.29 Body angles derived from limb vectors of a seated person as in Fig. 5.28 112

5.30 A semantic map illustrating the components that make valence; hand and body poses 115

5.31 Marginal values chosen for the valence to determine appropriate approach behavior of the robot 117

5.32 A comparison of the limbic arrangement for regular sitting posture with different chairs 121

5.33 Two occasions during experiment 3 125

5.34 This box and whisker plot shows the mean values of feedback scores received from users for the robot's proactive behavior with IDMM and without IDMM 125

5.35 A semantic map which shows the co-relation between various aspects of a human activity 128

5.36 Objects and aspects related to each semantic feature 129

5.37 The relationship between various steps involved in an activity . . 129

5.38 The decision flow of the approach during learning and recognizing activities. 130

5.39 Upper and lower limits of speeds related to each speed range . . 131

5.40 A typical conversation between the robot and its user while learning an activity 135

5.41 A map of the domestic area used for the study 137

5.42 Different occasions encountered during the experiment 140

5.43 Recognition accuracies of the 4 occasions considered in the experiments 143

5.44 Recognition accuracies when two occasions considered 143

6.1 Change of behavior encountered before and after approaching a human 148

6.2 Spatial vectors related to the body joints used to differentiate two poses 149

6.3 Input membership functions of the gaze parameters 150

6.4 Input membership functions of the gesture parameters 151

6.5 Output membership functions of each parameter 151

6.6 Output Membership Functions of second fuzzy evaluation 151

6.7 Rule Base to evaluate Gaze Level 152

6.8 Rule base for Attention level evaluation 152

6.9 Two occasions from a typical domestic environment 155

7.1 Approach of QASA 158

7.2 The top view of the HRI scenario 160

7.3 Distances marked considering maximum occupancy regions of the selected set of tasks 164

7.4 Two occasions encountered during PHN 165

7.5 This graph shows the mean values of feedback scores received from users for the robot's behavior in the three considered occasions . . 166

8.1 A semantic map which shows the co-relation between various aspects in human behavior 169

8.2 Features used as the inputs to the AR model are shown. 171

8.3 Marginal values to choose the *nature of conversation* and *proxemics* of the robot. 173

8.4 Marginal values to choose the *orientation* of the robot. 173

8.5 Values obtained for LOI_1, LOI_2 and LOI_3 during ordinary situations encountered 175

8.6 Occasions involved during SEA 179

8.7 The mean values of feedback scores received from users for the acceptance of robot's behavior in three occasions 179

9.1 The action-state model 182

9.2 Decision flow of the robot platform 182

9.3 The interrelations of semantic features during the Observation Plan and the Monitoring Plan. 183

9.4 Observational parameters considered in the system for robot decisioning. 184

9.5 Q-value test for 4 occasions while training 186

9.6 Q value tests for 3 occasions during training are shown. (The user feedback scores plotted against each user 187

LIST OF TABLES

4.1 Results of the experiment 36

4.2 T-Test for the comparison among each situation: while relaxing and while engaged when the person initiating the interaction was the robot and another human. 37

4.3 Some of the tasks selected for the study 43

4.4 A comparison of conversational preferences by the type of interaction when the user was alone and when with few people around 46

4.5 ANOVA test for the comparison of percentage frequencies of each interaction type in each area of the social environment 48

4.6 t- Test for the comparison among each type of interaction when the user was alone and with few people around 49

4.7 ANOVA test for the comparison of the frequencies of each conversational preference during each task 52

4.8 T-test for the comparison of gaze parameters among HHI and HRI-week 3 64

5.1 Experiment results 73

5.2 Results of the experiment obtained for two occasions: while users were posing for the sensor and recognition accuracy of hand pose while the users were seated. 80

5.3 Results of the experiment 87

5.4 An analysis of the confusion matrix generated from the results of experiment 2 102

5.5 Observations made through time 112

5.6 Valence assigned for each hand and body posture depending on the general interpretation of emotional state behind each posture 116

5.7 Results of experiment 01 119

5.8 Results of experiment 03 123

5.9 Observations made through time 141

6.1 Experiment Results 154

7.1 The set of zones and their corresponding probabilities of occupancy obtained for each task 161

7.2 Results of ten scenarios during the experiment and average feedback score received for each task 165

7.3 T-test for the comparison of user feedback scores in initial and final attempts 166

8.1 Calculated values based on observations 177

8.2 Results of the experiment 178