RESHAPING HUMAN INTENTIONS BY AUTONOMOUS SOCIABLE ROBOT
MOVES THROUGH INTENTION TRANSIENTS GENERATED BY ELASTIC
NETWORKS CONSIDERING HUMAN EMOTIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ORHAN CAN GÖRÜR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

APRIL 2014

Approval of the thesis:

RESHAPING HUMAN INTENTIONS BY AUTONOMOUS SOCIABLE ROBOT MOVES THROUGH INTENTION TRANSIENTS GENERATED BY ELASTIC NETWORKS CONSIDERING HUMAN EMOTIONS

Submitted by **ORHAN CAN GÖRÜR** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**   _____

Prof. Dr. Gönül Turhan Sayan
Head of Department, **Electrical and Electronics Engineering**   _____

Prof. Dr. Aydan M. Erkmen
Supervisor, **Electrical and Electronics Engineering Dept, METU**   _____

Examining Committee Members:

Assoc. Prof. Dr. Afşar Saranlı
Electrical and Electronics Engineering Dept., METU   _____

Prof. Dr. Aydan M. Erkmen
Electrical and Electronics Engineering Dept., METU   _____

Prof. Dr. Aydın Alatan
Electrical and Electronics Engineering Dept., METU   _____

Assist. Prof. Dr. Yiğit Yazıcıoğlu
Mechanical Engineering Dept., METU   _____

Assist. Prof. Dr. Sinan Kalkan
Computer Engineering Dept, METU   _____

Date: 11.04.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: ORHAN CAN GÖRÜR

Signature:

ABSTRACT

RESHAPING HUMAN INTENTIONS BY AUTONOMOUS SOCIABLE ROBOT MOVES THROUGH INTENTION TRANSIENTS GENERATED BY ELASTIC NETWORKS CONSIDERING HUMAN EMOTIONS

Görür, Orhan Can

M.Sc., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. Aydan M. Erkmen

April 2014, 124 pages

This thesis focuses on reshaping a previously detected human intention into a desired one, using contextual motions of mobile robots, which are in our applications, autonomous mobile 2-steps and a chair. Our system first estimates the current intention based on human trajectory depicted as location and detects body-mood of the person based on proxemics behaviors. Our previous reshaping applications have shown that the current human intention has to be deviated towards the new desired one in phases according to the readiness changes induced in the human. In our novel approach, Elastic network plans way points (intention transients) by searching trajectories in the feature space of previously learned motion trajectories each labeled with an intention. Our methodology aims at planning an "intention trajectory" (sequences of intention transients) towards the final goal. The initial way points possess destabilizing effects on the obstinance of the person intention making "the robot gain the curiosity of the person" and induces positive mood to the person making "the robot gain the trust of the person". Each way point found by the elastic

network is executed by moves of an adequate robot (here mobile 2-steps or chair) in adequate directions (towards coffee table, PC, library). After each robot moves, the resulting human intention is estimated and compared to the desired goal in the intention space. Intention trajectories are planned in two modes: the "confident mode" and the "suspicious mode". This thesis work introduces our novel approach of planning trajectories based on elastic networks following these two modes.

Keywords: Human-Robot Interactions, Sociable Robots, Intention Reshaping, Elastic Networks, Emotional Body-Mood Detection, Path Planning, Intention Estimation

ÖZ

ELASTİK AĞLARI KULLANARAK, İNSAN DUYGULARINA GÖRE ÜRETİLEN GEÇİCİ NİYET ROTALARINI İZLEYEN OTONOM SOSYAL ROBOTLARIN İNSAN NİYETLERİNİ ŞEKİLLENDİRMESİ

Görür, Orhan Can

Yüksek Lisans, Elektrik ve Elektronik Mühendisliği

Tez Yöneticisi: Prof. Dr. Aydan M. Erkmen

Nisan 2014, 124 sayfa

Bu tez, önceden tespit edilen insan niyetlerini, mobil robotların çevreye duyarlı hareketleri ile hedeflenen niyete yönlendirme üzerine odaklanmıştır. Robotlar uygulamalarımızda otonom 2 basamaklı merdiven robot ile sandalye şeklinde robotlardır. Sistemimiz öncelikle insan rotalarına bakarak niyetlerini ve sosyal etkileşim uzaklıklarına bakarak ("Proksemi") vücut ruh hallerini tahmin etmektedir. Önceki niyet yönlendirme çalışmalarımız bize mevcut insan niyetinin hedef niyete yönlendirilmesinin, kişinin istekliliğine göre fazlar halinde olması gerektiğini gösterdi. Bu özgün yaklaşımımızda elastic ağlar robotlar için, insanlardan öğrenilmiş ve her biri belirli niyetlerle etiketlenmiş rotaları kullanarak, ara noktalar (geçici niyetler) planlamaktadır. Bu ara noktaların her biri hedef niyete giden ara niyetler olarak nitelendirilir ve aslında kişiyi niyete yönlendiren rotalardır. Başlangıç ara noktalar insanın o anlık uğraşındaki dikkatini dağıtmak ve bu dikkati robotun üzerine çekebilmek niyetiyle planlanır. Yine bu noktalar insanın sosyal etkileşim alanına girerek kişiye pozitif ruh hali aşılamak ve robota güvenmesini sağlamak amaçlıdır.

Elastik ağlar tarafından planlanan her ara noktadaki rota, uygun bir robot tarafından uygulanır. Rotalar test ortamımızdaki niyetlere göre kahve masasına, çalışma masasına veya kitaplığa doğru olabilir. Her bir robot hareketinden sonra, insan niyeti hedeflenen niyet ile karşılaştırılır. Niyet rotaları (ara noktalar) insan ruh haline göre planlanmaktadır. Bunlar "rahat mod" ve "şüpheli mod"dur.

Anahtar Kelimeler: İnsan-Robot Etkileşimleri, Sosyal Robotlar, Niyet Şekillendirme, Elastik Ağlar, Vücut-Ruh Hali Tespiti, Yol Planlama, Niyet Tahmin Etme

To my precious family

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVATIONS

BDI	(Belief-Desired-Intention)
BG	(Back Ground)
DBN	(Dynamic Bayesian Networks)
EN	(Elastic Networks)
EM	(Expectation Maximization)
FG	(Fore Ground)
FPS	(frame per second)
HMM	(Hidden Markov Model)
HRI	(Human Robot Interactions)
OOM	(Observable Operator Model)
ROI	(Region of interest)
RGB	(Red Green Blue color codes)
TSP	(Travelling salesman problem)

CHAPTER 1

INTRODUCTION

1.1. Motivation

Recent developments on both artificial intelligence and hardware capabilities for the robots resulted in greater advances in the field of Human-Robot Interaction (HRI). Nowadays, robots are being used in both industry and our homes as assistances to humans (Erden & Tomiyama, 2010). The fact that humans express their intentions based on the high variety of unpredictable interactions, generates requirements on assistant robots to understand and model these interactions towards satisfying the needs of their interacting human agents (Yokoyama & Omori, 2010). Robots capable of recognizing intentions and emotions of other agents and interacting through social behaviors develop the field of socially intelligent robots (Kerstin Dautenhahn & Billard, 1999).

In social interactions between intelligent agents, estimating the intentions and emotions of one another, called social cognition (Fong, Nourbakhsh, & Dautenhahn, 2003), is required to infer the behavior of the opponent agent which eventually results in inducing their own intentions onto the other by compromises or persuasion (Heinze, 2003; K. A. Tahboub, 2006). Inducing intentions on one other is due to morphing actions of one agent onto the other; we term this induction as intention reshaping (Durdu, Erkmen, Erkmen, & Yilmaz, 2011) based on strategic moves of an agent for the purpose of attaining a desired change on the other agent, relying on the statement that one's intentions direct one's future planning (Bratman, 1999). A pioneer study on intention reshaping in HRI field is by (Terada, Shamoto, Mei, & Ito, 2007) which only focuses on the behavioral changes of humans according to different designs and stances of the robots. However, in this study, the robots do not

behave to induce intentional changes on the humans; the work only approaches from the human perspective. In our laboratory's previous study (Durdu et al., 2011), a new approach is introduced on intention reshaping by developing context dependent robots (a chair and a 2-steps robots) that induce changes on the human intentions according to the contextual moves of these robots that are commanded with a remote controller. Our novel advancement to the same topic pertains to the full-autonomy of the robots that plan their own trajectories through the generation of intention transients (way points) using elastic networks aiming to break the obstinance of the person and attracting human intention towards the desired one. Elastic net generates these way points according to the detected human-body mood and the estimated current intention of the person, aiming first to increase confidence of the person towards the robot and the environment then to impose reshaping actions. This approach emulates a social interaction between humans, increasing the chance of the robots to understand human behaviors and react proactively.

Our main motivation on this study is to develop sociable behaviors for robots equipped with enhanced social cognition abilities to be applicable to real life scenarios. Assistant robots helping people in their daily lives in smart homes or industry are the ultimate aim where these robots can guide people in emergency situations where verbal communication is impossible by classifying them as being confident or suspicious and treating them accordingly. In addition, these robots can be used commercially, catching the attention of the people and leading them towards their shops.

1.2. Problem Definition and Our Approach

In our novel approach of intention reshaping, we develop fully autonomous robotic systems which are able to carry social cognitions from the on-line visual observations of human headings and trajectories. Our robots are able to estimate the current intention of the human, able to break the obstinance on his/her intention by detecting the emotional mood of the person according to proxemics in HRI as in (Christensen, Pacchierotti, & Hgskolan, 2005), and are able to decide whether the

person is confident enough to start an interaction. Based on these recognized behaviors, our robots do pro-active movements until a *confident mood* is detected on the person. This is realized by a successful induction of *confident mood* by appropriate robot moves, which brings the interaction to the final phase of intention reshaping. All of the paths for the robots are planned using intention transients generated by Elastic Networks among previously learned motion trajectories of human subjects each labeled with an intention.

Towards our goal, we used an experimental room similar to the one in (Durdu, 2012), where there are two context dependent robots being a chair and a 2-steps robot and the contextual objects such as: a library, a working table, a PC and a coffee table. The human and the robots are tracked via a ceiling camera on-line, the current intention and the body-mood of the person are estimated and the reshaping is performed with the aim of attaining a desired intention. Human subjects entering the room are not being told about the existence of any robots and of any interactions and are faced suddenly with our robots trying to reshape the estimated current intention of the person towards the desired one without the human being aware of this trial. The phases in our system are: choosing a desired intention from the feature space of previously learned human motion trajectories through Hidden Markov Model (HMM); estimation of the current intention of the person; detection of human body-mood as *confident* or *suspicious* based on proxemics; generating transient intentions ("way points") by elastic networks according to the detected human body mood. Our system uses Elastic networks as a search algorithm for a way point in the feature space of learned intention trajectories. The initial way points are generated in dense areas (intention areas in the feature space "familiar" to human subjects) around the current intention until a *confident mood* is induced on the human. The aim here is to have the robot "break the obstinance of the person" on what s/he was doing and "gain the curiosity and the trust of the person" relying on the psychological research that a *confident mood* results in more external-focused attention (collaborative person) (Fredrickson, 2003; Grol, Koster, Bruyneel, & Raedt, 2013; Sedikides, 1992). Inducing a *confident mood* is realized by gently approaching the person by intention transients lying in his/her familiar region, making him/her feel more

comfortable with the robot as in the studies of (Butler & Agah, 2001; Huettenrauch, Eklundh, Green, & Topp, 2006; Mead & Matarić, 2011). Finally, after a successful mood induction, the next way point is searched around the desired intention and eventually converging to that desired intention. Each found way point is realized by an adequate robot (for example, 2-steps robot for library and chair robot for working table related desired intentions). After the robot moves, human intention is estimated and compared with the desired one. A mismatch results in resuming a new detection of the human body-mood and a new way point (learned trajectory) search according to this mood being *suspicious* or *confident* which moods are also studied in (S. Lee & Son, 2008). We named these moods as execution modes of our reshaping since they decide upon the path planning strategy.

1.3. Contribution

The contributions of this thesis are given as:

- Reshaping human intentions into a desired new one by fully autonomous moves of sociable robots planned by elastic networks in real-time.
- Developing robots which have social cognition abilities and proactive reactions in real-time scenarios according to the behaviors of interacting person.
- Detecting human emotional body-moods on-line with a new approach utilizing proxemics behaviors.
- Planning paths for the robots to break the obstinance of the human and gain his/her curiosity and make that person confident based on proxemics behaviors.
- Estimating human intention *on-line* in a closed loop algorithm which is then compared to the desired intention and checked if intention reshaping is successful.
- Planning paths for the robots to guide the confident person towards the desired intention to be reshaped into.

4

- Adaptation of Elastic Networks as generator of intention transients based on the currently detected emotional mood and the estimated current intention trajectory of the person.

- A novel approach to traditional elastic networks using one dynamic node and changing effects of attractive forces at the iteration called *critical iteration* where oscillations of dynamic node settle down.

1.4. Outline of Thesis

This section outlines the thesis chapters and the topics to be covered. Starting with this chapter, we introduced the objective of the thesis with our motivation and contributions to the robotics field of research.

Chapter 2 gives the related works in the literature on the focus points of this thesis, which are intention estimation, emotional body-mood detection, effect of positive mood on external-focused attention of a person, emotional mood induction on a person and intention reshaping.

Chapter 3 details the methodologies in our study which are: intention estimation with HMM and trajectory estimation; execution mode decision according to the detected emotional body-moods of the person; our developed elastic networks for generation of intention transients and their execution through adequate robot moves.

Chapter 4 provides the results and discussions of our work. Simulation results on the usage of elastic networks are detailed together with the simulated intention reshaping scenarios and sensitivity analysis under parameter changes. In addition, results of real-time experiments with two different human subjects are given with snapshots from the recorded videos analyzing each methodology in Chapter 3 and picturing and discussing the entire closed loop system.

Finally, in Chapter 5 we conclude our thesis with a very brief summary on what we aimed and what we obtained from this study. Future works that are feasible towards extending this work are also explained.

CHAPTER 2

LITERATURE SURVEY

The milestones of this thesis work are: 1) intention and trajectory estimation; 2) body-mood detection of a human subject; 3) planning a feasible path for a robot agent to socially interact with the human; 4) reshaping his/her intention into a desired new one. We thus aim through these milestones to establish a natural and sociable interaction between human and robots by planning actions considering both the current intention and the mood of the human agent, just as in a human-human interaction. We further go beyond natural interactions towards intention engineering by generating intention transients from estimated current intention towards the ultimate goal of reshaping: the desired targeted intention. Intention estimation is a detection of "what the human subject intents to do now" by a facing agent, here a robot. The robot, after estimating current intention interacts to the human obstinance on that intention. Thus, readiness to intention change is based on detecting the body-mood of the human agent which is a crucial part of the thesis work because the interacting agent is a robot to which we are not generally familiar in our daily lives. In other words, in order for the human subject complies with the robot interactions; the robot should gain the trust of its interacting agent, making him/her confident (collaborative). In this section, a literature survey of three of the milestones namely, the intention estimation, body-mood detection and intention reshaping are given. Literature survey on our path planning algorithm using elastic networks is detailed in the methodology Section 3.4.1.

2.1. Intention Estimation

Human-Robot Interaction (HRI) is an emerging field both in industry and in our daily lives. Developments on both robot intelligence and hardware capabilities accelerated their usage as human assistants in unstructured environments such as helping the elderly (Dario, Guglielmelli, & Laschi, 2001; Erden & Tomiyama, 2010). However, designing robots in human assistance is a non-trivial problem since human beings cannot be satisfied due to high variability of their intentions and their unpredictable interactions (Koo & Kwon, 2009; Yokoyama & Omori, 2010). In order to have a qualified interaction between human and robot, robots should act much more like human beings, understanding mutual actions and reacting accordingly (Jenkins, Serrano, & Loper, 2007). Therefore, the design process of interacting robots with humans requires prediction of intention for a natural and intelligent cooperation with a human agent.

From the perspective of our thesis topic, Bratman claims (Bratman, 1999) that, one's intention directly affects one's future planning. Estimating the intention of one another between two interacting agents eventually results either in morphing their actions to that of the other's intention by compromises, or in imposing their own intentions onto the other by persuasion (Heinze, 2003; K. A. Tahboub, 2006). Therefore, in order for a robot to induce a desired change on the intention of another agent, which we term intention reshaping that is the main focus of this thesis (Durdu et al., 2011), the robot should begin to estimate the intention of that agent beforehand, and act accordingly.

Researches on intention estimation have been conducted by classifying the problem characteristics relative to the event in question. Durdu (2012) in his work gave an example about estimating a probable fight between two people, in which a good analysis is required by observing and discriminating the actions of the people prior to the fight and leading to it. In other words, intention estimation problem requires first the characterization of all of the actions that the agent underwent prior to the event and their classification related to event based intentions.

The first requirement which is the characterization of actions yielding certain intentions plays a crucial part in their estimations. Social psychologists have been stating for many years that human behavior is goal-directed, that is one's action reflects intentions (Baum, Petrie, Soules, & Weiss, 1970; Bratman, 1999; Dennett, 1989; Lewin, 1952). Most of the actions a human realizes are designed in advance and executed as the plan proceeds. Moreover, some of these actions may become routine which is performed automatically such as driving a car (Ajzen, 1985). If one can become aware of the characteristics of the actions required to attain a certain goal, classification process will be easily handled via Machine Learning (ML). In the literature, action characterization based prediction of the intentions includes mimics, body movements, hand gestures, facial expressions (Adolphs, Tranel, Damasio, & Damasio, 1994; Horstmann, 2003), walking or running (Chouchourelou & Matsuka, 2006; Roether, Omlor, Christensen, & Giese, 2009), pointing (Dittrich, Troscianko, Lea, & D, 1996; Manera, Schouten, Becchio, Bara, & Verfaillie, 2010; Sato, Yamaguchi, & Harashima, 2007) and dancing (Dittrich et al., 1996; V Sevdalis & Keller, 2009; Vassilis Sevdalis & Keller, 2010). In addition, conducted researches in psychology in the field of "self-recognition" portrays the ability of humans recognizing their own acts by kinematic displays of gestures (Daprati, Wriessnegger, & Lacquaniti, 2007), drawing (Knoblich & Prinz, 2001) and body movements (Cutting & Kozlowski, 1977; K. K. Lee & Xu, 2004; Meltzoff, 1995). Although, indicative works are on body movements, and gestures, there are also studies on auditory manners by listening audio outputs resulted from actions (Haggard, Clark, & Kalogeras, 2002; Kohler et al., 2002). Actions based on habituations in the fields of neuroimaging and neurophysiology, such as grasping for eating (Fogassi et al., 2005) and drinking (de Lange, Spronk, Willems, Toni, & Bekkering, 2008) are also characteristic works in intention detection and estimation. To conclude, action types form the characteristics of intentional attitudes of human beings correlated with their environmental contexts. As a result, a good analysis of these characteristics including human orientation, positioning, posture, gestures or facial expressions together with their contextual information results in reasonable estimation of the human intentions. In our case, we utilize orientation and posture of human subjects for the purpose of intention estimation, which are detailed in the upcoming chapters.

Classification of action characteristics towards intention estimation is generally performed using Machine Learning (ML) techniques. Since actions of human or robot agents leading to intentions are sequential processes, the classical sequential supervised learning can be used to construct a classifier which can generate intention classes (Mitchell, 1997). Sequential supervised learning is closely related to time-series analysis and sequence classification (Dietterich, 2002). A good example of describing the need of classification for intention estimation process can be given for the case of using gaze and gestures to control mouse and keyboard inputs on a PC. In order to understand what the person intents to do, actions such as making hand gestures like moving arms for keyboard inputs, gazing different directions or winking eyes for mouse control are generally classified (Ali, Khan, & Imran, 2007; Arai & Mardiyanto, 2011; Qi, Wang, & Huang, 2007; Vlasenko & Wendemuth, 2009).

Intention recognition with computational learning methods was studied earlier as plan recognition problem. In the earlier studies of this area, merging artificial intelligence (AI) and psychology, it was stated that in order to infer the goal of a human, the actions need to be structured into plans (C. Schmidt, Sridharan, & Goodson, 1978). In the work of Schmidt et al., the authors discussed the application of a psychology driven theoretical system called the Believer, which is concerned with single-actor sequences that reports goal-directed actions with the possibility of fail or succeed in achieving the goal (C. Schmidt et al., 1978). The work in (Wilensky, 1983) plans and understands human behaviors in common-sense problem solving and body language understanding based on plan recognition. Later on in 1986, Kautz and Allen approach the problem with a new theory that recognizes a plan resulting from complex actions based on certain hierarchies drawn from the class of possible actions performed, which they called action taxonomy (Kautz & Allen, 1986). By this way, they can link complex sequential actions to a general plan. Another approach to plan recognition was made by modelling uncertainty based on Bayesian probability theory (Charniak & Goldman, 1993). In this work, inferences of an agent's plan were realized by choosing the most likely interpretation for the set of observations using Bayesian updating. In another work using Bayesian networks, the

authors monitored pedestrians' intention in traffic, and developed a system warning drivers for possible accidents (S. Schmidt & Färber, 2009). This work is an example of learning possible plans of humans by observing and classifying their actions from recorded videos of the traffic in question.

There are also significant amount of researches carried out for structuring people interactions during meetings. The work (Dielmann & Renals, 2004) aims at segmenting meetings into actions such as, dialoging between participants, note taking, presenting and doing monologues using audio information gathered from the lapel microphones of each participant. The authors used two-level HMM, which is a Dynamic Bayesian Network (DBN), concatenating acoustic features. In the work of (Zhang, Gatica-Perez, Bengio, & McCowan, 2006), the authors studied recognition of sequences during meetings as in (Dielmann & Renals, 2004) but using both audio and visual features. They used a two-layer HMM with one layer modeling individual actions as in the work of Dielmann et al. (2004) and the second one, modeling people interactions.

Inferring human intentions from actions using collected visual data is another field of intention estimation. In the study given with (T. Mori, Segawa, Shimosaka, & Sato, 2004), the authors focus on the recognition of human actions in daily-life such as sitting, lying, standing etc. They model these actions using a different type of HMM called continuous HMM which recognizes actions from a tree representation. They start from the high level of the model recognizing if the person is sitting, lying or standing. Then, by using parent-child relation of nodes in the tree of hierarchical actions, next level HMM models the action according to its parent. For example, if the first level estimates the action as sitting, current level gives probabilities on whether the person is sitting on the chair or on the floor. The hierarchical representation of actions makes the recognition process easier and more realizable by simplifying estimations at low-level with decreased amount of states, and by classifying the actions at coarse level. The works in (Taketoshi Mori, Shimosaka, Harada, & Sato, 2005; Shimosaka, Mori, Harada, & Sato, 2005) are also based on the recognition of basic human action classes of sitting, running, walking, standing etc.

but the methodology differs using kernel vectors based on switching linear dynamics, which is a probabilistic model generally utilized on tracking and classifying complex motions of human skeleton.

So far, the exemplary studies given are about estimating human intentions while human subjects are not interacting with an object, or are only verbally interacting with another human. The studies below are about estimating intention during a human-machine or robot interaction, which is our focus in the intention estimation part of this thesis. In (K. Tahboub, 2005), the author simulates a scenario with two robots where one robot is controlled by a human operator while the other one estimates operator's intention by observing the actions of the remotely controlled robot. Although this work is a simulated one and does not contain a real physical human-robot interaction, it involves real-time intention estimation which is the case in most human-robot interaction scenarios. The work defines a model of intention recognition based on a four-level decomposition of intentional behavior as in Figure 2.1. The levels on the left hand side in the figure are classical intentional levels realized by people starting from planning a desired state for intended action whereas the other two levels reverse engineered the first two, reaching to a recognized plan or intention. Modeling intention states with action scenarios are realized by using Dynamic Bayesian Networks.
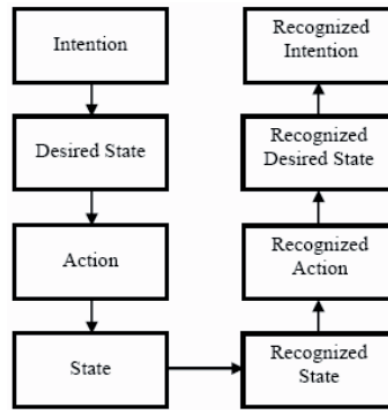
Figure 2.1 Intention recognition scheme decomposed into four-level of intention-action states (K. Tahboub, 2005)

Another work directly from real life is given in (Koo & Kwon, 2009) where the aim is to recognize four basic human actions in the area of public services which are approaching, departing, bypassing and stopping. The authors prepared an experimental environment for human subjects to interact with a mobile robot capable of measuring distances within 360 degree circumference via IR-scanners. K-means clustering method is used to detect human subjects with Extended Kalman Filters measuring their velocity. For the inference of actions traditional HMM observation model is used. Although sensory information provides better performance in real-time application scenarios, the work here is only able to recognize basic actions due to the limitations in tracking people. The authors of the work in (Kelley et al., 2008) however approach the real-time intention estimation problem in HRI systems by using visual data gathered from cameras attached on the interacting robot. Since there is one camera used on the robot, it can only get position information relative to itself, resulting in limited action recognition such as, meeting people, following, passing by, picking up or dropping off an object.

The work of (Z. Wang et al., 2013) proposes a method which models movements aimed at certain intentions probabilistically. The authors used Bayes' theorem in their online algorithm providing real-time intention estimation. To test their system

performance, they create two different HRI scenarios. In the first, one robot and a human opponent play table-tennis which requires fast reactions. In order to generate these reactions on time, a decision making on the type of action (angle, speed) before hitting the ball should be carried out almost spontaneously. This is realized by estimating the target, to where the human intended to send the ball, observing his/her actions. The second scenario reflects the closest work to our aim of intention estimation, in which a humanoid robot is used to detect natural interactions of the human thereafter allowing the robot to react proactively. However, the estimated intentions are again basic actions like, jumping, crouching, kicking-high or low in a table-tennis game. In our study, we design a system where our robot is interacting with the human in their daily lives in a simple living room environment. Literature survey on pioneering works of intention estimation helps to clarify that, the more qualified and realistic the collected sensory data are, the more precise and advanced intention estimations can be realized.

2.2. Emotional Body-Mood Detection with the Aim of External-Focused Attention

Our ultimate goal of this project being to reshape human intention into a desired new one by autonomous robot moves in the cases where people are not familiar with the environment and where no verbal interaction occurs. The critical point in our work is to create the readiness of the person to any intention change. The first point to break obstinance in a prior intention is to catch his/her curiosity and make the person follow these robot moves. The question of how to break this obstinance is, to lead the person to more external-focused attention. This has been studied intensively in social researches carried for groups of human examining the relationship between human mood and attention in social behaviors. Works of (Cunningham, 1988a, 1988b) emphasize the effect of people's positive and negative mood on sociability. In conducted experiments (Cunningham, 1988a), subjects are partitioned into two groups: subjects in one were induced with Velten Mood (positive thought experiment); the other subjects were received depressed mood induction. The

resultant experiments showed that the first group with positive mood indicated significantly greater amount of social behavior than the second. It is stated that, subjects in the depressed mood preferred thinking while sitting or napping alone rather than being in favor of socialization.

Similar to the works of Cunningham (1998a,1998b), the work in (Wood, Saltzberg, & Goldsamt, 1990) studies the effects of the mood on self-focused attentions. In this work Wood et al. (1990) state that inducers for self-focused attention serve as a depression state is reveals a correlation with depression to be a negative mood. It was tested that sad mood induced self-focused attention while happy or neutral moods did not. The work in (Watson & Tellegen, 1985) explains the positive and negative moods mentioned. According to the authors of this work, positive mood states can be exampled as happy, cheerful, confident and relaxed whereas the negative ones are anger, anxiety, suspicion and sadness. In our work, we generalized and named the negative mood states as *suspicious* and positive mood states as *confident*.

The work inspiring our study the most from the perspective of mood effects to attention, was the work in (Sedikides, 1992) which examined not only the inducers of self-focused attention but also the moods resulting in an external-focused attention. As we mentioned, our aim is to induce an external-focused attention on the human subject. In that study, the author aims to prove that the assumption of positive moods may lead to external-focused attention which is logically the exact opposite of the relation between negative moods and self-focused attention. Sedikides (1992) partitioned three groups of subjects which are to be conditioned by one of three moods (sad, neutral and happy). After inducing these moods on the subjects by asking them to recall related emotional events, the author wanted them to write a story either about themselves or someone else they know well. The classification was done according to the story clause and if it is about self, the subject is labeled as self-focused or else external-focused. The experiments showed that comparing neutral and happy moods, sad mood induced a great deal of self-focused attention. In addition, the author demonstrated a new fact that positive mood (happy in that study) resulted in external-focused attention more than the other two moods. Similarly, in

the works (Fredrickson, 2003; Grol et al., 2013; Wadlinger & Isaacowitz, 2006), it is stated that positive moods or emotional states broadens perceptions of visual attention when compared to neutral or negative emotions.

These findings guided us in the balance of this thesis to break the self-focused attention leading to obstinate intention in human subjects and direct their focus towards our robots which is an example of external-focused attention. If the person has a negative mood, named as *suspicious* about the robot, we first aim to induce a positive mood on that person, which is in our case *confident* mood. We claim that a confident person can be more externally-focused triggering curiosity about our robot moves. These moods of confident and suspicious are inspired from the work in (S. Lee & Son, 2008) where the authors simulate an emergency situation in a crowded scenario and create a Belief-Desired-Intention (BDI) model for humans who are classified according to their confidences in the environment.

With the perspective of emotional mood detection, Cowie and Douglas-Cowie (1995) state that humans communicate through two channels, being explicit and implicit. Messages via the explicit channel are physically detectable messages such as; speaking, gesturing, moving, posing etc. and are largely researched under the topic of intention detection from actions, as discussed earlier. The implicit one is described by the authors as the task defining how to receive the messages transmitted through explicit channel (Cowie et al., 2001). In our daily interactions, we use same words or exhibit same actions that may mean differently according to the implicit emphasizes on words or emotions hidden in these actions. Likewise, we make sense of actions of the other agent by detecting or predicting emotions of the speaker. Due to the fact that human interactions includes these implicit channels, researches in the field of HRI try to develop robots capable of modeling these social cognitions in order to generate proactive interactions with humans satisfying implicit social aims such as: emotions, intentions, etc. (Fong et al., 2003). Fong et al. (2003) called these robots "socially interactive robots" which can basically detect emotions, recognize models of human agents and develop social relationships. In the literature, usage of these robots are introduced in areas where either robots are used for persuasion,

16

changing the feelings, attitudes or behaviors of the interacting human agents (Fogg, 1999, 2002), where either robots act as mediators teaching social interaction skills in autism (Werry, Dautenhahn, Ogden, & Harwin, 2001) or robots are themselves an avatar representing a human agent (Paulos & Canny, 1998).

Here the question is asked on how the robots can model emotions or implicit messages of the human agents. Since 1872, starting with Darwin's famous research (Darwin, 1872), there are a lot of researches conducted in psychology on behavioral expressions of emotions in humans. Apart from researches on emotions with facial expressions (Beaudry, Roy-Charland, Perron, Cormier, & Tapp, 2013; Keltner, Ekman, Gonzaga, & Beer, 1993; Reisenzein, Studtmann, & Horstmann, 2013; S. Wang, Liu, Lv, Lv, & Wu, 2010; Wehrle & Kaiser, 2000) and emotions with vocal characteristics (Banse & Scherer, 1996; Jessen & Kotz, 2011; Juslin & Laukka, 2003; Sauter, Panattoni, & Happé, 2013; Scherer, 2003), there are also researches on bodily expressions of emotions which mostly interest our study pertaining to observing human actions with a camera in a non-verbal situations. In the work of (Wallbott, 1998), the question of whether body gestures, movements, postures or quality of movements reflect human emotions is detailed starting with the work of Darwin (1872) which is also concerned with emotions exhibiting bodily movements. Wallbott (1998) collected all of the emotion models of Darwin based on certain posture and movements pattern within a table such as: motionless, passive actions yielding the emotion of sadness; purposeless movements, clapping hands, jumping and dancing yield the emotion of joy; head and body held erected yielding the emotion of pride etc. (Walbott, 1998, pp.880). Although in the work of (Ekman & Friesen, 1974) it is stated that bodily-movements only give the intensity of the emotion and that there is not any specific body movement or gesture for an emotion, there are researches indicating that emotional state of a person may influence his/her bodily-movements (Chouchourelou & Matsuka, 2006; Hatfield, Cacioppo, & Rapson, 1993).

There are few works on computational emotion modeling based on visual data, where most are emotion detection from facial expressions. In the work of (Breazeal,

2003), a humanoid robot was developed emulating a human face and having the capability to learn human facial emotions. This robot learning these facial emotions can apply these expressions while interacting with a human subject. In the experiments, the robot called 'Kismet' can adapt its facial emotional mood to the one of its interacting human subject. Another work detecting mood from facial expressions and postures is given in (Wada, Shibata, Saito, & Tanie, 2004) with a robot called 'seal robot' aiming to interact with elderly people to overcome their stresses and improve their feelings. Similarly, the work in (Kozima, Nakagawa, & Yasuda, 2005) develops a creature-like robot detecting emotions from facial expressions of autistic people used in therapy. Therapy through emotional moods are also handled in (K Dautenhahn & Werry, 2000; Mazzei et al., 2011; Robins, Dautenhahn, Boekhorst, & Billard, 2005; Vanderborght, Simut, & Saldien, 2012; Werry, Dautenhahn, & Harwin, 2001; Werry, Dautenhahn, Ogden, et al., 2001) each using differently shaped robots with different interaction capabilities.

A study on emotion detection system using speech data is given in (Scheutz, Schermerhorn, & Kramer, 2006). The authors use speech emotion filter introduced in (Burkhardt & Sendlmeier, 2000) to synthesize the affective state on the speech of the humans. In this work, the robot is able to detect emotions of sadness, anger, fright and happiness from the speech of its interacting human and react accordingly by speaking in the tones of these given emotions. For example, if the detected tone is fright, the volume of the voice of the robot will be higher with theatrical pitch swings.

The work in (Barakova & Lourens, 2010) models emotional movements during games with robot companions authors used Laban movement analysis to characterize human motions as emotional states. Laban movement analysis is a method to observe and describe human bodily movements categorized as strength, directness, speed and flow of motion. An example given in the study is to analyze the difference between punching a person and reaching a glass. Here, the strength and the speed of the former movement are clearly higher than the latter one. By using these categories of the human body, the authors are able to project body-movements of children during a

game with a social robot into four basic emotions which are sadness, fear, anger and joy.
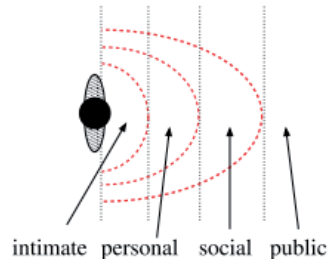
Figure 2.2 Spatial interaction zones in HRI through a corridor (Pacchierotti, Christensen, & Jensfelt, 2005)

In our study, unlike most of the works using off-line body movement detection stated above, we are detecting on-line emotional feelings (mood in real-time) of the subjects about our robots from human headings (direction of motion). That is, we are measuring the confidence and trust of the person to our robots by using only visual data tracking human from a ceiling camera. Our work utilizes a psychological work called '*proxemics*' studying the spatial interactions between humans (Hall, Birdwhistell, & Bock, 1968). Study on "proxemics" is well examined in depth within the perspective of HRI in the works of (Christensen et al., 2005; Pacchierotti et al., 2005) by dividing spatial distance around the person into four categories of interactions which are intimate, personal, social and public as given in Figure 2.2. The authors examined the same spatial relation between a robot and human subjects in a hallway passage where an interaction is unavoidable and observed that proxemics is valid in HRI and robots should be aware of these relations. Additionally, in the work (Butler & Agah, 2001), the authors examined human moods when a robot is approaching a subject. In that work, it is stated that if this approach is slow and direct, the observed mood of the human is comfortable.

Similarly, in the more recent works in HRI (Huettenrauch et al., 2006; Mead & Matarić, 2011; Takayama & Pantofaru, 2009; Walters, Oskoei, Syrdal, & Dautenhahn, 2011), the authors studied proxemics behaviors by assuming that a human approaching a robot has to be comfortable with it.

By putting all those studies together, we designed a novel on-line mood detection system inferring from studies of proxemics behaviors stating that, a person heading towards a robot allowing it to enter his/her intimate region is detected to be comfortable with the robot (we named *confident*). On the other hand, a person showing no interest in our robot even after robot moves towards him/her is detected to be in an uncomfortable mood (*suspicious mood*). According to the studies of relation between attention and mood, we contribute a novel conclusion that a confident person (positive mood) gives more attention to our robot (external-focused attention) making him ready for an interaction. The application of mood detection system and the statement of "confident person gives more attention to the robots" are experimented in real-time, details of which are given in Sections 3.3 and 4.1.2.

2.3. Inducing Emotional Mood on a Human through Intention Reshaping

As we discussed, we should detect confident or comfortable mood (in general positive mood) in the person to establish the readiness to reshaping his/her intention. Therefore our first path for the robot is planned to induce this positive mood on the person making him confident with our robots. After a confident mood is detected on the person, the next path will be to reshape the intention of the confident person into a desired one.

2.3.1. Inducing Emotional Mood on a Human

For the purpose of inducing confident mood (positive mood in general), our robots should socially interact with the person in a trustful and comforting manner. In the literature, there are a few works covering the related topic. In the work of (Suzuki,

Camurri, Ferrentino, & Hashimoto, 1998), the authors clearly showed the effect of approaching a person on the induced emotional state. The factors affecting the emotional states are the velocity of the robot, the distance between the robot and the human and the gestures of the robot. The most influential work on our study is given in (Butler & Agah, 2001). In this study, the authors examined human-robot interactions with different robot behavior patterns such as: the robot approaches the person and trans-passes the person. They analyzed the level of human mood (in this case mood is measured as the level of comfort) during the interaction experiments based on the speed, the distance and the design of the robot where the experimenters were not informed about the robots. According to the survey results from the human subjects in the case where the robot approaches them, a slow and direct approach resulted in more comfortable and attention taking, which is exactly what we desired. In addition, parallel results were found in more recent studies (Huettenrauch et al., 2006; Mead & Matarić, 2011; Takayama & Pantofaru, 2009). However, none of these works approached the problem from the robot's perspective, that is, they only considered psychological changes of the human subjects towards different robot moves. After examining the results of these studies, in our novel approach, we plan paths for the robots according to the currently detected human intention. Robots' following these paths slowly is expected to increase confidence of the human subjects and gain their curiosity afterwards.

2.3.2. Intention Reshaping

Intention reshaping, being the final stage of our study, aims to change the current human intention into a desired new intention.

In the literature, some exemplary works are introduced in psychology mainly examining the effects of intention change on behavioral change (Webb & Sheeran, 2006) and intentional change in children under the effect of what adults do (Carpenter, Call, & Tomasello, 2005; Meltzoff, 1995). In the field of HRI, these behavioral or intentional changes are examined according to the different types of robotic interactions with human subjects. The study in (E. Wang, Lignos, Vatsal, &

Scassellati, 2006) surveyed the subjects intentional behaviors in accordance with the variations of head movements of a humanoid robot such as: motionless, human face tracking smoothly, tracking fast and turning away from the human face. The survey results showed that the subjects stated different perceptions and behaviors for these head motions. That is, changing the head movements induced notable changes in intentional behavior of humans.

The works in (Terada & Ito, 2010; Terada et al., 2007) studied how the intention attribution is affected by different artifacts (non-humanoid robots). In this study, the authors used chair and cube as robots and observed the attributed intentions of human subjects to the movements of these robots. The results showed that, the human subjects attributed different intentions to these reactive movements depending on the shape of the artifact, and the perceived goal of the artifacts by the subjects. Similarly, the work in (Parlangeli, Guidi, & Caratozzolo, 2013) concluded that, the attribution of mental states or intentions to the mechanical structures or artifacts is affected by personal and contextual differences in human-robot interactions.

Intention reshaping idea was first introduced in our previous work in (Durdu et al., 2011). In this study, the aim was to observe the intentional changes on the human subjects after the pro-active and contextual robot movements of non-humanoid robots (robot-like chair and 2-steps). Comparatively using HMM and Observable Operator Model (OOM), estimated intentions before and after the robot movements showed intentional changes in the human subjects in real-scenarios. For example, a human preparing coffee in the experimental room, is distracted by the movement of 2-steps robot in front of the library, and changed his/her intention to take a book from a library. This is a clear example of intention attribution to the 2-steps artifact robot that reshapes accordingly the intention. However, in this work the robots are commanded from outside with a joystick and intention estimation and intention comparison were realized off-line, that is after the videos were recorded.

Our main contribution in this study is creating a closed loop system which reshapes previously recognized human intention into a desired one by fully autonomous

22

sociable robot moves planned by generating intention transients through Elastic Networks in real-time experiments. Literature survey on elastic networks for path planning is given in Section 3.4.1. In our novel approach, we first detect body-mood of the human subject by observing his/her heading whether it is towards the robot or not. Any failure yields to *suspicious mood,* planning a path headed towards the location of the estimated current intention location by HMM at this phase of the experiment. Here another novelty is to gain the curiosity of the human by breaking his/her intentional obstinance. The subject is expected to approach our robot upon the detection of *confident mood* in the person, making him/her ready to start an interaction and accept intentional changes. Chapter 3 gives detailed methodology of our approach and the results of the entire close loop system in on-line experiments are given in Section 4.1.2.

CHAPTER 3

METHODOLOGY

Our system for reshaping a current intention into a desired one by robot triggered interactions is illustrated as a block diagram in Figure 3.1. As seen in the figure, the ultimate goal is to generate a path, or a trajectory, for the robot to follow until the person realizes the desired intention, given to the algorithm beforehand. Each trajectory is generated from a feature space learned by previously recognized intention trajectories of humans in training sessions. These trajectories designate suitable robot moves in adequate directions and each one is designated by the intention they are driven with. These intentions are transient ones, called a way point towards the desired intention. Transient intentions are planned through way points calculated using Elastic Networks in two modes: "*confident*" and "*suspicious*". These modes are depicting currently detected body-moods of human subjects. The reason of the introduction of these modes can be explained as such: we should note that, our first aim is to attract the attention of human subjects by our robots. We try to gain their curiosity and make them confident with the environment by suitable robot moves. The research works in the literature revealing that a person in a confident mood elicits more external-focused attention than a suspicious one (Cunningham, 1988a, 1988b; Sedikides, 1992; Wood et al., 1990) directly support our attempt to break any obstinance with a current intention. Our robots manage to achieve the aim of making the human confident by getting close to the subjects while mimicking their previously observed actions (Kerstin Dautenhahn, 1999). That way, the human subject can carry out more external-focused attention (Sedikides, 1992) and be curious about the robot.

We gave the name of "*execution modes*" to the detected two different body-moods of the person. According to these modes, elastic network works with different

parameters to generate way points. If the person is in suspicious mood, firstly the robot should break the obstinance of the person on what he is intending to do and gain his curiosity to create a readiness to intentional change. On the other hand, if the person is confident, it means the robot successfully caught the attention of the person on itself; hence, elastic networks easily generates for a way point (transient intention) starting from the current intention towards the desired intention. The mechanism behind body-mood detection is based on the reactions of the human subject to the robots which is further detailed in the upcoming sections.

This chapter begins by introducing our experimental setup. We then provide the human/robot detecting and tracking approaches explained through demonstrative examples. After that, we proceed by detailing the generation of intention space called feature space by learning intentional actions of human subjects in training trials with the same experimental set up. Intention estimation and body-mood detection are explained next in sections 3.2 and 3.3.

Transient intentions are generated based on the estimated current intention, observed previous intention before the current one and detected body-moods using Elastic Networks. We first overview the classical methodology and then explain our adapted elastic nets to generate transient intentions (way points) within the intention space (feature space).
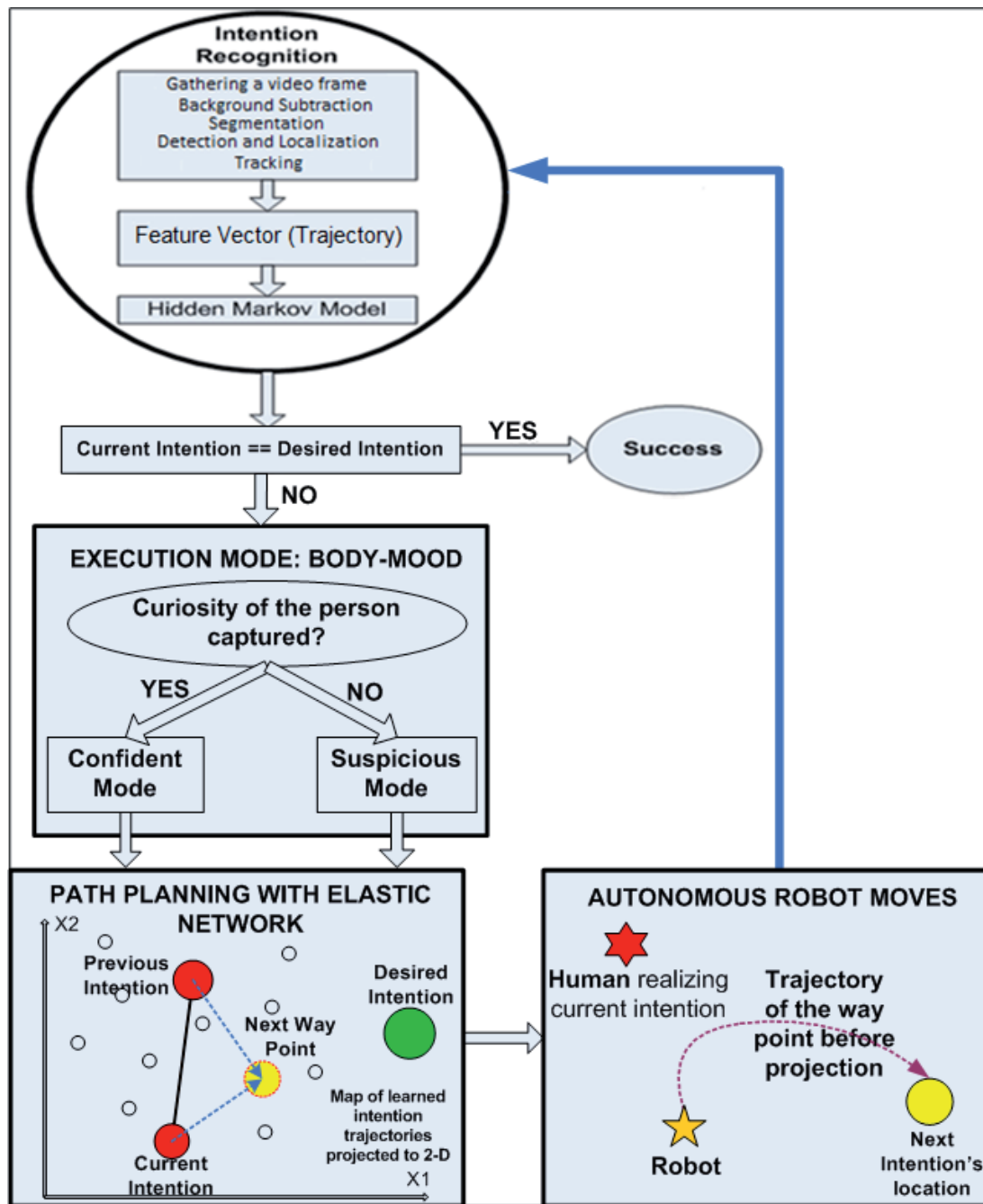
Figure 3.1 Flow chart of the proposed methodology

3.1. Intention Space (Feature Space) Extraction

3.1.1. Experimental Setup

In order to realize and test our progress we prepared an experimental room and used two robots adapting the idea from the previous work of (Durdu, Erkmen, Erkmen, & Yilmaz, 2012). The room is equipped with a coffee table, a working table with a PC and a library. In addition, to interact with our human subjects we have two contextual mobile robots which are a mobile 2-steps and a mobile chair both autonomous. All of the objects that a person entering the room can interact are shown in Figure 3.2(a). For the sake of proving the intention reshaping concept, we purposefully installed minimal amount of objects that demonstrates distinct independent intentions. All of the possible actions that a human subject can undergo in the room are listed in Table 3.1 against the intention they can carry.

Table 3.1 List of observable actions and their labeled intentions in our application

#	Observable Actions	Labeled Intention	Short Versions
A1	Discovering the environment	Discovering	Discovering
A2	Going to the Coffee Table	Drinking Coffee	Coffee
A3	Going to the Library	Getting book from Library	Library
A4	Going to the Work Table	Sitting on the Table	Table
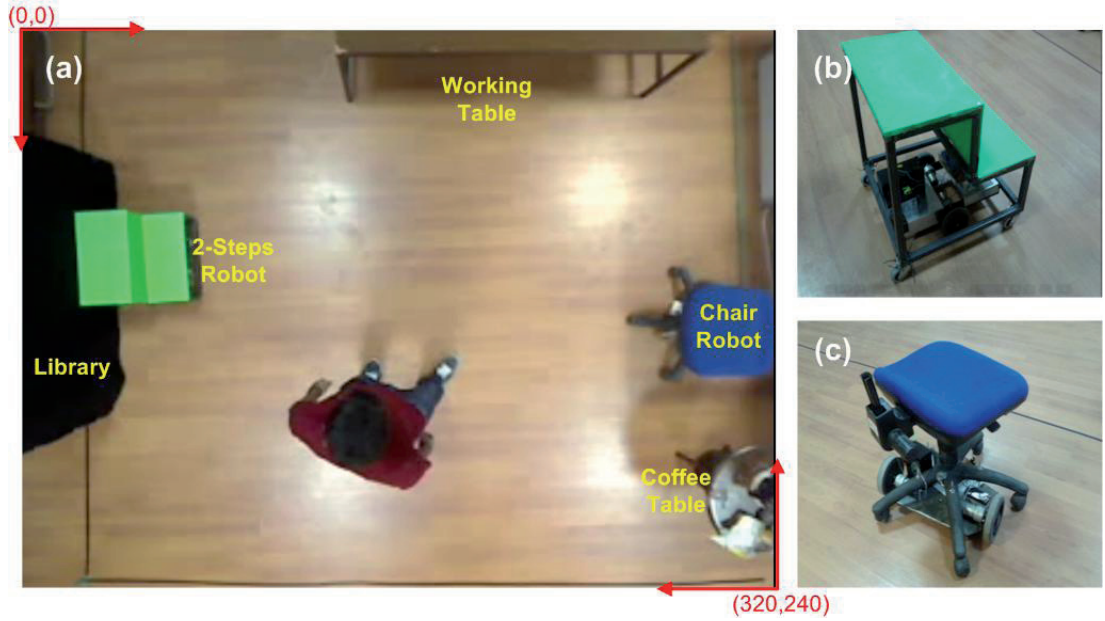
Figure 3.2 (a) Objects in the experimental room seen from the ceiling camera; (b) 2-steps robot; (c) Chair robot.

The robots, 2-steps and chair in Figure 3.2(b) and (c) respectively, can do planar motions on two wheels each driven by 12V DC motors. These motors are controlled by PIC 16F84 microprocessor upon the received control signals from a PC through RF. Robots have a load capacity for mobility which is 10 kg. Detailed technical specifications on the hardware of the robots can be found in (Durdu, 2012).

A camera placed at the center of the ceiling of the room tracks human and robots and their trajectories. Region of interest (ROI) of the camera is 240x320 with a frame rate of 15 fps (frame per seconds) for the training experiments. Here, we should note that we had two phases in our experiments and the frame rate mentioned was for the training part. The real-time tests in intention reshaping phase has a frame rate of 8 fps approximately (detailed in Section 4.1.1). Every human and robot actions either learned to generate the intention space or used to reshape a current intention is based on the tracking of human and robots. Tracked trajectories of human define their carrying intention. Detecting and tracking of human and robot that lie at the heart of

our system are detailed in section 3.1.2. Tracking is also used to estimate the current intention (section 3.2) and the heading and localization are used for determining the body-mood of the person (section 3.3), both estimated after processing the received frames from the camera. The main computer plans transient intentions using elastic networks and sends appropriate directional commands according to these transients to the robots via radio signals in order to realize adequate trajectories.

3.1.2. Human/Robot Detection, Localization and Tracking

As mentioned in section 3.1.1, there are three mobile objects to be detected, identified and tracked, which are a human subject, chair robot and 2-steps robot. In each frame as a feature vector, the locations of all of these three objects are to be determined. Feature extraction is one the most crucial component of our study. A false localization of human subject might have led to a false estimation of intention resulting in irrelevant path planning for the robots. Therefore, in this section we give brief analysis of the parts through the feature extraction given in Figure 3.3 including problems encountered and some tricks to overcome these problems.
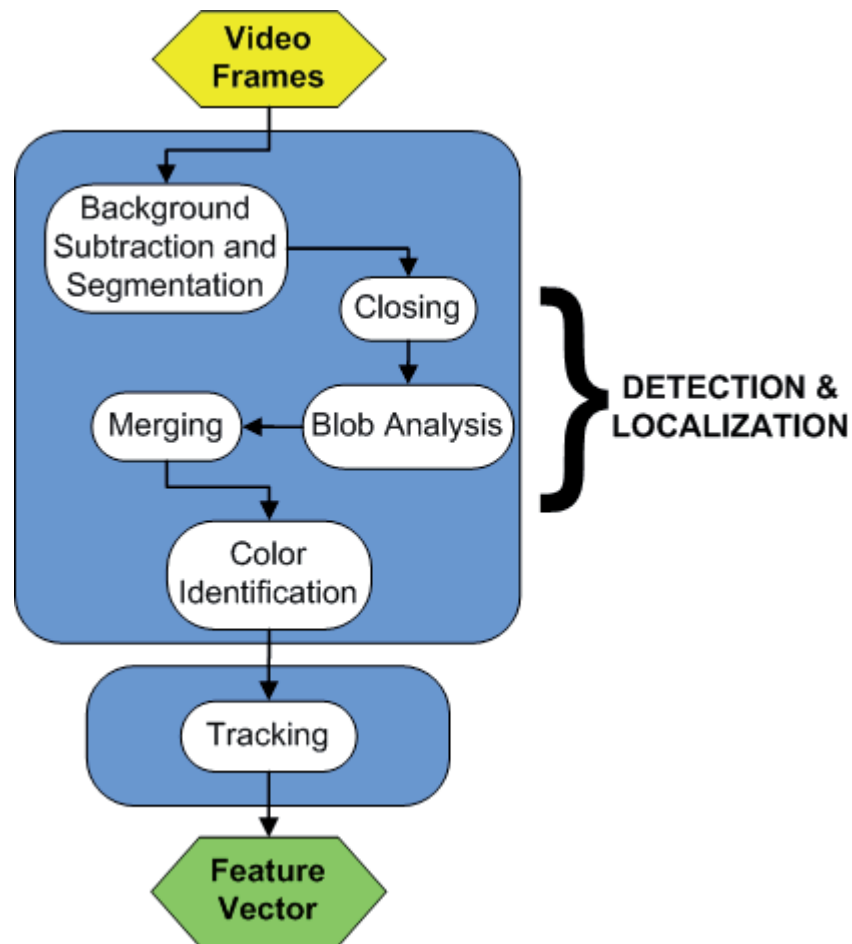
Figure 3.3 Block diagram for feature extraction system

3.1.2.1.   Object Detection and Localization

Image processing steps starts with background subtraction which is the major component of computer vision systems utilized for detecting moving objects using a stationary camera. A stationary background is subtracted to detect moving pixels by taking the difference between an input image, being current frame gathered from the camera, and a background image. Then, by thresholding the difference, pixels corresponding to moving objects in the scene are located.

Since the illumination in the environment can change between frames, defining a stationary background is not a trivial task in image processing. In the literature, it is usually preferred to use a method modeling distribution of intensity value of each pixel. In this study, we used a Gaussian Mixture Model on the first few frames to estimate a mean for the background pixel values as in (Stauffer & Grimson, 1999). For the image thresholding, we used the most common method called Otsu's Thresholding Method. It is basically an auto-thresholding method performing clustering-based image thresholding. Using these well-known segmentation methods mentioned, subtracted background can be seen in Figure 3.4(c) with the original image (Figure 3.4(b)) and the stationary background image (Figure 3.4(a)) from our applications.
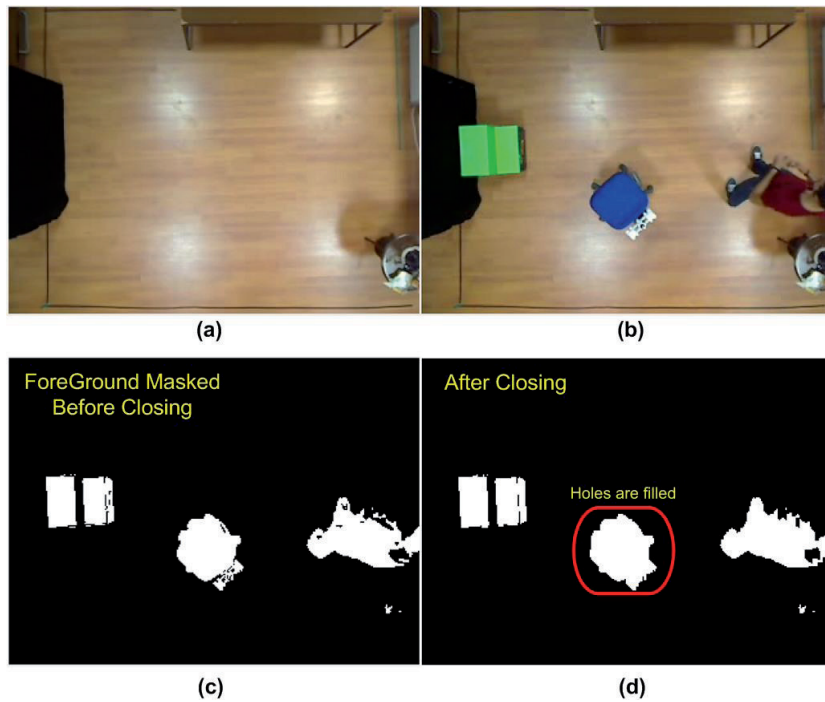
Figure 3.4 (a) Modeled background; (b) An input image; (c) Foreground masked view of the image in part (b).

After segmentation, we utilize a famous morphological operation called "Closing" (Soille, 2003). The aim is to fill holes between object pixels that are close to each other. That way, blobs are created for each object in the scene. As the structural element we use a rectangle with sizes 7x1 pixels. An exemplary image before and after "closing" operation is given in Figure 3.4(c) and (d). After "closing", holes on surface the object images in the foreground masked view are filled, which is a necessary step to define a blob around the objects. Furthermore, the resultant image after "closing" is scanned with blob analysis algorithm adapted from (Chen, Lin, & Chen, 2007). The obtained blobs are analyzed with respect to statistical values which are: a bounding box around these blobs, centroids of these boxes, label matrices and total amount of blobs in the image. Exemplary blobs detected are demonstrated in Figure 3.5(a). The green rectangle bounding box indicates a blob belonging to the 2-steps robot (originally seen in Figure 3.4(b)) and the blue rectangle is for human sitting on the chair robot.
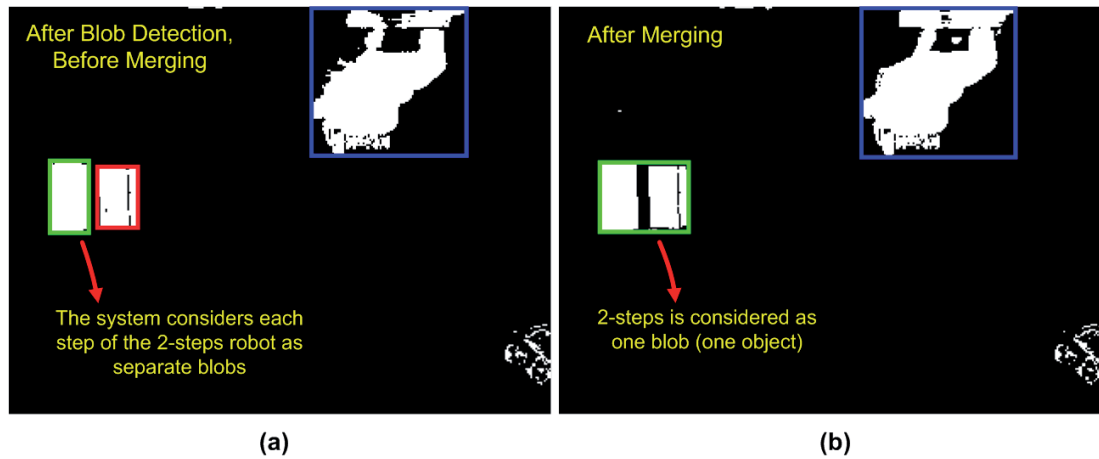
Figure 3.5 (a) Snapshot was taken right after blob analysis. The blob detected around the 2-steps robot is a green rectangle which considers each step of the 2-steps robot as a different object; (b) After the "merging" algorithm, 2-steps robot could be fully detected as one object.

After the blob analysis algorithm extracted the centroid and bounding box coordinates of the blobs in the image, we used "merging" method to merge the blobs belonging to the same target. As demonstrated in Figure 3.5(a), each step of the 2-steps robot is detected separately because of the dark region separating steps as seen in an original view in Figure 3.4(b). Merging algorithm takes the rectangle coordinates of the bounding boxes of all blobs detected and measures the distances between each. If the distance between two blobs is smaller than a threshold and this value stills the same for the consequent frames, they are considered to be the same target a merged to be one blob (see Figure 3.5(b)).

During the applications we encountered many problems on detection due to luminance, angle of the camera, shadows of the objects. To overcome, we apply parameter changes while testing the performance of the system. Although detection and localization is not the main focus of this thesis, we provide solutions offered to each problem briefly:

- *Merging of the blobs belonging to different objects:* When two objects, especially human with robots, got close enough to each other, they were claimed to be one blob detected by the merging algorithm causing loss of actual locations of both agents. As for the problem, we adjusted the box merging threshold according to our experimental room.

- *Misdetections:* In some experiments, misdetections occur especially on the coffee table as seen in the right bottom corner in Figure 3.6(b). This is due to the fact that while human subjects were preparing coffee, they misplaced the coffee machine, coffee cup, sugar can or even hit the coffee table which is light enough to move. Since these materials are considered to be background at the beginning, these misplacements cause ghost formations big enough to be detected as a blob. Increasing the minimum area for a bounding box of the blob detected seems to be a good solution; however, this time we wouldn't be able to detect even real objects of our concern. For example, each steps of the 2-steps robot are sometimes detected as separate blobs at first (see Figure 3.6(b)). Then they are merged as mentioned. Any further increase in the limit

caused ignoring of these two steps. Instead, we developed a new algorithm at the end of the blob merging algorithm which was the last step of our detection process. This algorithm takes all of the detected blobs in the image (along with the coffee table misplacements), calculates the areas of their bounding boxes and outputs the biggest three blobs. As seen in Figure 3.6(b), we have three moving objects to be detected and they cover by far the biggest three areas comparing the misdetection of the coffee table.

- *Shadows of the human subjects:* When human subjects are in certain locations, their shadows become separated from the background and fall on to robots due to the non-homogenous lighting in the room (See Figure 3.6(b)). As demonstrated in Figure 3.6(b), the biggest problem caused by this is to merge the two blobs: human subject together with the shadow and robot, although the human and robot are far away. In these situations, they are claimed to be one blob and one centroid calculated for both of the objects. That causes incorrect positioning for them. Instead of detecting one blob and ending up with a wrong data, we preferred no detection at all and preserved the previous locations of the bounding box of two objects detected correctly right before shadow merging problem. We managed to achieve this by decreasing the maximum blob area to be detected in the blob analysis algorithm.
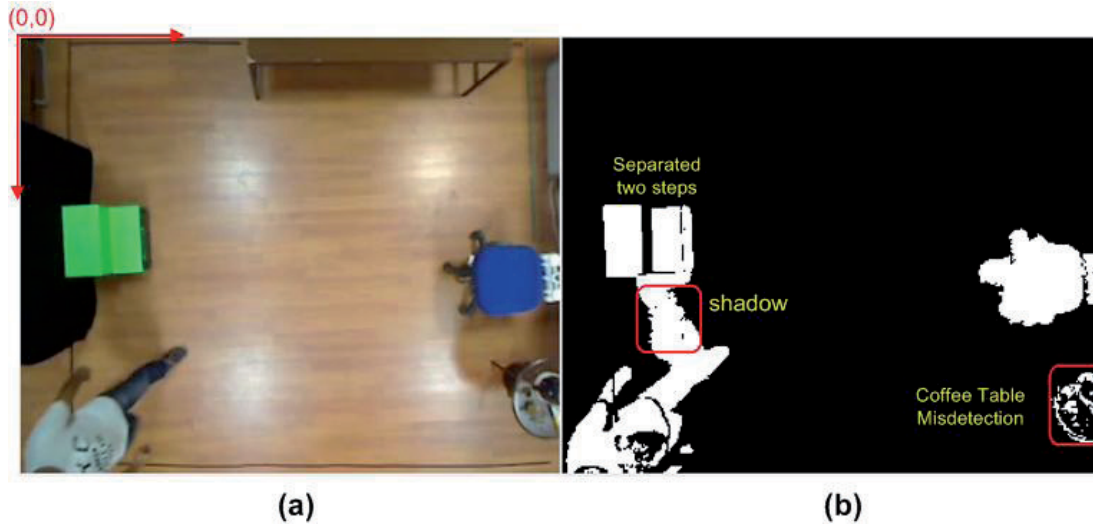
Figure 3.6 (a) Original image with origin of the pixel coordinate system being at the top left corner; (b) Foreground masked view with the mentioned problems of: shadow causing false merge of blobs, misdetection on the coffee table, separated two blobs belonging to the same object, 2-steps robot.

Along with the problems mentioned, there were still some bottlenecks in the detection. The most important one is that we couldn't do anything when the objects actually overlapped as seen in Figure 3.8(b). For those kind of cases happened a lot during the experiments, we ignore the merged blobs as we did in shadow problem and preserve the previous bounding boxes of the overlapped objects as demonstrated in Figure 3.8(b). We set a maximum area limit for the bounding boxes such that, ignoring merged blobs of two different objects continues until a near full overlapping achieved (e.g. the person almost sits on the chair robot or the person almost climbs to 2-steps robot) as in Figure 3.8(d). In this figure, detected area of two merging objects got below the maximum blob area limit and a bounding box is assigned. However, there are two problems to be solved, what the label of this merged blob is (is it assigned to the robot, to human or both) and what happened the location data between Figure 3.8(b) and Figure 3.8(d). These problems are covered by tracking detailed in the next section.

Color Identification: As stated in Figure 3.3, we have color identification for the last step of the detection and localization process. To identify the object detected as a blob with a bounding box, we check the RGB color values of each centroid pixel of the bounding boxes. This step is also necessary to track each object separately. To start with, Figure 3.6(a) clearly indicates that we have different colored robots in the scene. The methodology is to check the color value of the pixel on the centroid coordinates outputted from detection algorithm. Input and output centroid values of the color identification algorithm can be seen in Figure 3.7(a) and (b) respectively. In part (a), a random labeling with bounding boxes around the blobs is seen. On the other hand, in part (b), each object was labeled and localized correctly where green bounding box and initial 'S' for 2-steps robot, blue box and initial 'C' for chair robot and always red colored box with initial 'H' for human subject.

As for the color identification process, the only critical point was the choice of the colors of the robots. We decided on two of the main colors in image processing, namely green and blue (two of RGB) for the robots. In other words, chosen colors and their default RGB values are highly separated from each other. After labeling two robots, the remained last centroid was claimed to be belong to the human subject. That way, human subject labeling is made independent from the colors s/he wears (colors except the green and blue tones are preferred). However, even if the human subject wears green or blue, once a correct labeling was achieved, separate target tracking algorithms of each object ignores false labeling occur during the system works.
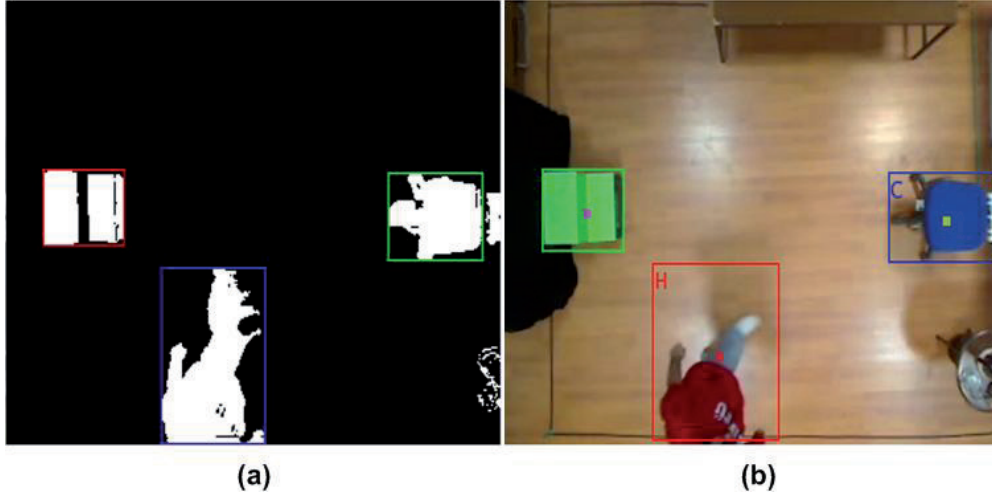
Figure 3.7 (a) Random labeling after detection algorithm before color identification with foreground masked view; (b) Corrected labeling can be seen from the initials of the objects at the top-left corner of the bounding boxes of the blobs. 'H' stands for human and always in color red whereas 'S' for 2-steps robot with green color and 'C' for chair robot with blue color.

3.1.2.2. Object Tracking

For the last but the most important part of our feature extraction system, we used tracking algorithm. A tracking algorithm is needed as the most crucial part of a localization problem. We implemented three separate tracking algorithms with three different tracking thresholds for three different moving objects detected and labeled in Section 3.1.2.1. As in conventional tracking systems, Kalman filter was used for its proper performance on estimating past, present and future states by using series of measurements containing noise and other nonlinearities (Welch & Bishop, 1995).

To start with, a feasibility check to ensure the correctness of color identification is handled by a comparison algorithm beforehand. That is, we compare detected locations of each object with their previous locations to check if the new labeling is feasible relying on the maximum travel distance can each object realize between two

consecutive frames. For the purpose, we calculated a target tracking threshold value for each object by taking their maximum speed into account. That is, we assume that a human can move 20 pixels whereas a robot can shift 10 pixels between consecutive frames. If the feasibility check of color labeling fails for an object, previous coordinates of this object are forwarded to Kalman filter.

Kalman filter takes the locations of the bounding boxes of each object detected in the previous frames, measures the effect of noise and predicts the locations of these boxes in the current frame. This is called state estimation of a dynamic system and it is the first mode of our Kalman filter algorithm. As for the second mode, within the same sampling time, Kalman filter enhances its current state estimation by using the predicted state and the detected current state. In our applications, these states are the coordinate values of left-hand corner of the bounding boxes surrounding each object. For the final application of the filter, it reduces the effect of noise in the detected coordinates (Welch & Bishop, 1995).

We use separate tracking algorithms for each object. The point of this usage is not to lose any location information of the objects especially when the two of them overlap as in the given scenario with Figure 3.8. In this figure, the left hand side of each part is a snapshot of the intensity image of that frame before color identification and tracking algorithms, whereas the right-hand side is the resultant snapshot of all processes mentioned throughout Section 3.1.2. If a merging occurs due to an overlap situation, our detection algorithms do not assign a blob around this overlapped image. However, as soon as the overlapped region gets lower than the maximum blob area limit, it is detected as a blob shown with yellow bounding box in the snapshots on the left-hand of Figure 3.8(c) and (d). Then, the color identification algorithm labels this blob with human, since the human is mobile during an interaction with a robot. As will be detailed in Section 3.5, during an interaction (as the human sits on the chair or climbs onto the 2-steps robot), the robots are not allowed to move due to safety reasons. Therefore, overlapping situations are considered to be interaction moments and robots actually stand still until the human finishes the interaction.

In Figure 3.8(a), blobs of the chair robot and the human subject started merging and Figure 3.8(b) shows the prevention of the blob detection of these merged objects. In the part (b), intensity image shows the ignored blob (there is no bounding box) where the colored image (after tracking algorithm) still preserves the last locations of both the human and the chair. As the human keeps approaching to the chair, merged blob area decreases below the maximum blob area limit resulting the detection of the blob again as seen in the intensity image of Figure 3.8(c). The colored image of this figure shows that, the detected blob is labeled as human because the bounding box of the human started moving towards newly detected box estimating the path the human covered between the last detection in Figure 3.8(b) and Figure 3.8(c). Figure 3.8(d) and Figure 3.8(e) show the full settlement of the bounding box of the human subject based on the tracking algorithm. Estimated path by the Kalman filter is given in Figure 3.8(f) with indicated starting and finishing points. This is a satisfactory performance for our localization process.

After detection, localization and tracking, a $[3x2]$ matrix holding centroid pixel coordinates of the bounding boxes of each object is resulted in each frame, as demonstrated in the equation (3.1). The initials '*C, S, H*' are for the chair robot, 2-steps robot and the human respectively. Feature descriptors of our system utilize these coordinates which are to be used in Hidden Markov Model for learning and estimation (Section 3.2), execution mode decision (Section 3.3) and Elastic networks (Section 3.4).

$$\begin{bmatrix} C_x & C_y \\ S_x & S_x \\ H_x & H_y \end{bmatrix}_{3x2} \tag{3.1}$$
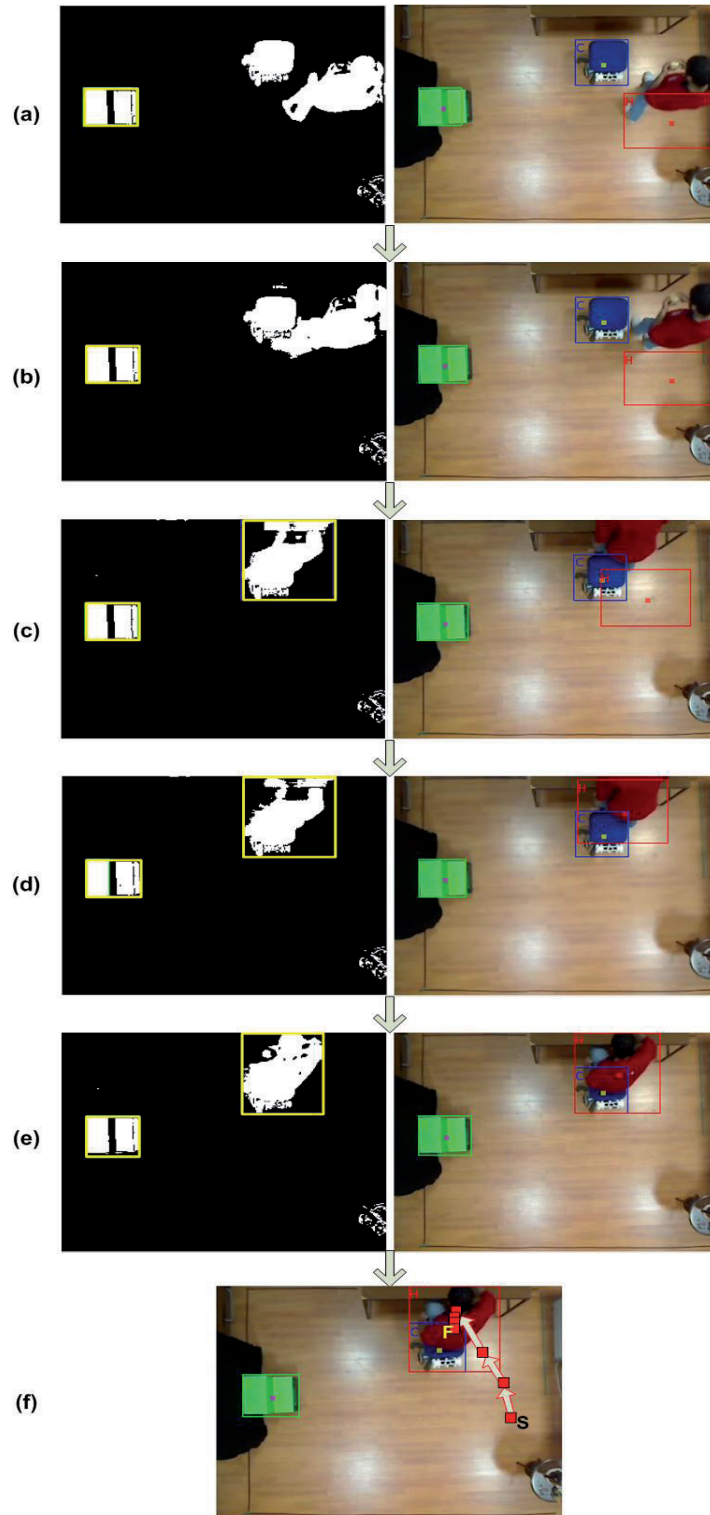
Figure 3.8 (a)-(e) Exemplary demonstration of overlapping of the two of the objects. In each part, a snapshot from video frame is given with an intensity image before the color identification and tracking algorithm (left-hand side), and a colored image after all processes (right-hand side) for a comparative analysis; (f) Estimated path between

part (a) and part (e) of the Kalman filter is drawn indicating the performance of our tracking algorithm.

3.1.3. Generating the Intention Feature Space (Training Experiments)

The aim of the training experiments is to collect training data to construct intention space (feature space). This feature space is used for learning process to construct estimation model with Hidden Markov Models (HMM) and elastic networks detailed in the upcoming sections. Experiments were conducted with 6 human subjects in the experimental room with the same setup explained in the Section 3.1.1. A ceiling camera with the resolution of 240x320 tracked the humans with 15 fps and recorded human actions in the room. To construct the estimation model in the learning phase detailed in Section 3.2.1, intention feature space should include intention trajectories for all of four intentions given in Table 3.1. To record as much trajectories as we can for all intentions, there were some milestones about the content of the experiments stated as follows:

- First of all, we should note that, our final testing experiments require that the people are not told about any objects and the robots in the room in order to test our trust and curiosity gaining performance. Therefore, we started training experiments with uninformed humans about the context of the room including the robots, to have a consistent intention model with the final experiments. That way we could collect trajectories with completely natural behaviors.
- During the experiments we noticed that if all of the human subjects were uninformed about the context of the room, all we could observe from the actions was discovering the room. However, to construct the database for all of the intentions, we needed the people to realize other three intentions mentioned in Table 3.1, as well. For that purpose, we had only 3 human subjects uninformed, while remaining three were told about the robots and that they should pay attention to robot moves. By paying attention, we could manipulate the actions of the subjects to collect more data for the other

intentions. For example, moving chair reminded these three informed subjects to sit on the table and work with PC giving us a trajectory realizing the intention of 'sitting on the table'.

- Moreover, to utilize Hidden Markov Model, there should be transitions between the intentions observed (detailed in Section 3.2.1). To have the transitions, we told all of six human subjects that they were expected to return the objects they took, back to their original places. For example, as soon as a person quits reading a book while sitting on the table, s/he should return it back to the library giving us a transition from 'sitting on the table' to 'getting/dropping a book from the library'.

With the aforementioned rules about the experiments, all of the human subjects spent 5 minutes of time in the experimental room and their actions were recorded. Exemplary tracking information from recorded actions is illustrated in Figure 3.9(b) with its originally captured last frame given in Figure 3.9(a).
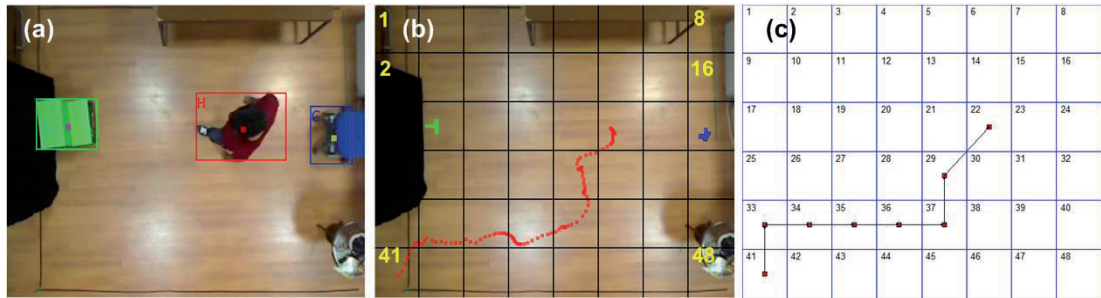
Figure 3.9 (a) Original image captured at the last frame of a tracking sample; (b) An exemplary tracked trajectory with 7 seconds long shown on a grid map which has its last frame shown in (a). This trajectory is an intention feature descriptor labeled with 'discovering' intention. Labeling is done by the experimenter based on his observations; (c) Grid numbered view of the trajectory in part (b).

As for our intention feature space, it consists of feature descriptors with tracked trajectories each labeled with an intention as in the example trajectory labeled with 'discovering' intention given in Figure 3.9. It should be noted that, labeling is realized by the experimenter himself based on his observations, meaning that a trajectory ended up with which intention. After recording trajectories containing pixel coordinates of each location of the person, there are three requirements for our intention feature space for the future applications and the solutions applied to them:

1. As will be detailed in Section 3.2.1, HMM requires a finite number of emissions (outputs) in order to have realistic probability functions for states. Therefore, we needed to divide the map into grids, in our applications being 48 grids with the same size. Each pixel values are also recorded with the number of the grid they fall into as illustrated in Figure 3.9(c).

2. Intention feature space should consist of feature vectors having trajectories with equal length. The reason is that, since Elastic networks have a search space with nodes having same dimensions, intention feature descriptors, which are nodes of elastic nets, should have same dimensions. Experiments showed that the average time for a human subject to realize an intention was 7 seconds in our room. As a result, we fitted all of the intention feature descriptors (recorded trajectories with a labeled intention) to 7 seconds being 105 frames (7 secs x 15 fps). Normally, HMM doesn't have to have equally length features for training; however, for Elastic networks this adjustment is need to be handled.

3. Since HMM is a method calculating conditional probabilities of an intention occurring on a certain grid given that the previous intention, in order to calculate output (grid) probabilities consistently, there should be same amount of feature vectors labeled for each intention. After the experiments, we could extract 25 different 'drinking coffee' trajectories as the limiter. Therefore, 25 trajectories were assigned for all of four intentions.

After these requirements satisfied, total amount of 100 discrete intention feature vectors (trajectories each labeled with an intention) are extracted from the videos each having $7x15fps = 105$ frames with human pixel coordinates (x, y) and grid numbers for each frame. Each feature descriptor is in the size of $[3x105]$ with the row matrix: $\begin{pmatrix} x \\ y \\ grid \end{pmatrix}$ and column matrix:$(1^{st}\ frame \quad ... \quad 105^{th}\ frame)$.

3.2. Intention Estimation and Real-Time Prediction of Human Trajectories

Intention estimation is realized based on the idea that human direction of motion (heading) reflect an intention (Clarke et al., 2005)(Bratman, 1999). That is to say, that humans plan their actions according to their set intentions. We, thus, monitor actions of the human agents with the context of a scenario defined by a room equipped with limited number of realizable intentions given in Table 3.1 . These actions are recorded according to the exhibited context dependent final intention as feature vectors of our system (see Section 3.1.3). For the machine learning, we utilize these feature vectors in the intention estimation approach based on Hidden Markov Models (HMM) as in PhD thesis work of Durdu (Durdu, 2012).

Intention estimation is required in our system because we are comparing the current intention estimated with the desired one if we could change successfully. Moreover, for the search space of our Elastic Networks algorithm, we introduce feature vectors, which are the trajectories of the person leading to a certain intention given in Section 3.1.3, as nodes of the network. Therefore, the input to elastic network should be a "*trajectory*" of the current intention (see Sections 3.4.4 and 3.4.5). In other words, after estimating current intention, we predict the trajectory of the human which will lead to that intention details of which are given in Section 3.2.2. For the intention estimation, we constructed a model characterizing intention actions of Table 3.1 as a set of sequences of emissions using HMM. The model thus finds the most probable set of state (intention) transitions corresponding to a certain observed action. Since

45

we have hidden states being intentions and observable actions being sequences of locations, HMM suits well to our system.

3.2.1. Hidden Markov Model for Estimation

In order to estimate the intentions, there should be constructed a model, which is called the learning process of a stochastic system. Hidden Markov Model provides a general framework for sequential decision making in situations where states are hidden and actions (observations) are stochastic. In this model, there are finite set of states which are linked to finite number of outputs with probability distribution functions (Rabiner, 1989). Formal definition of a HMM is given as:

$$\lambda = (A, B, \pi) \tag{3.2}$$

There are N hidden states expressed with alphabet set S, and M observations with the set O:

$$S = (s_1, s_2, \dots, s_N), O = (o_1, o_2, \dots o_M) \tag{3.3}$$

Transition array A, supplies the probability of state j following state i and defined as:

$$A = \{a_{ij}\}, \ a_{ij} = p(s_i | s_j) \ satisfying \ \sum_{j=1}^{N} a_{ij} = 1 \tag{3.4}$$

Observation array B, holds the probability of observing output k at the state j defined as:

$$B = \{b_{kj}\}, \ b_{kj} = p(o_k | s_j) \ satisfying \ \sum_{k=1}^{M} b_{kj} = 1 \tag{3.5}$$

For the last component, π is prior information defining a probability array of the state i being the initial state:

$$\pi = \{\pi_i\}, \quad \pi_i = p(s_i) \ satisfying \ \sum_{i=1}^{N} \pi_i = 1 \tag{3.6}$$

The learning task in HMM is to find: Given emission (output) sequences and a model find the most probable set of state transitions results in that output and compute the emission probabilities. As for the testing task(finding the most likely states), it can be declared as: Given an emission sequence and a model (emission and transition probabilities calculated in learning session) find the most probable set of state transitions results in that output. These tasks have been studied well and solved using dynamic programming techniques. For the learning task, Baum-Welch algorithm was introduced finding the (A, B, π) parameters which are state transition, observation and initialization arrays of the model by Baum et al in 1970 (Baum et al., 1970). For the latter task, testing, Viterbi in 1967 introduced Viterbi algorithm calculating the most likely states (Viterbi, 1967).

As for the learning process with HMM, all database of extracted feature descriptors in Section 3.1.3, which are trajectories of humans labeled with an intention, are used to construct the model. These feature descriptors train our HMM that is subsequently used to estimate current intention when viewing a certain human heading adapted from (Durdu, 2012). In its traditional usage of HMMs, there are two sub-models to be well defined beforehand: state transition model and observation (emission) model (Yamato, Ohya, & Ishii, 1992). Our HMM uses four possible intentions given in Table 3.1 as being states in the transition model to calculate state transition matrix while the grid information is utilized in observation model to calculate emission matrix.

3.2.1.1.   Generating Emission and Transition Matrices (Baum-Welch Algorithm)

This part is the training part of HMM that is, given a set of sequences; estimate the parameters of the model. Baum-Welch Algorithm, which is actually an Expectation-Maximization algorithm, is used to calculate the maximum likelihood of the model parameters from emissions and states (Dempster, Laird, & Rubin, 1977). For the purpose of estimating state transition and emission probabilities of our HMM, all of

the trajectories were examined. As mentioned in Section 3.1.3, to have consistent state (intention) probabilities for each grid location in our ROI, equal amount of trained trajectories should be used for each intention. In our application, 25 trajectories from each intention, meaning that total amount of 100 trajectories (with 105x100=10500 grid information (observations), see Section 3.1.3) were loaded to the Baum-Welch algorithm and recursive calculations were handled until a convergence to the model parameters was satisfied adapted from (Dempster et al., 1977).

Calculated model parameters being transition probabilities and emission probabilities are shown in Figure 3.10 and Figure 3.11 recursively. In Figure 3.10, all of the emission probabilities of each intention are demonstrated with bar graphics. In the graphics, each bar represents the probability of observing the related intention in the related grid number. It should be reminded that our ROI in the scene is divided into 48 equally sized grids (see Section 3.1.3). As for the analysis of emission probabilities, we can deduce that 'discovering the environment' the first intention spreads nearly homogenously to the grids as demonstrated in Figure 3.10(a). On the other hand, the other intentions heavily occur at their located places. For example, by looking at Figure 3.10(b), (c) and (d) respectively, it is clear that experimenters drink coffee near the grid numbers of 39, 40, and 47, whereas they take book from 17, 18 and they sit near 5, 6, 13, and 14. All of these deductions from the graphics can be compared with the original image in Figure 3.9(b). As for the last words on emission probabilities, the model satisfies the common probability rule that all of the grid probabilities of each intention sum up to 1.

Transition probabilities are given in Figure 3.11 with bar graphics. The horizontal axis states the four intentions with numbers (explained on the figure) whereas the vertical one indicates the probabilities of transitions from each numbered intention to its different-colored bar represented intention. If we look at the highest bars for each intention, it is clear that experimenters generally leaned to go on with what they were doing, that means breaking the obstinance of the experimenters on what they are doing is very hard. In addition, experimenters generally passed to the intention of

'discovering the environment', after they had finished their previous intention. This explains that, since the experimenters were not familiar with the environment and suspicious about it, they usually do not plan what to do next after they finished what they were doing. In other words, they observe the environment first, and then plan what to do next. Moreover, the common probability rule was also satisfied here that, all of the transition probabilities of each intention summed up to 1. Finally, the last parameter being the probability of initializations of the states given with Eq.(3.6) was not given as a graph because all of the experimenters started to discover the environment first, resulting in the probability of this intention being '1' and others are '0'.

In conclusion, with the analysis and the calculations made on the graphs given, we can deduce that our model is a workable and feasible model to be used in the estimation of intention, which is our ultimate goal for this section.
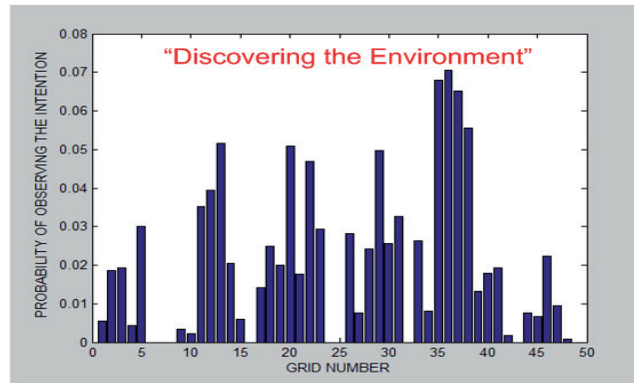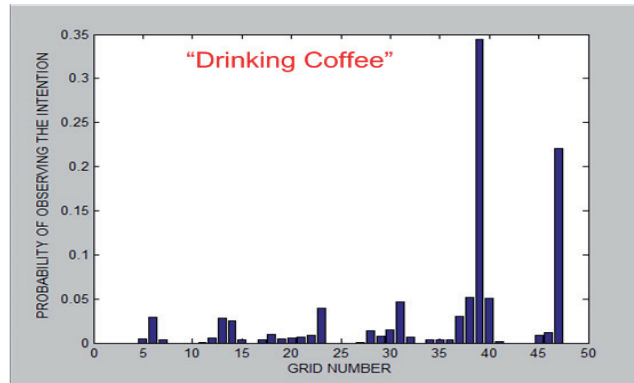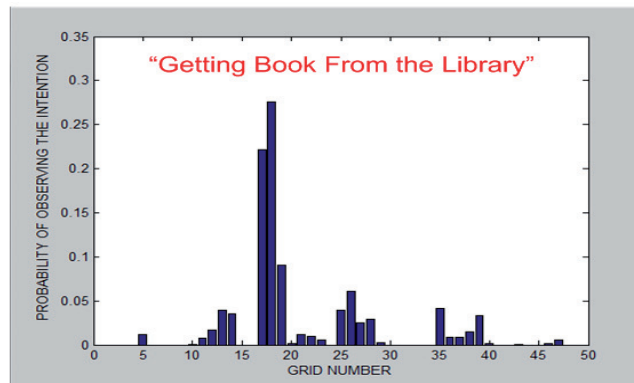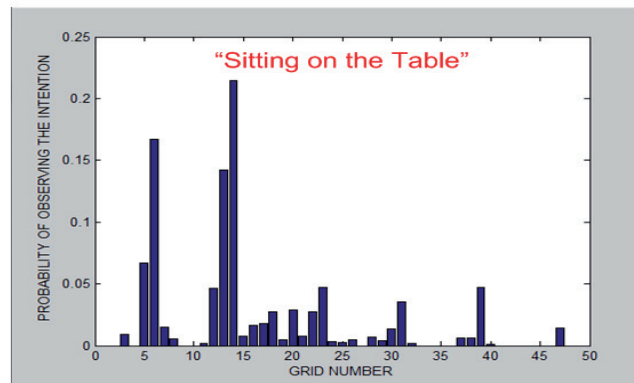
Figure 3.10 Emission probabilities of each intention, each bar in the graphics refers to probability of observing the related intention in the related grid.
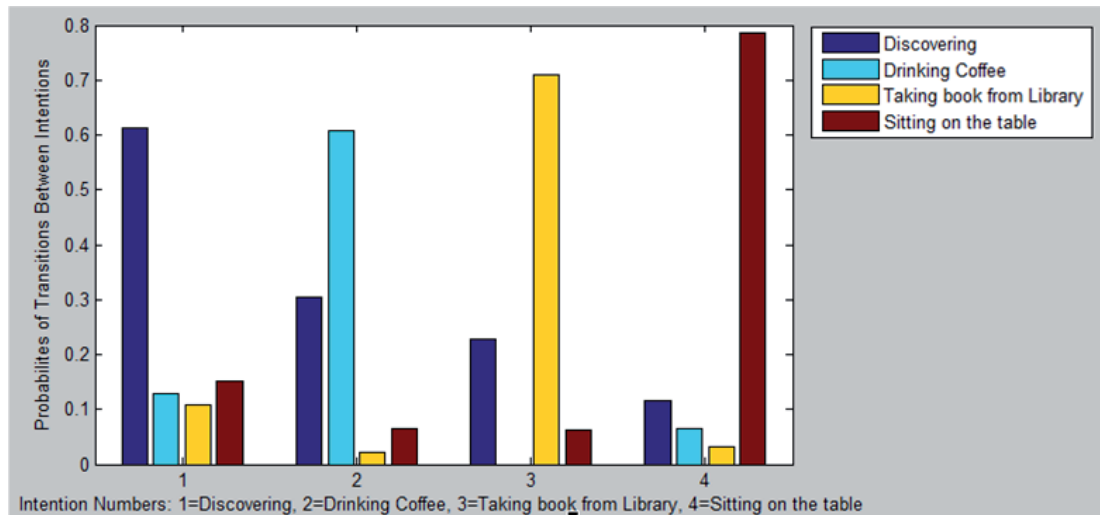
Figure 3.11 Transition probabilities of each intention, each bar in the graphic refers to the probability of transition from the intention stated with the number below the bar in the horizontal axis, to the intention stated with the color of the bar.

3.2.1.2. Finding the Most Likely States, Estimating the Intention (Viterbi Algorithm)

After estimating the model parameters of our system, the final part being intention estimation takes place. It is actually the estimation of the most likely state and explained in the literature as; given a sequence of observations and a model, including emission, transition and initialization probabilities stated in the previous section, calculate the most probable sequence of states using Viterbi algorithm (R Durbin, 1998).

The most important tip to notice using Viterbi algorithm is that, changing the length of the sequences given to the Viterbi algorithm to estimate the most likely states results in different estimations. The reason is that, maximum likelihood calculations are handled by not only utilizing both emission but also transition probabilities between states. Therefore, according to the previous consecutive locations before the

last location (emission) in a given sequence, found most likely state for the last emission may vary. As a result, the longer the sequence given to the algorithm, the more reliable estimations we get.

In our application, close loop algorithm being reshaping intention into a desired one by autonomous robot moves in human-in the-loop experiments starts here. Path planning is done according to the currently estimated intention in real-time experiments. Therefore, we had to consider the real-time performance of our system before estimating intention. At real time due to the CPU performance limitations and many computational costs in the algorithm, video frame can be maximum 8fps while the constructed intention feature space with training experiments was 15 fps. This doesn't affect our HMM parameters since they are related with the probabilities of observations and state transitions which are independent of the frame rate. However, there should be modifications on the intention feature space and its descriptors, all of which were constructed with 15 fps, because they are going to be used by trajectory estimation and path planning in real-time experiments with 8fps. This modification being the reduction of lengths of trajectories in the feature space is detailed in the Section 3.2.2.1. We decided to estimate the intention by looking at the grid observations of the human subject of the last 3 seconds with 24 sequences (estimation will be in real-time, $3\text{secs} \times 8\text{fps} = 24$ sequences of observed grids) recalling our claim that a person realizes an intention in about 7 seconds. Viterbi algorithm takes these 24 up-to-date grid sequences observed at all iterations and gives 4 different probabilities belonging to four different intentions for each sequence. Finally, we declared the state which has the biggest probability among four of the last sequence, as the current intention estimated of the human subject. We demonstrate the estimated intentions on the top-left corner of the image frame as seen in Figure 3.12. In the figure, emission and transition matrices constructed are also shown. More results on this topic can be found in results section (Section 4.1.2).

As for the last words in intention estimation, recalling flow chart of the system in Figure 3.1, the estimated intention is compared with the final desired intention. If they are equal, the system terminates here claiming success; on the other hand, real-

time trajectory estimation takes place before moving on to the execution mode as detailed in the proceeding sections.

3.2.2. Real-Time Human Trajectory Prediction

As mentioned before, our intention feature space consists of trajectories labeled with an intention. This map of intentions is actually our search space in which we are going to generate trajectories for our robots to follow by using elastic network methodology. As it is detailed in Section 3.4.5, our elastic network needs currently estimated intention of the human to generate a way point close to it. However, it is obvious that intention estimated (one of four intentions in Table 3.1) is meaningless for elastic network model utilizing trajectories in the intention feature space. In addition, elastic network's way point generation is in real-time in our application. For the purpose, we predict a trajectory for the current intention of the human subject among the intention feature descriptors in real-time. Therefore, before the explanation of trajectory estimation, real-time adaptation of intention feature space is detailed.

3.2.2.1.   Real-Time Adaptation of Intention Feature Space

As it is first stated in Section 3.2.1.2, our close loop algorithm is a human-in-the-loop algorithm in real-time. Processing video frames and having a search algorithm running in the background at the same time allocate great amount of CPU memory leading to a slowdown in PC performance. Although we were able to record videos with 15fps for training part of the project, due to CPU performance it decreased to 8fps at maximum in real-time analysis. This reduction in the frame rate doesn't affect any of the previous steps, none of them frame dependent; however, it affects trajectory prediction and path planning algorithm, both of which is in real-time, because of the different sizes of the currently predicted trajectory of the human recorded with 8 fps and intention feature space constructed with 15 fps used as a search map. Therefore, all of the constructed feature descriptors labeled with

53

intentions each having 105 locations (7sec x 15fps) should be scaled to 56 locations (7sec x 8fps). We could not record these feature descriptors in real-time at the training experimenters because in real-time recording a video while processing each frame is impossible due to the limited cache memory. Therefore, we recorded training videos first then process it with image processing tools. In conclusion, by shrinking each feature descriptor of the intention space to 56 sequences of location, we adapted the feature space to real-time conditions.

3.2.2.2. Predicting the Current Trajectory

Predicting the current trajectory means finding the most possible trajectory that the human subject may follow. The reason why we did not predict the trajectory at the beginning instead of estimating intention is that, trajectory prediction with a lot of possibilities is not feasible in our small experimental room. We are trying to estimate the trajectory by observing the first 3 seconds of the human action which afterwards may proceed to many different trajectories in the room where all four intention locations are very close to each other also. In other words, a lot of different trajectories leading towards different intention locations have the same initial trajectories. By estimating intention beforehand, we decrease the search space to only trajectories ending up in the estimated intention because of the fact that all of the trajectories are labeled with one of four intentions. In fact, there are 100 different trajectories in our search space and estimating intention beforehand decreases this number to 25 possibilities (recalling that there are 25 trajectories for each intention). In addition, intention was going to be estimated anyway in order to check if we could reshape intention of the person into what we desired. Therefore, it is a good enough solution for us to utilize intention information and then do the trajectory prediction.

Trajectory prediction comes right after intention estimation and it is realized by comparing the track of the person with all of 25 trajectories in the estimated intention space. In comparison, the last 24 sequences (last 3 seconds in real-time, see Section 3.2.1.2) are used as the recent track information of the human subject. The aim is to estimate the entire trajectory (7seconds long) by looking at these recent 24

sequences. For the purpose, we use a shift comparator which first calculates distances between the first 24 location of trajectories in the current intention space and recent 24 sequences of human location. After the closest one found, the most recent 24 sequences is updated with the new frame and is compared with the previously found closest trajectory's location sequences between 2nd and 25th, omitting the first one. This is for checking if the person is still on the estimated route. The algorithm does this shifting comparison for 4 consecutive frames as long as the distance of the closest trajectory to the recent 24 sequence decreases or stays the same. If it does not, the algorithm searches for a new prediction. Either way, an independent counter counts and after 28 sequences (3.5 seconds) passed, lastly labeled closest trajectory is delivered to the elastic network model as the current trajectory. An exemplary result of trajectory prediction is shown in Figure 3.12. Yellow trajectory shows the previous trajectory that the person actually followed within last 7 seconds while the red one belongs to currently predicted one selected from the feature space of the intention 'discovering', which is the currently estimated intention written on the top-left corner. Finally, previous and current trajectories stated are passed to our elastic network model. More results on this method are given in results section (Section 4.1.2).
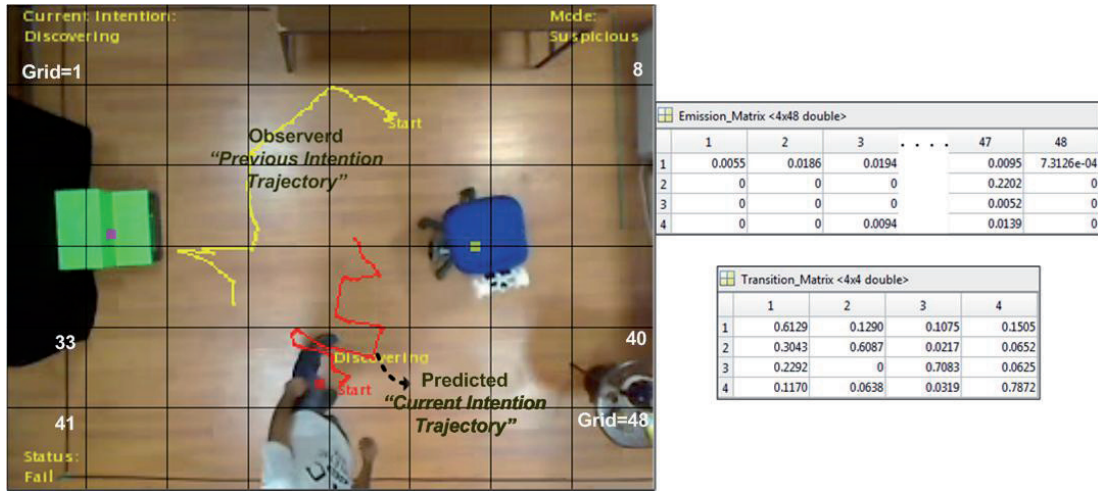
Figure 3.12 Demonstration of current intention trajectory prediction. Previous trajectory is the observed trajectory of the human recorded in the last 7 seconds where the currently estimated trajectory is the one human subject is expected to follow next. This trajectory is chosen among the trajectory space of the intention of 'discovering' which is the currently estimated intention written on the top-left corner of the snapshot.

3.3. Execution Mode: Human Body-Mood Detection

Execution modes: *confident* and *suspicious,* yield different strategies in planning way points for the robots to follow. These modes are depicting currently detected body-moods of human subjects which states the emotional aspect of a person effecting his/her decision making in an unknown environment (S. Lee & Son, 2008). Naturally, a robot should catch the curiosity of the human if we want to reshape the human intentions by autonomous moves of robots. Therefore, the robot firstly aims to win this curiosity and make the human confident with the environment, based on the research that a person in a confident mood elicits more external-focused attention than a suspicious one (Fredrickson, 2003; Grol et al., 2013; Sedikides, 1992; Wadlinger & Isaacowitz, 2006). Briefly, before reshaping the intention to a desired one, we aimed to make the human confident with the robot and trust it. Robots will

56

not be focused on the ultimate desired intention until the body mood of the person is detected to be *confident*. As for this part of the project, we can say that detection of the body-mood is a decision making process for the robots to act or plan their paths accordingly.

In this section, the methodology behind the body-mood detection is given. By looking at the works of Huettenrauch et al. (2006), Christensen et al. (2005), Butler et al. (2001) and Mead et al. (2011) examining spatial relation between a robot and a human subject, we made an inference that if a person heads towards a robot allowing it to enter his/her intimate region (which is closer than 30cm), that means s/he feels comfortable putting him/her in a confident mood and ready to start an interaction. On the contrary, staying or moving away from the robot reveals suspicion or unconcern of the person. These ideas are the milestones for our approach on body-mood detection. Basically, we are tracking the heading of the human body, whether the person moves through the robot or not. Since we are observing the reaction of the person against the robot moves, the detection of the body-mood takes place right after each robot move assuming *suspicious mode* at the very beginning of each experiment. After the robot makes its move, we compare the location of the human before and after the movement. If the direction of motion towards the robot and the person is close enough to the robot, we state the person is *confident*. As a result, we claim that the person can give more external-focused attention to our robot as Sedikides (1992) stated making our execution mode of path planning to switching to *confident mood* after which we plan a path for the robot aimed at reshaping the intention of the person to the desired one. On the other hand, if the heading is not concerning the robot, *suspicious mode* starts, where the aim is making the mood of the person *confident* and gaining the human curiosity by roaming around the current intention of the person regardless of how far they are from the desired intention.

The detection of the heading of human subjects is realized by geometrical approaches on the 2-D image coordinate plane with the coordinates of the human just before the robot moves, after the robot moves and the robot itself in the last frame (see Figure 3.13). In the figure, all of the coordinates mentioned is shown. Actual

heading is the direction of motion of the human between before and after the robot moves whereas ideal heading is the heading leading to an interaction between human and robot switching the human body-mood to *confident.* However, of course the ideal heading couldn't be achieved in a coordinate system. In addition, human may approach to the robot from different sides of the robot. Therefore, we defined an error margin called *angle of curiosity* stating the curiosity of the person in the robot. If the distance (proxemics distance on Figure 3.13) between the last locations of the human and the robot is below 75 pixels (close enough to interact within a second) and if the human displaced more than 10 pixels (standing still not yields confidence), a comparison of this angle with a threshold of 45 degrees decides the mood of the person. Below the threshold yields the mood is *confident* whereas above means *suspicious.* Calculation of this angle is realized by the conventional law of cosines. Calling *actual heading* as *a*, *ideal heading* as *i,* and the distance between the last human and robot locations as *d, angle of curiosity* is calculated as follows:

$$angle\ of\ curiosity = arccos\left(\frac{a^2+i^2-d^2}{2\times a\times i}\right) \qquad (3.7)$$
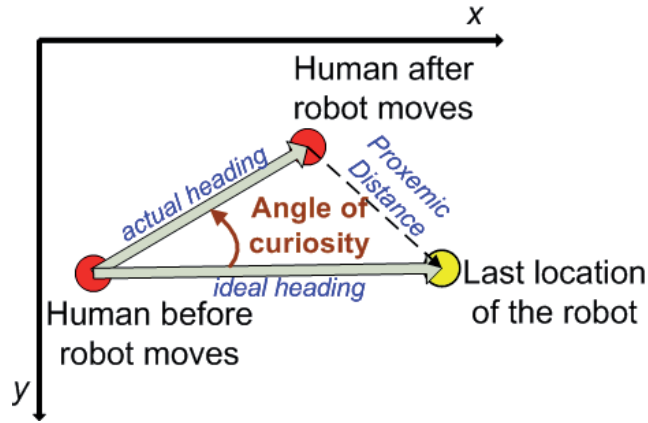
Figure 3.13 Geometrical approach of deciding human-body mood. The line connecting the location of the robot and of the human before robot moves is the ideal heading whereas the other line between human locations before and after robot moves is actual heading. Angle of curiosity is the angle between these two lines stating the error from the ideal heading leading to confident mood for the person.
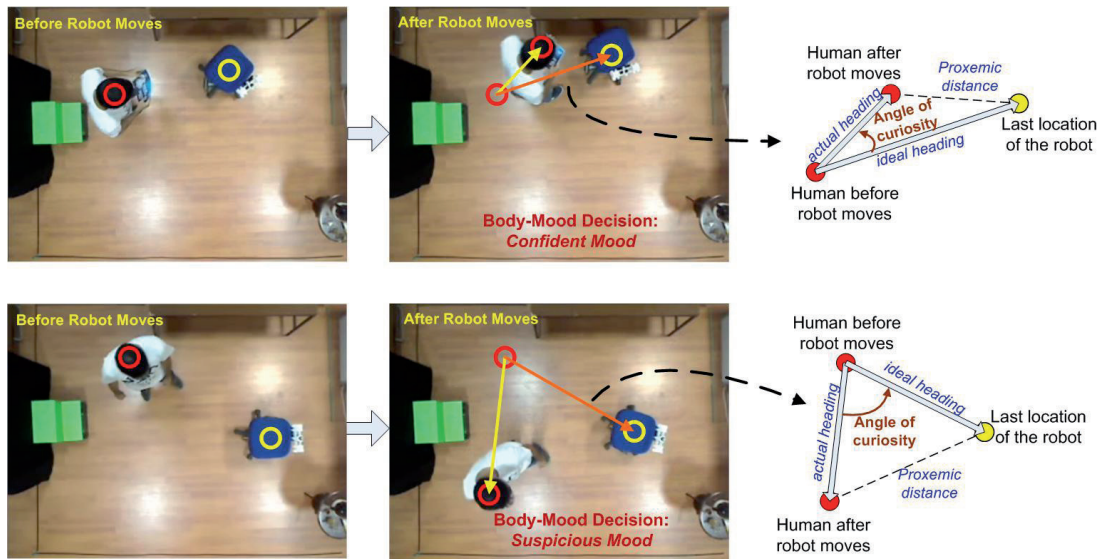
Figure 3.14 Examples of different execution modes. (a) Human subject approached robot with about 35 degrees of angle of curiosity leading to confident mode; (b) Human subject stood still not caring about 2-steps robot move switching the execution mode to suspicious mode.

There are two examples from actual scenarios given in Figure 3.14. The former one is claiming that the person is *confident* and the execution mode is *confident mode* while the latter clearly reveals the unconcern of the person for the robot declared as *suspicious* while he was sitting without caring the movement of it. In the scenario given in Figure 3.14(a), the robot will be realizing the desired intention of 'sitting on the table'; on the other hand, in Figure 3.14(b), the robot will be still trying to gain the curiosity of the person making him *confident.* Different strategies in path planning according to these two modes are realized by changing one parameter in Elastic networks described in the Section 3.4.5. More results on this topology within the real-time scenarios are given in results section (Section 4.1.2).

3.4. Generating Intention Transients Using Elastic Networks

The conceptual novelty in our work is the planning of intention trajectories towards the desired intention consisting of way points in intention feature space created in Section 3.1.3 with two different modes: *confident* and *suspicious*. The former mode is adopted when the human keeps following the direction of the interacting robot. In this mode, robots directly generate a way point (trajectory) passing near the current intention of the person pointing out the desired intention. However, in the latter mode, the robot cannot quite destabilize the obstinance in the current intention of the human and cannot capture his/her curiosity. In this mode, since the human is unconcerned about the robot and has suspicions, establishing an interaction between human and the robot is very difficult as clearly mentioned in(Christensen et al., 2005). Therefore, our robots managed to achieve the aim of making the human confident and easily interact with themselves by slowly getting close to the human subjects as stated in (Suzuki et al., 1998) while mimicking one of their previous actions observed (Kerstin Dautenhahn, 1999). That way, the human subject can carry out more external-focused attention, be curious about the robot and ready to start an interaction (Sedikides, 1992). For the purpose, the way point is generated around the current intention regardless of how far we are from the desired intention. Additionally, the found way point is in the dense areas of intentions observed in the

feature space (intentions that are previously realized by the experimenters), which are familiar to the person. Basically, we can state that the ultimate goal in *suspicious mode* is to destabilize the obstinance and gain the curiosity of the person switching the mood of the person and the execution mode to *confident mode*. Each way point found by the elastic network is executed by moves of an adequate robot (2-steps or chair robot) in adequate directions stated by the trajectory in the way point.

These requirements of planning way points as transient intentions through the desired intention in the intention feature space are realized by adapting the elastic networks of Durbin and Willshaw introduced in 1987 (Richard Durbin & Willshaw, 1987). In this section, a detailed literature review on elastic networks, advantages of using elastic networks, adaptation of our intention feature space to elastic network and utilization of it are given with an example explaining our methodology.

3.4.1. Literature Review on Elastic Networks

3.4.1.1. Elastic Networks in Macromolecules

Elastic networks were first studied as Gaussian Network Models (GNM) or Normal Model Analysis to understand and characterize the mechanical aspects and long-term dynamics of biological macromolecules simulate chemical interactions between atoms by Tirion in 1996 (Tirion, 1996) one year later in amino-acid level by Bahar et al. (Bahar & Rader, 2005) and Haliloglu et al. (Haliloglu, Bahar, & Erman, 1997). The method proceeds from complex semi-empirical potentials characterizing the covalent and non-covalent interactions between atoms. In this study, atom interactions were represented as a mass and spring elastic network. Figure 3.15(a) explains the elastic networks structure in GNM where every node (atom) is connected to its spatial neighbors by uniform springs (interactions).
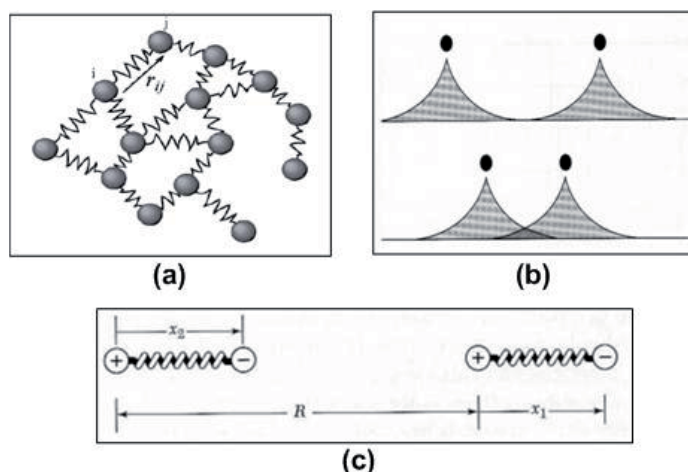
Figure 3.15 (a) Elastic Network model with schematic representation. Every node is connected to its spatial neighbors by uniform springs. r_{ij} being a distance vector between nodes i and j (Chennubhotla, Rader, Yang, & Bahar, 2005); (b) Solid circles being nuclei, electronic charge distributions overlap as atoms approach representing repulsive interaction (Kittel & McEuen, 1996); (c) Representation of the atoms as oscillators in Van der Waals-London interaction modeling (Kittel & McEuen, 1996).

In chemistry to investigate the crystal structures, inert gases are mainly used since they form the simplest crystals. There are two major interactions between inert gases constructing these crystal structures or bindings. These are repulsive interaction and Van der Waals-London interaction. The former interaction is explained as such; when the atoms brought together, their charge distributions gradually overlap, results in changing the electrostatic energy of the system. At sufficiently close separations the overlap energy will be repulsive due to the fact that two electrons cannot have all their quantum numbers equal as demonstrated in Figure 3.15(b). The latter one describes the interaction when the separation between atoms is large. Referring to the Figure 3.15(c), at a separation R large in comparison with the radii of the atoms, the interaction due to the charge distribution on the atoms (covalent or non-covalent) will be zero. However, this time their nucleuses will carry an attractive force between them. Nucleuses induce dipole moments in each other, and the induced moments cause an attractive interaction between the atoms. An analogy can be formed as such;

if we consider the two atoms as two identical linear harmonic oscillators as in the figure, it is clear that when their opposite charges are aligned horizontally, there will be a coulomb interaction energy of these two oscillators (Kittel & McEuen, 1996). Energy equations happened due to repulsive and Van-der Waals attractive interactions are given respectively as follows:

$$U_R = \frac{\sigma}{R^{12}} \, , U_A = -\frac{\sigma}{R^6} \tag{3.8}$$

In 1924, these two energy equations were first introduced together as a mathematical model that approximates the interaction between a pair of neutral atoms or molecules by Jones (Jones, 1924). The formulation called Lennard-Jones potential is given as:

$$U(R) = 4\varepsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right] \tag{3.9}$$

In the equation, ε is the depth of the potential well, and σ is the finite distance at which inter-particle potential is zero (a property that changes from the type of atoms). R is the distance between two atoms. In the later studies, since the first term has no theoretical justification as the second one, repulsive energy equation was approximated by an exponential function given as:

$$U_R = \exp \left(\frac{-R}{\rho} \right) \tag{3.10}$$

where, ρ is used to indicate the maximum range of the repulsive interaction takes place. In Figure 3.16, the total potential energy between two atoms and its change according to the distance between them can be seen clearly. The minimum occurs at the state of equilibrium. After ρ, as the atoms keeps getting close to each other, the curve will be very steeply increasing meaning the repulsive force, the first term, will be very high. Finally, the resultant force applied on an atom can be calculated as $- dU(R)/dR$.
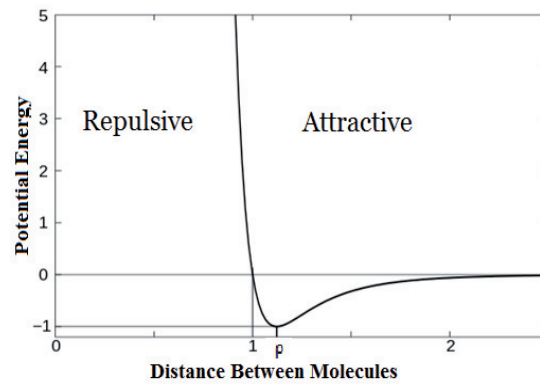
Figure 3.16 The characteristics of the Lennard-Jones potential which also describes the interaction between the nodes in an elastic network model (Kittel & McEuen, 1996).

3.4.1.2. Elastic Network in Control Applications

As detailed, elastic network methodology has two kinds of interactions, each takes effect at different ranges. In addition, these interactions are defined such that, resulting forces are integrable. That means, there will be an energy minimization model in which system will try to find the point where the derivative of energy function will be zero, or the forces will added up to zero where the system will be at equilibrium. Moreover, these energy minimization models seek for the energy equation to be optimum. All of these specialties make the Elastic Networks a preferable method which can be used in the control world optimization problems.

Although there are not many applications using this method in optimization systems, Durbin and Willshaw in 1987 approached the "*Travelling Salesman Problem (TSP)*" using the elastic nets, which was a pioneer work in the control area being our main inspiration sources (Richard Durbin & Willshaw, 1987). Later on in 1989, Durbin analyzed the formulation and the methodology they created by doing sensitivity analysis on each parameter introduced in the energy function of their elastic network

(Richard Durbin, Szeliski, & Yuille, 1989). In these studies, an elastic network model used for the establishment of topographically ordered projections in the brain (Wurtz & Albano, 1980) was utilized in TSP, a classical combinatorial optimization problem. The problem can be defined as, given the positions of N cities, they may be lie in a plane, find the shortest closed tour in which each city is visited once. Durbin and Willshaw created an algorithm in which the procedure was the successive recalculation of the positions of dynamic points (moving atoms) describing a closed contour encircling cities (steady atoms) in the plane. This contour being initially a small circle is gradually elongated non-uniformly to fit elastically around cities by tracking the minimization of the energy equation. In Figure 3.17**,** this circle, called as rubber band, can be seen clearly from (a) to (f) while it elongates and captures all of the cities. The two forces of this energy equation are defined as a force attracting dynamic points towards cities (steady points) while the second one is between dynamic points holding them together. More to speak on these forces, the former force aims to find all of the cities whereas the latter one is for the shortest path requirement, keeping the neighboring dynamic points together to minimize the total path length. Initially, all cities have nearly equal influence on each dynamic point. As the iterations proceed and the rubber band elongates, the influence of the city on a dynamic point closest to it gets much bigger than those of other cities while the attraction between the dynamic nodes weakens. This change of specificity guides each dynamic node to the closest city while the first force still favors the shortest contour perimeter. The gradual increase of specificity is obtained by the gradual decrease both in the distance where the second force becomes dominant and on the effect of the first force. This change in the balance of the forces is controlled by a length parameter called "K" that eventually tends towards zero. This reduction is equivalent to lowering the temperature in *optimal simulated annealing* method (Kirkpatrick, Jr, & Vecchi, 1983).
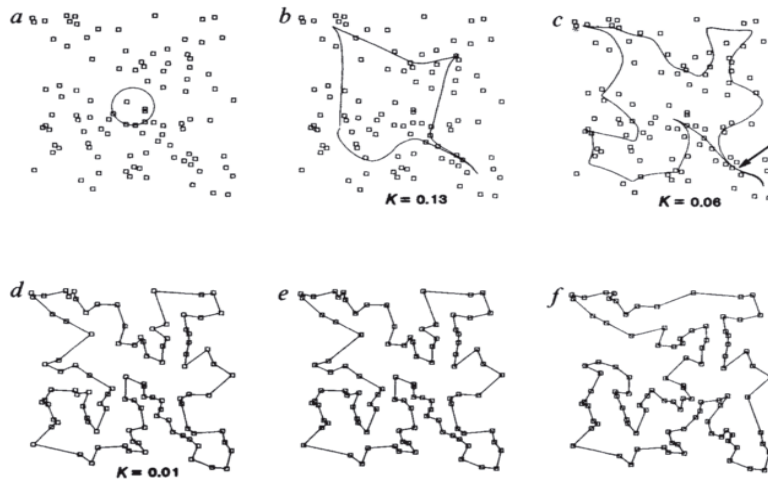
Figure 3.17 An exemplary progress of the elastic network mathematical model solving a TSP problem. Initially a rubber band circle is defined at the mass center of the cities. With the forces applied on the dynamic points forming the rubber band, the band elongates through the cities starting from part (a) to covering all of the cities in part (f).

Simulated annealing analogy with thermodynamics simulates cooling the metals. Temperature is lowered slowly allowing thermal equilibrium to be attained at each stage. At high temperatures, thermal mobility is high but as T goes down it is lost and molecules tend to line themselves up in a rigid structure. This rigid state is a state of minimum energy. As long as T is decreased slowly, the nature is almost certain to find it itself. Durbin and Willshaw introduced the parameter K as an analogy to the parameter T of the simulated annealing method(Richard Durbin & Willshaw, 1987). They keep K parameter or temperature high in the beginning, resulting in the energy landscape being smooth and showing only coarse features (see Figure 3.18). By this way, the system moves into a globally low energy region bypassing local minima. When the gain is increased, it reveals all the details in the energy surface allowing to find a specific local minimum. To conclude, since this methodology offers a good solution to combinatorial optimization problems, it is commonly used in the literature (Geng, Chen, Yang, Shi, & Zhao, 2011; Nahar, Sahni, & Shragowitz, 1986).
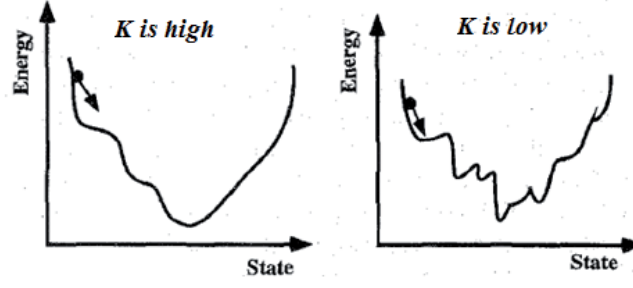
Figure 3.18 Energy landscapes for high and low K parameters. When K is high, the surface is smooth and the system can move toward globally better solutions whereas lowering K reveals all the details along with the local minima in the energy surface (Shams, 1996).

After explaining the *K* parameter effect, we will give details on the formulations of the mathematical model that Durbin and Willshaw used because in our applications, we modified these formulations as for the solution of our problem. Their rule for the change in the coordinates of dynamic nodes (force function in conventional elastic nets) was given as:

$$\Delta y_j = \alpha \sum_i w_{ij}(x_i - y_j) + \beta K(y_{j+1} - 2y_j + y_{j-1}) \qquad (3.11)$$

where α and β are for the relative strengths of the force of cities and the force between the dynamic nodes. x_i is the city location and y_j is the location of a dynamic node. Coefficient w_{ij} gives the influence of city *i* on the dynamic point *j*, which can be stated as the strength of the connection between them and it is formulated as:

$$w_{ij} = e^{-|x_i - y_j|^2/2K^2} \Big/ \sum_k e^{-|x_i - y_k|/2K^2} \qquad (3.12)$$

This coefficient is a normalization of the force of the cities, so that the total influence of each city is equal. This force equation is integrable; therefore, an energy equation of the system can be introduced as:

$$E = -\alpha K \sum_i \ln \sum_j e^{-|x_i - y_j|^2/2K^2} + \beta \sum_j |y_{j+1} - y_j|^2 \qquad (3.13)$$

67

It is clear that, the first term of the energy equation is the repulsive interaction defined in Eq. 3.10 with a minus sign making it act like an attractive force. The second term is surely an attractive interaction. Since the repulsive effect among the molecules occurs when the atoms are very close to each other (ρ in Eq 3.10), the authors wisely used the repulsive formula to define the interaction between the cities and the points. The K parameter states the maximum distance between a city and a point to initiate a force (much stronger) between them.

Another application of elastic networks other than TSP was introduced by Shams in 1996 as a solution for the problem of localizing multiple targets(Shams, 1996). In this study, in order to solve the problem of localizing the position of multiple targets as they travel inside a surveillance area based on the bearing angle of info received by an array of S sensors, the author suggested a mathematical model similar to the one Durbin and Willshaw introduced. It is again a combinatorial optimization problem where the total amount of target is not known and there might be ghost formations according to the false sensor readings or multi-targets introduced at the same time (one sensor received signal from one of the target, while the other sensor gets from another target at the same time. If these signals intersect, a ghost formation occurs). The optimality occurs when all the targets are correctly labeled by selecting intersection points of different types (different intersection combinations from different sensors) which form a small cluster in space. In the system, nodes are the intersection points and a bistate neuron is introduced if an intersection point is associated with a certain target or not. Since there are lots of targets, there are also different groups of dynamic points (in this problem three point in each group formed a rubber triangle as given in Figure 3.19) associated with each target. These triangular rubber bands are attracted by the intersection points. As one triangle gets close to an intersection, the specificity of this point on this triangle increases by using temperature constant in simulated annealing method, resulting in the encircling of it. It is clear that the idea and the mathematical model is the same as of Durbin and Willshaw while the application scenario differs.
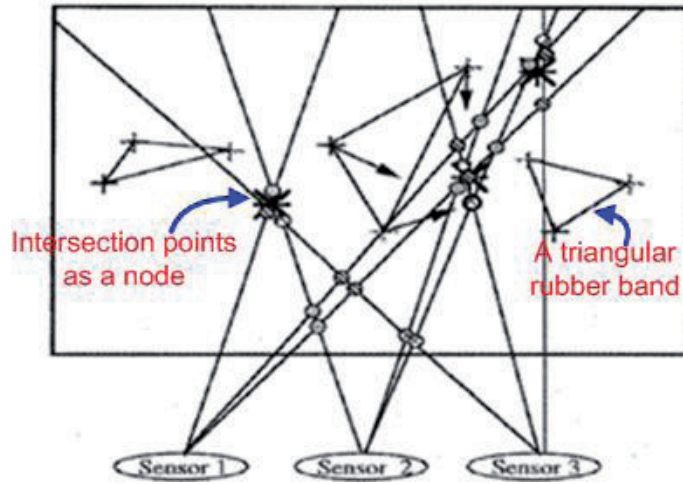
Figure 3.19 Three sensors and three targets are shown where targets are depicted by large asterisks and intersection points between pairs of bearing angles are depicted by textured circles with each texture represent different type of intersection. Three triangular elastic modules are introduced where each vertex associated with only one of the intersection types. The arrows show the direction of the dynamic elastic module nodes.

In the control field, the usage of elastic network is not very common. Although the method can be used in many different combinatorial optimization problems, in the literature, the only different examples are given above (even Shams (1996) uses the same formulation as the first approach of Durbin and Willshaw (1987)). The other studies as in (Boeres, de Carvalho, & Barbosa, 1992; Geng et al., 2011; Vakhutinsky & Golden, 1994) cover mainly different sensitivity analysis of the formulation and its parameters of the elastic network method again in the *travelling salesman problem* of Durbin and Willshaw (1987). Our approach on the elastic network is the same as the one of these researches. We have, as well, utilized this common model on our scenario with different structures of dynamic nodes, steady nodes and rubber line instead of rubber band, details of which are given in the proceeding sections.

69

3.4.2. Advantages of Elastic Networks

Significant advantages encouraged us to adapt this method to our application. Since neurons are sparsely connected in elastic nets, so that each neuron is connected to only two neighboring neurons, this method can scale well to overall increasing number of neurons. In addition, the K parameter makes the system move towards global minimum of energy (equilibrium point); therefore, the system avoids getting trapped by possible local minima. Indeed, since the system progress is biased on gradual increase of specificity, the global minimum of the energy function is guaranteed to be the optimum solution. These advantages are demonstrated on TSP (Richard Durbin & Willshaw, 1987) and a comparative analysis between elastic networks and Hopfield's neural network (Hopfield & Tank, 1985) is conducted (Roy, Sarma, Soumyadip, & Maity, 2013). The performance of elastic net is remarkable compared to Hopfield net which generated shortest contour perimeters 19% longer than those of elastic net. In addition, Shams in (Shams, 1996) tried to merge the energy minimization ability of Hopfield net and simulated annealing method on a combinatorial optimization problem as in elastic nets of Durbin (Richard Durbin & Willshaw, 1987); however, the system falls short of finding good solutions even with relatively small number of neurons.

In addition, our problem of intention reshaping is a framework on a dynamic map where there may be nodes newly introduced (newly detected intentions) on the map during the system works. Therefore, a classical search algorithm method as A* algorithm by Hart, Nilsson and Raphael (Hart, Nilsson, & Raphael, 1968), which takes the node map as a tree and search for the optimum node according to a cost function, will not be suitable for our purpose. Since all of the nodes are connected to each other as parents and their children, the tree structure will not be updated. As a result, an introduction of a new node to the system will not be recognized by the algorithm.

In our problem, since we are creating a trajectory of intentions consisting of way points both, connected to other way points (candidate intermediate intentions), as

well as the current and previous intentions in a space with large number of training intentions (neurons or nodes), elastic neural network provides a suitable framework to our applications. The details of our approach on elastic networks are given in the proceeding sections.

3.4.3. Statement of Purpose

As discussed in the previous sections, we are generating way points consisting of trajectories for the robots to follow. This generation is handled in two different modes of execution: *confident* and *suspicious.* Each mode requires different path planning as; the former one aims to generate a way point consisting of a trajectory directed towards the desired goal whereas the way point found in the latter one should be close to the current intention of the human subject in the dense area of observed intentions (familiar places to the person). From now on, we will analyze these requirements of two modes with the perspective of elastic networks.

Our elastic network tries to generate a way point in the previously constructed intention search space in Section 3.1.3. Before the system starts, a desired intention trajectory to be reshaped into is selected from this map of trajectories (called as nodes in elastic nets). In iterations, estimated current intention and previous intention of the person are labeled on the map. According to the execution mode, decided upon the estimated body-mood of the human subject, elastic network algorithm is started to generate a way point (a trajectory node) for the robot to follow. In the *confident mode*, found way point should be close to the desired intention which is a trivial problem. On the other hand, in *suspicious mode*, our aim is to generate a way point in the closest and most frequently sampled areas of the intention nodes around the currently detected intention node of the human which results in the main difficulty to be solved in our project. This area of most frequently sampled intention nodes consists of previously observed intentions from the training experiments. In other words, we are trying to guide the person through these familiar and close areas to the current location of the person to be able gain curiosity and trust.

71

Before we explain the created map, our approach and the formulations for the path planning with elastic networks, we should link our intention topology to elastic network search map and create an elastic network model. For this purpose, the next section explains necessary adaptations on the mechanisms discussed in the intention estimation in Section 3.1.

3.4.4. Adaptation of Intention Feature Space to the Elastic Network Model

In order to link our intention topology with elastic networks, we need to detail the adaptation of our intention feature space to elastic network search space. As it is detailed in the literature search of elastic nets (Section 3.4.1), elastic network model is applicable on topographical maps and simulates physically realizable forces causing displacements on nodes in 2-D maps. Therefore, our search map (intention feature space) should consist of intention feature trajectories as 2-D nodes. In our approach, we introduce trajectories used in training intentions as nodes creating the search space of elastic nets. To have our dynamic node successfully converging to a steady node (feature vectors) under the effects of these forces of elastic nets, resulting in the foundation of a way trajectory point, we needed to reduce our multi-dimensional learned feature trajectories into 2-D by averaging with a weight vector. For the purpose, each feature descriptors in size $[2xN]$ is multiplied by constructed weight vector in size $[Nx1]$, with N being the frame length a human subject is observed. It is better to recall that sequence length of trajectories, N, equals to 56 sequences (7seconds x 8fps for real-time). Resulting matrix after multiplication is in size $[2x1]$ and depicted as a 2-D node in our Elastic network system. We should clarify that each reduced node also keeps its original trajectory information because the chosen way point among them will be the trajectory to be followed by the robots.
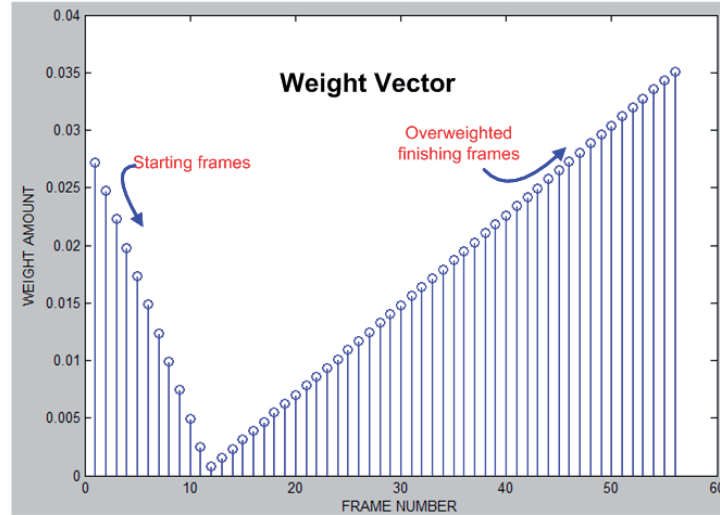
Figure 3.20 Weight vector with horizontal axis being video frame number and vertical axis is for weight amount. Averaging a trajectory of 56 sequences of locations with this weight vector over weights locations of the trajectory at the starting and finishing frames.

The key point on weight vector is, it puts bigger weights on starting and finishing frames of the trajectories (see Figure 3.20). Since our main aim was to plan trajectories for the robots starting around the current intention to gain curiosity and pointing out the desired one especially in *suspicious mode*, starting and finishing locations of the trajectories are crucial for the system. In addition, finishing locations of the trajectories are pointing out the places where its labeled intention takes place. Therefore, we emphasize the locations at the starting and finishing frames in the trajectories. In other words, averaging a trajectory of 56 sequences of locations with the weight vector given in Figure 3.20 increases the emphasis on the locations of the trajectory at the starting and finishing frames. An exemplary formation is illustrated in Figure 3.21 where a trajectory with 56 sequences of location leading to an intention of 'drinking coffee' given in part (a) is reduced to a two dimensional node circulated in the whole reduced map of 100 intention trajectories in part (b). In the figure, it is clear that exemplary intention trajectory was close to the coffee table because of over weighted finishing locations and not far away from starting locations

73

due to considerably big weights in starting locations. Finally, a deduction can be made from Figure 3.21(b) that, nodes close to the objects such as; coffee, table or library belong to the trajectories labeled with related intentions.
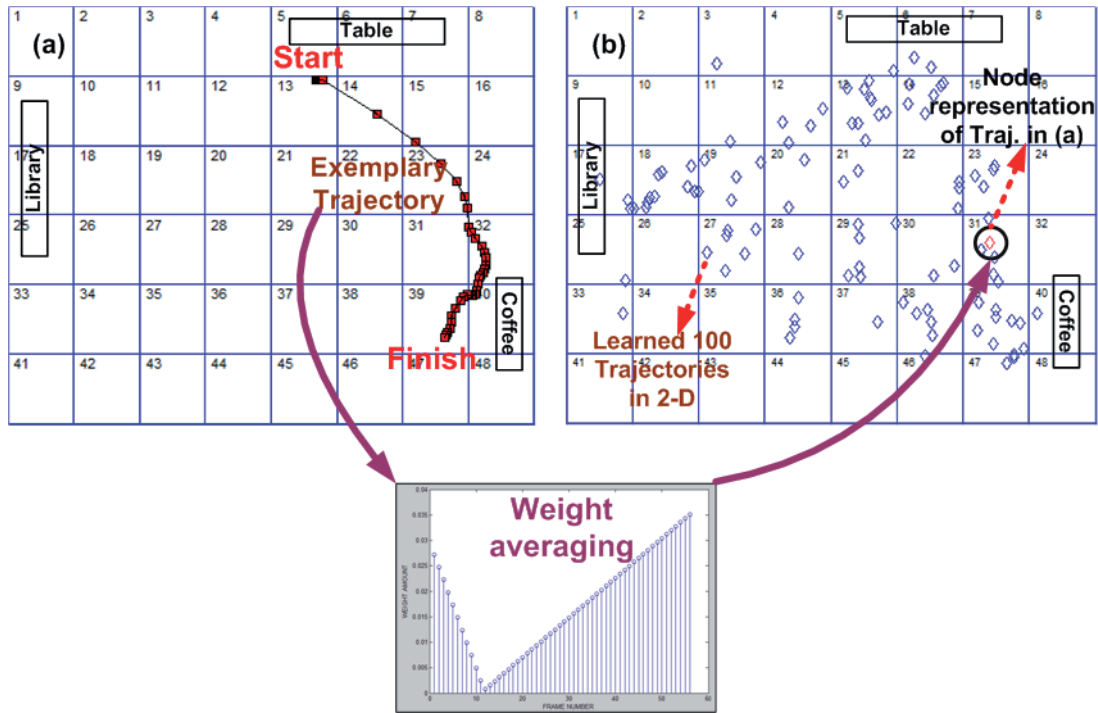
Figure 3.21 (a) An exemplary trajectory leading to the intention of 'drinking coffee' with starting and finishing locations are mentioned; (b) All of 100 trajectories after multiplied by the weight vector in Figure 3.20 are demonstrated as elastic network nodes in 2-D. Circulated node belongs to the trajectory in part (a). It is clear that this node is close to the end locations of the trajectory not far away from starting.

3.4.5. Our Approach on Elastic Networks

To better understand the method of planning way points and elastic network model created with nodes and forces, we are going to explain the formulations on exemplary drawings given in Figure 3.22. Starting with the introduction of nodes, current intention, previous intention and desired intention are the special nodes of the model whereas the other nodes are the trained set of intention trajectories, all of which are illustrated in Figure 3.22(a). Current intention is the estimated trajectory found in the Section 3.2.2.2 and previous intention is the trajectory observed starting from the first move of the robot and lasts 7seconds. We should note that, 7 seconds or 56 sequences with 8fps is crucial for the length of this trajectory because before the trajectory is introduced in the elastic networks search map, it is going to be averaged with the weight vector in Figure 3.20 having the size of [56x1]. Right after 7seconds passed, current intention and trajectory estimation starts again. Current trajectory is selected from the pool of trained intention trajectories; however, previous trajectory is the actual trajectory of the person followed between two successive current trajectory estimations. Therefore, previous trajectory is always a newly added node to the map at iterations. Finally, desired intention in the figure is the final intention aimed for the human to realize after reshaping actions. This intention is chosen by the user and it is a point directly showing the location of the desired intention decided. For example, if the desired intention is selected to be 'drinking coffee', desired intention node becomes the location of the coffee table.

In our mathematical model approach, we have a training set of intention trajectories in feature space is a set of n nodes $X = x_1, x_2, \ldots, x_n$ and one dynamic point y. At iterations, the current and the previous intention of the person and the desired intention are eliminated from the set of nodes X, and labeled separately as y_k, y_{k-1} and $x_{desired}$ according to iteration k. At the beginning of the algorithm, the dynamic point is introduced as the middle point on a rubber line connecting nodes y_k and y_{k-1} as shown in Figure 3.22(a). Then, this rubber line is stretched toward the next way point to be found by minimizing the energy function defined as:

$$E = -\alpha \left(K \sum_i^n e^{\frac{-\|x_i - y\|^2}{2K^2}} + E_{desired} \right) + \beta(\|y_k - y\|^2 + \|y_{k-1} - y\|^2) \quad (3.14)$$

where K is the length parameter that eventually tends to zero as in temperature constant in *simulated annealing method* (Richard Durbin & Willshaw, 1987; Geng et al., 2011; Kirkpatrick et al., 1983). The terms in Eq.(3.14) can be defined as such: the first one is an attraction of the dynamic point on the rubber line to the nodes in set X and to $x_{desired}$(with $E_{desired}$ to be stated later) while the second one is an attraction of the dynamic point to the current and previous nodes favoring the next way point to be found around the current intention. Exponential term in Eq.(3.14) states the influence of each node in set X, and K parameter in that term measures the maximum distance from a node to the dynamic node in order to create a strong attraction between them.

The attraction of the dynamic node to the final one is calculated separately in the additive term $E_{desired}$ as:

$$E_{desired} = K_{desired}e^{\frac{-\|x_{desired} - y\|^2}{2K_{desired}^2}} \quad (3.15)$$

where $K_{desired}$ is first initialized as:

$$K_{desired} = \|x_{desired} - y\| \quad (3.16)$$

but decreases in each iteration according to a predefined rate, as K in Eq.(3.14). Here $x_{desired}$ is the coordinates of the desired intention as mentioned previously. Initially, the exponential term in Eq.(3.15) acts as a constant and bigger than that of other x_i's in Eq.(3.14); however, as $K_{desired}$ decreases stepwise in iterations, this term decreases. This separate calculation enables initially the orientation of the dynamic node to be directed towards the desired goal. However, depending on the execution mode detected, if the person is *suspicious,* this effect is gradually but sharply decreased due to a bigger reduction ratio for $K_{desired}$ than that of K when tending to zero, resulting in the influence of other nodes taking over. On the other hand, if we

are in *confident mode,* this reduction ratio is set the same as of *K*, satisfying the biggest attraction is always towards the desired node and resulting in the dynamic node to approach or even end up on the desired one.

Since all other nodes are static except the dynamic one, the energy equation in Eq.(3.14) becomes the work done by the dynamic point. It is known that, negative work gives the kinetic energy in the mass and spring systems; therefore, minus the derivative of Eq.(3.14) multiplied by *K* ($E_{desired}$ is multiplied by $K_{desired}$), which is $-K\frac{\partial E}{\partial y}$, gives us the position change in the dynamic point at iterations as:

$$\Delta y = \alpha \left(\sum_i^n e^{\frac{-\|x_i - y\|^2}{2K^2}} (x_i - y) + e^{\frac{-\|x_{desired} - y\|^2}{2K_{desired}^2}} (x_{desired} - y) \right)$$

$$+ \beta K (y_k - 2y + y_{k-1}) \quad (3.17)$$

where the multiplication of *K* is for attractions that gradually vanish resulting in the stable convergence of dynamic node.

Each static node is applying a force; therefore, the superposition of these forces gives a strong attraction of the dynamic point towards dense areas (where a large number of nodes exist) directed to the desired node as illustrated in Figure 3.22(b) with bold red arrows. By looking at the figure it should be noted that, since our main goal is to direct the dynamic node towards the desired for either of the execution modes, only the attractive forces of static nodes on the dynamic one in the direction of the desired one are taken into account ignoring the others. Moreover, the attraction of the dynamic point by nodes y_k and y_{k-1} (yellow arrows in Figure 3.22(b)) favors the closest dense areas around the current intention. At the beginning of the process, these two forces of static nodes and current and previous nodes at opposite directions make the dynamic point oscillate between the current intention and the dense area of the nodes. In the limit where *K* tends to zero, the second attraction gradually loses its effect while the first exponential term in Eq.(3.17) has an increased specificity by *y* tending towards the closest node at these dense areas, thus lowering $\|x_i - y\|$ for low

K. At a *critical iteration*, which is the iteration that we contributed, the second force term, holding the dynamic point near the current node, becomes negligibly small allowing the dynamic node to end up on the closest node. Characteristics of the oscillations until the *critical iteration* are demonstrated and discussed in the results Section 4.2.1.

In conclusion, any change according to Eq.(3.17) in *y,* which is the dynamic point, results in a reduction of energy reaching the global minimum. In the limit where *K* tends to zero, in order to keep energy bounded on the global minimum, there must be as well a term $\|x_i - y\|$ tends to zero. Assuming that we are in *suspicious mode,* as illustrated in Figure 3.22(c), that node will be the next chosen way point of our concern which guarantees to be the closest node in the densest region of intentions (the region intentionally more familiar to the person) around the current one directed towards the desired node and called intermediate next intention in the intention trajectory. As for the *confident mode*, next chosen node is the desired final intention directly as demonstrated in Figure 3.22(d). Finally, found way node containing is transferred to our robots to be followed which is discussed in the next section being our final step of the methodology. More examples on the generation of way points for both modes are given in Section 4.2.3 with simulations and in 4.1.2 with real-time experimental applications.
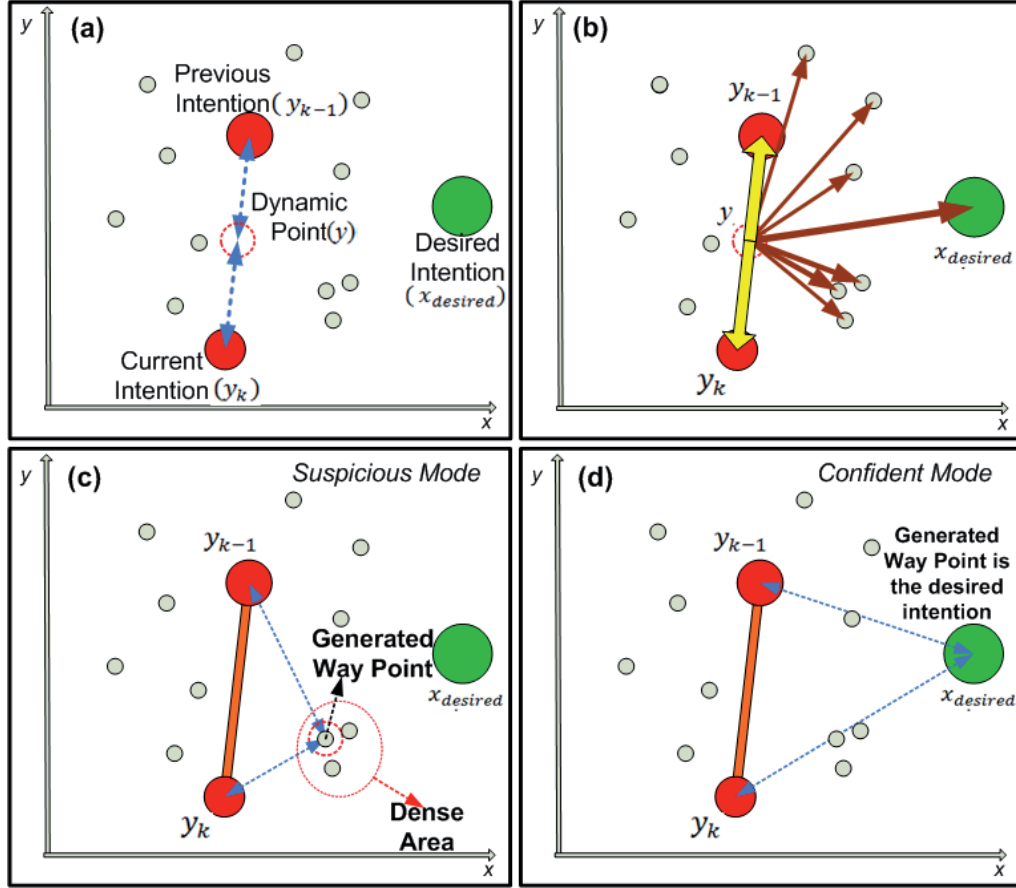
Figure 3.22 (a) Initialization of the path planning is shown. x and y are pixel coordinates. All nodes are intention trajectories reduced in dimension with weight vector in Figure 3.20; (b) Attraction demonstration. The bold red arrows are the attractive forces exerted by the nodes (which are in set X of Eq. (4)) together with $x_{desired}$ on the dynamic node. The yellow arrows show the attraction applied by previous and the current node; (c) Way point is found in the dense area in *suspicious mode* after that oscillations induced by the two forces in Eq.(3.17) vanish; (d) In *confident mode*, next generated way point is the desired intention node directly.

3.5. Autonomous Robot Moves

Found way point is actually a learned trajectory in the feature space of intention trajectories. This trajectory has 56 sequences each of which contains a pixel

coordinate on the image frame (recall that every node in the search map are learned trajectories projected to 2-D by averaging in Section 3.4.4). This information is transferred to our robots and they follow each location consecutively, ending up realizing the way trajectory found. An exemplary way point drawn on the image is demonstrated in Figure 3.23 with a green path. As soon as it is found, according to the current location of the robot, directive commands are transmitted via radio signals to make the robot move to the start location of the path and follow the rest. Here keeping final location constant (the last sequence in the way trajectory being 56th location), starting point of the trajectory is taken as the closest point on the trajectory to the person. The reason is that, since it is *suspicious mode* (as stated on the top-right corner of the figure); our purpose is to gain the curiosity of the person by getting close to him/her. In addition, the rest of the trajectory points, far away from the human subject would take the attention of the person to the places out of concern. In this example, desired intention is 'getting a book from the library' and the robot was trying to put the system into *confident mode*. Therefore, having 2-steps robot to go to the initial locations (near coffee table in the figure) drives the attention of the person to these places.
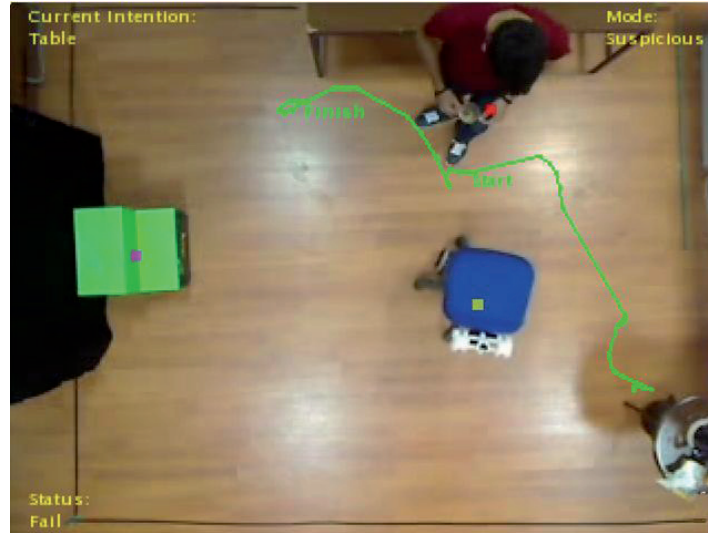
Figure 3.23 Green trajectory shows the way point found by elastic networks. This is the whole trajectory belongs to the way point found with elastic nets. However, start and finish locations that robot will follow are mentioned on the figure. Keeping the finish location constant (the last one in the trajectory sequence), starting point is taken as the closest point to the person for the purpose of gaining the curiosity. The rest of the trajectory is useless for our purpose.

Only one robot will realize the found way trajectory at a time. Therefore, in iterations we need to choose which robot to move. First of all, it is clear that our robots are contextual, meaning that they are one of the objects belong to the room context. Each robot can be used by the human subjects according to their intentions. As a result, we are using contextually the most appropriate robot according to the desired intention decided beforehand. For example, if we desired the intention of 'sitting on the table', from the beginning of the project until we claim success, the chair robot realizes all of the trajectories found by the elastic networks, gradually leading the person to the table and make him sit on the chair. For 'drinking coffee' intention, we also use the chair robot because it is more appropriate for the purpose than the 2-steps robot. May be the person wants to sit while preparing the coffee. If desired goal is chosen to be the library, it is obvious that the 2-steps robot takes the job and follows the path. As for the last remaining intention, 'discovering the environment', which is the easiest

intention to be reshaped into. This time, the closest robot to the starting point of the path realizes this trajectory.

As for the last two rules on robot moves, we should mention that, the next intention recognition (and trajectory recognition if the desired intention is not achieved, see Figure 3.1) is conducted after robots finish their movements. In addition, if the person interacts with the robot, the algorithm makes sure that this robot does not move at all until the interaction is ended. What we mean by interaction is a full interaction involving the person sits on the chair robot or climbs on the stairs robot (2-steps). The system understands this interaction if detected two centroids belonging to the person and the robot get nearly on top of each other. This rule is to ensure the safety of our experimenters

CHAPTER 4

RESULTS AND DISCUSSIONS

In this chapter, we firstly discuss results obtained from real-time experiments demonstrating the performance of the whole close loop intention reshaping system. We then conduct a detailed sensitivity of our system generating intention transients depending upon parameter changes.

4.1. Intention Reshaping Experiments in Real-Time Scenarios

In this section, we aimed to create robot actions analogous to a hungry pet desiring food from its owner. Experiments showed that our robots first try to gain the curiosity of the person by continuously getting closer to him/her, until the person's attention is caught and he confidently approaches the robot. Then robots reshape his current intention into the desired intention while continuously triggering the attention of the person by their moves. Any attention lost results again in trying to catch the attention of the person by the robot. Resultant performances of the system with snapshots from the experiments are discussed for several scenarios after giving brief information on real-time experiments.

4.1.1. Information on Real-Time Experiments and Recorded Videos

The same experimental room mentioned in Section 3.1.1 was also used in these experiments. Our aim was to test the close loop system we created on human subjects in real-time. Below are our milestones each measured and tested in these experiments:

- Estimating current intention of the person at all iterations, and comparing it with the desired one to claim success after a successful reshaping action

- Estimating the current trajectory of the person

- Detection of the body-mood of the person

- Planning a trajectory according (generating a transient intention) to current and previous trajectories, current body-mood of the person and the desired intention.

In order to check the consistency of the way point (transient intention) whether it is related with the current intention and body-mood of the person, we printed out all of these mentioned steps above on the recorded video during real-time experiments. As seen in the snapshots given in Figure 4.2, each corner has each up-to-date information being current estimation, execution mode (body-mood) and success status of the system. In addition, during the experiments, we popped up resultant elastic network map at iterations with current, previous, desired and generated way point are highlighted similar to the simulations given in Section 4.2.

For the scenario, we assigned some rules suiting the conceptuality of the project and exhibiting the performance of our system the best. These rules and their reasons are given below:

- Our experimenters shouldn't be informed about the project or context of the room beforehand. The reason is that, we need the experimenters initially to be suspicious about the room and the robots, so that we could make them confident, which is our first purpose in the project. In other words, we expected our human subjects to exhibit natural emotions so that our concept of reshaping intention reflects more realistic results.

- We should decide a desired intention among four possible intentions in the room, 'discovering the environment', 'drinking coffee', 'taking a book from library' and 'sitting on the table'. Nodes of elastic nets are the learned intention features trained in Section 3.1.3 and adapted to elastic nets in 3.4.4.

Among 100 trained intention features only the ones belonging to the desired intention and the intention of 'discovering the environment' are included in the elastic search map (total amount of 50 intention trajectories). Nodes of 'discovering' intention are taken into account in order to break the obstinance of the person making him/her have more externally-focused attention. On the other hand, nodes of the desired one are included for reshaping. The generation of remaining intention nodes belonging to the other two intentions may lead the robot guiding the person to wrong directions, therefore; they are excluded.

- All of the experimenters must stay within the black lines surrounding the room, which describes the ROI of the camera.

- Experimenters were informed not to move the coffee table and the desk suited in the room. Since they are detected as background information, any displacement on their places could result in the detection of them as moving objects (ghost formation) as detailed in Section 3.1.2.

Real-time performance of the main PC used in the experiments allowed us a frame rate of 8fps at maximum. However, this frame rate was enough for us to observe our challenges, system responses to these difficulties and enough to print the outcomes of the algorithms during video processing.

The timeline of our implementation for real-time experiments is given in the flow of a chart in Figure 4.1. A software interrupt was developed estimating the current intention of the person by using the last 24 sequences of location information at each frame. After the estimation, a comparison between the current and the desired intention is carried out. If a mismatch is detected, the interrupt returns a null value and the processing of that frame continues. However, if these intentions are equal to each other, the system claims 'Success' returning to 'Start' state, skipping any other algorithms by restarting the frame counter. After 'Success', the system is not terminated because we wanted to show how stubborn our robots are on their missions: even if the person realizes the desired intention and then acts with another intention, the suitable robot again tries to reshape the current new intention back into

the desired one. As the mismatch between the current and the desired intention continues, the algorithm starts with collecting location information from the human subject for 7 seconds. After the first 56 sequences are recorded and the trajectory formed is labeled as previous intention, the sequence counter is reset to re-start collecting data. This time, after 24 sequences of tracking the human subject, formed trajectory is saved and the current trajectory estimation is realized using these 24 sequences. Then, our shift comparator for current trajectory prediction starts working, and the best match at the 30th sequence is labeled as current trajectory (see Section 3.2.2.2). In the same iteration, body-mood detection of the human subject decides upon the execution mode. It was detailed in Section 3.3 that, execution mode is determined by how much the person gets close to the robot. Finally, elastic network algorithm takes necessary inputs shown in the figure and generates a way point (transient intention). Right after the generation of the transients, that takes approximately around 2 seconds, our system generates the solved elastic map on the PC screen. After a way point is found, the most suitable robot, which we call target robot, (see Section 3.5) starts realizing this trajectory, re-starting the algorithms.
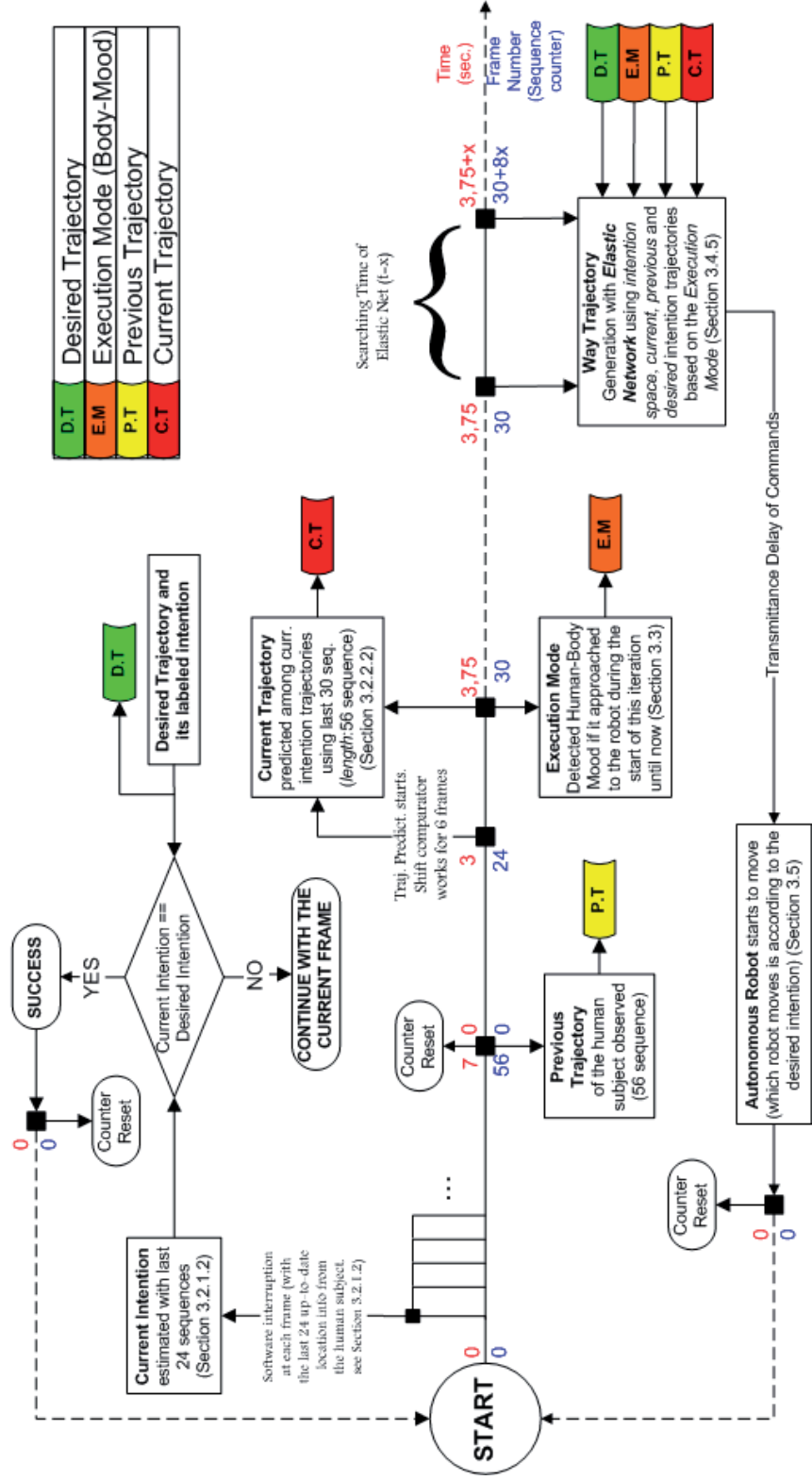
Figure 4.1 Close loop flow of the system used in human-in-the-loop experiments is given with timeline

4.1.2. Results and Discussions

In Figure 4.2, we gave sequential snapshots belonging to an experiment with a stubborn and initially highly suspicious person are provided in order to picture our system performance in both execution modes. In addition, we demonstrate in the same figure our results on intention and trajectory estimations, body-mood detection (execution mode) and the generation of transient intentions (way points) according to these execution modes. In each snapshot, the top-left corner depicts the estimated current intention of the person printed with their abbreviations given in Table 3.1 , whereas the top-right corner indicates the detected body-mood. Moreover, red trajectory drawn on the snapshots indicates the currently estimated trajectory that we expect human subject to follow, while the yellow trajectory is the previously observed trajectory that human subject actually did followed and the green trajectory is the found way point by our elastic network model. If a way point is newly generated in one of the figures, we give its elastic map at the right-hand side of the related snapshot. Finally, we need to mention that in this example all of the way points found were realized by the chair robot since the desired intention being 'sitting on the table' is related to this robot (see Section 3.5). Necessary explanations about the snapshots taken from certain phases of the experiment are given in the legend of the figure.

Starting with Figure 4.2(a), the system detected that the current intention of the person is 'coffee (drinking coffee)' which can be seen from the snapshot that the person is actually preparing a coffee. Since he wasn't interested in the robot at all, the system detected body-mood of the person to be suspicious and a way point was generated accordingly, which is demonstrated on the elastic map given in the same part. The starting location of the found trajectory's being near the person indicates that the system aimed to break the obstinance of him on coffee. From the aspect of elastic network model, since the nodes on the map are the trajectories weighted average emphasizing ending and beginning locations (ending points more weighted), a node between two intentions is likely to be a trajectory starting from one through

another (see details in Section 3.4.4). Since we excluded intention nodes labeled with intentions other than the desired one and 'discovering', found way point between the location of coffee and the desired one is a trajectory starting from the coffee table towards the desk. In part (a), the snapshot was taken just before the way point (trajectory) found was realized by the chair robot. In the next frame, the robot starts to move and a new loop was initiated (see Figure 4.1). In Figure 4.2(b), we jumped to the end of the loop where a new way trajectory was found. Blue trajectory in the snapshot illustrates the approximate path robot had followed between the frames in part (a) and part (b). It is clear that, the person grabbed his cup of coffee and stepped away from the robot *suspiciously,* where this mood is correctly detected again by the system shown on the top-right corner of this snapshot. In addition, the current intention of the person estimated correctly as 'discovering the environment'. Since the robot could not gain the curiosity of the person still, a new way point found shown by green trajectory close to the current trajectory of the person predicted (red trajectory) regardless of how far it was to the desired intention location.

In Figure 4.2(c), we aimed to demonstrate our current trajectory estimation performance. This frame was taken right after the robot started to move to realize the newly found trajectory in part (b). By comparing two snapshots in part (b) and part (c), the person actually followed our estimated trajectory with red color in part (b). Predicting the current trajectory correct increases the chance of the robot to approach the person correctly because each way point in *suspicious mode* is found close to the predicted current trajectories. Indeed, a successful approach to the person was realized as given in Figure 4.2(d). By comparing the locations of the human subject in part (c) and part (d), we can see that the person stepped back anxiously due to approaching robot towards him. This is actually the first interaction of the robot and the person in this experiment and again he was not comfortable with the robot which was also successfully detected by the system (printed *suspicious mode* on top-left corner).

Since the person kept his distance between him and the robot, again a new way point found in Figure 4.2(e) right after a current trajectory predicted. According to the

generated way point, the aim is still gaining the curiosity as in Figure 4.2(a) with the starting point being close to the person and finishing point showing the desired location. The next part being Figure 4.2(f) is given in order to reveal another correct prediction of the current trajectory which is drawn with red color in part (e). Frame in part (f) was taken while the algorithm still collects location data to construct previous trajectory node while the robot had already started to move (between 0-56 frames in Figure 4.1). Since current intention estimation is being realized at each frame (at all iterations), the system found out that current intention changed to 'Library'; however, a new trajectory prediction was not handled yet in part (f). In the snapshot at the left-hand side of Figure 4.2(g), a new current trajectory was predicted according to the current intention being 'Library'. At the same time, the robot was still realizing the previous trajectory found in part (e); however, the person did not care about the robot again. Therefore, a new way point was found in *suspicious mode*. Since our experimenter was very stubborn on his intentions, two more trials passed trying to climb the 2-steps robot while the chair was still trying for his attention.

Finally, in the snapshot given at the right-hand side of part (g), chair robot followed a trajectory close to the person. Realizing this trajectory resulted in the person closely examining the robot, this time not anxiously. As seen at the top-right corner of this snapshot, our system was successfully detected this interest and his approaching the chair robot claiming for *confident mood* for the emotional body-mood of the person. As it was stated for the elastic network model, once the execution mode switched to *confident mode,* our model does not decrease initially high attraction of the desired node ($K_{desired}$ in Section 3.4.5) and expects its dynamic node to end up on the desired one, which is the case given with the elastic net map in Figure 4.2(h). The way point found was shown on the snapshot with green trajectory which is a constantly located in front of the working table, directly showing the intention of 'sitting on the table'. This snapshot was taken right after the robot started to move. While it was moving, our experimenter was still examining the robot and observing its actions just like we expected. This is actually a proof for our idea that, once we gain the curiosity, the person becomes interested in the robot and starts to observe its

actions as in the case on this snapshot in part (h). Another breaking point on this part is that, continuous intention estimator had already estimated that the person was going to the table (see top-left corner on the snapshot) changing the status of the system to 'Success'. As it was stated in Figure 4.1, as long as the system keeps claiming 'Success', observation of previous trajectory, estimation of current trajectory, and planning of a way point are bypassed by the algorithm. At the snapshots in part (i), chair robot finished its trajectory and the person sat on it. From now on, system only estimates the current intention of the person and compares it with the desired one. If a mismatch occurs (in this example, if the person stands up moves away from the table), the algorithm again tries to reshape the intention of the person into the desired one ('sitting on the table') stubbornly.

Figure 4.2 Starting from (a) to (i) important moments are demonstrated with sequent snapshots. Desired intention was 'sitting on the table'. Reactions of the person and appropriate robot moves are explained for both two execution modes on each part
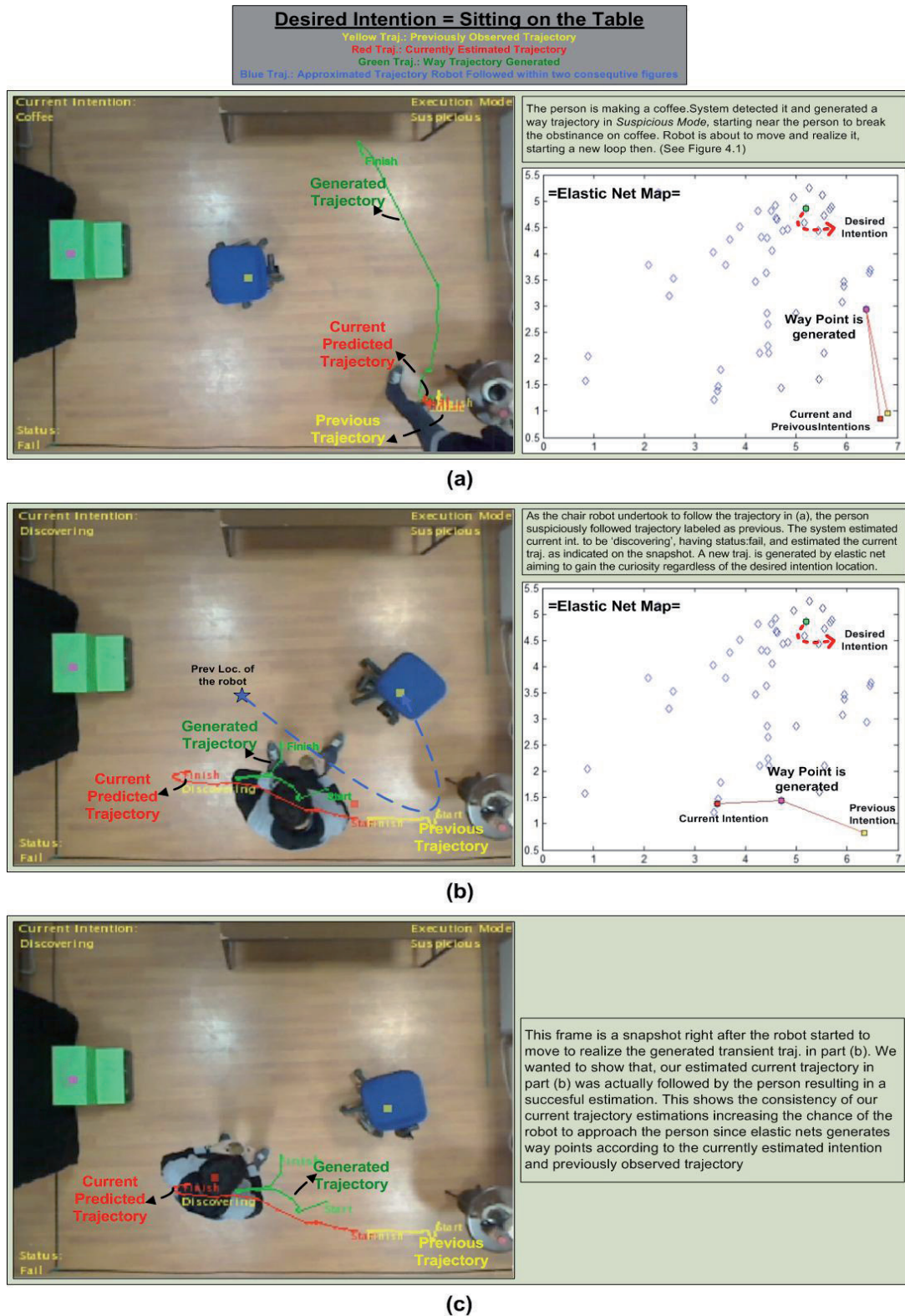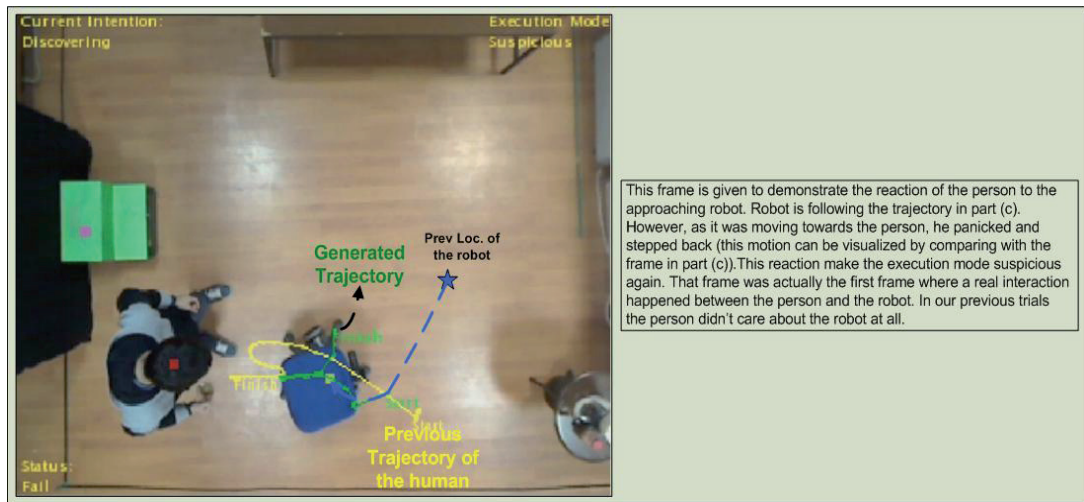
(d)
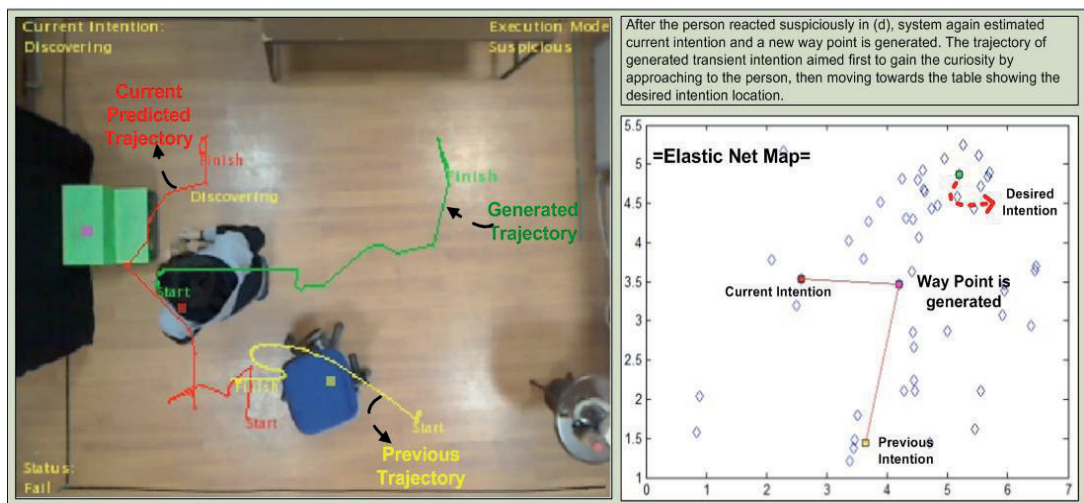
(e)
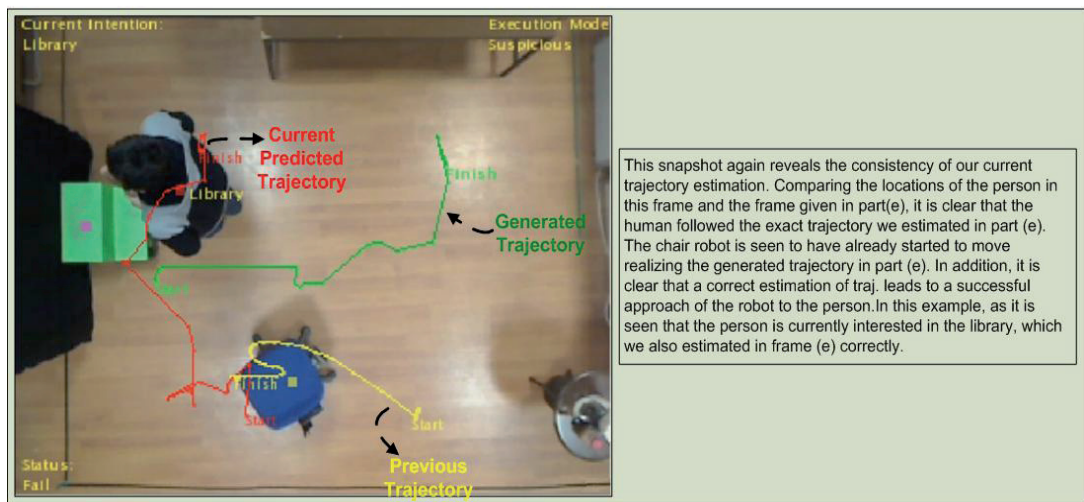
(f)

Figure 4.2 Continued

While the robot was following the green trajectory generated in part (e), the person didn't care about the robot at all. As a result, again a *suspicious mood* was detected on the person making the system generate a new trajectory aiming to gain the curiosity of the person in the next iteration as seen in the left-hand sided snapshot below. At the next trial he person was still interested in 2-steps robot trying to climb it to reach the library. After two more trials, a trajectory was generated again near the person as seen in the snapshot at the right-hand side. This time the person headed towards the robot and started to examine it. At that moment, our system detected the interest of the person and detected the body-mood of the person to be *confident*, switching the execution mode to *confident mode* (it was stated at the top-right corner).

(g)

After *confident mood* detected on the person, elastic networks directly found the desired intention node. This node is a constantly located trajectory as given in the snapshot at the left-hand side. Robot was moving towards this location while the person was still curious about the robot as expected. In addition, intention estimator had already found that the person was going to realize the intention of 'sitting on the table', claiming success accordingly.

(h)

In the snapshot at the left hand side below, it is clear that the system estimated the current intention to be 'sitting on the table' before the person actually realized it. Therefore, we claimed success as printed at the bottom-left corner of this snapshot making the algorithm follow the 'success' path given in Figure 4.1. In the right-hand sided snapshot, the person actually sat on the table, which approves the accuracy of our system. From now on, the system only estimates the current intention of the person and compares it with the desired one. If any mismatch occurs, the system will again generating a way point for the robot to reshape the intention of the person to the desired one (sitting on the table).
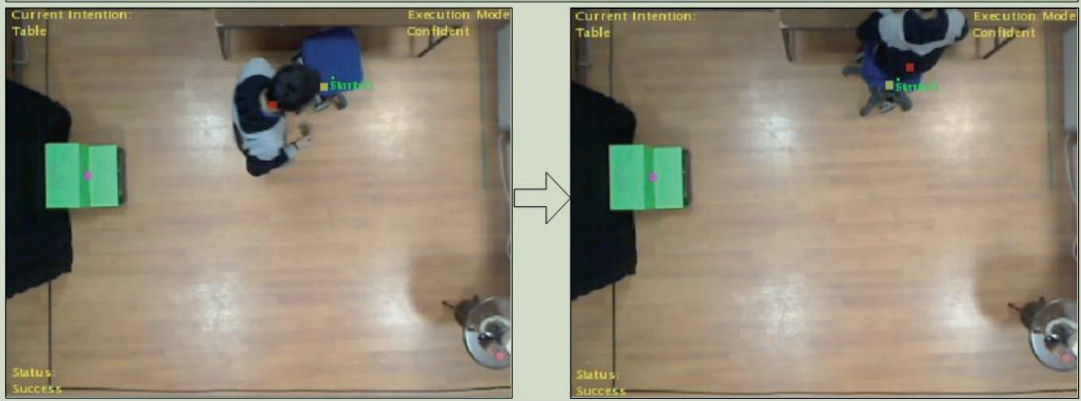
(i)

Figure 4.2 Continued.

94

In another example given with Figure 4.3, we tried to test our system with two serial desired intentions. For the purpose, we created a scenario in which the first desired intention was 'getting a book from the library', as soon as the person realizes it switching to the second one being 'sitting on the table'. In Figure 4.3(a), 'table' intention was estimated on the person and a trajectory was found for the 2-steps robot (it is now the target robot since the desired intention is related with the library) detecting that the person is *suspicious*. In that experiment, the experimenter was more confident with our robots comparing the experimenter given in Figure 4.3; therefore, after 2-steps robot realized its trajectory, he started to examine the robot curiously, making the execution mode *confident* as shown in part (b). If a *confident mood* detected on the person, elastic network model directly generates the way point on the desired intention node as shown in the elastic network map figure in part (b). The trajectory that 2-steps robot should follow were drawn on the snapshot with green color.

As seen in the next snapshot at the left-hand side in part (c) of Figure 4.3, experimenter followed the robot as we expected and started to climb on it to reach the library at the top. Since the system was able to estimate current intention to be 'getting book from the library', status had turned to 'Success', claiming the success of the first state of this scenario. Then, the system switches to next desired intention being 'sitting on the table'. The right-hand sided snapshot of the same part shows that the status changed to 'Fail' again since now the current intention was not equal to the new desired intention. Therefore, a new trajectory was found as given in the elastic map and it was followed by the chair robot this time. The reason was mentioned in Section 3.5 that, the intention of 'sitting on the table' is related with chair robot. Finally, in the left-hand sided snapshot of the last part (d), we directly gave the moment, when the person approached to the chair robot confidently, skipping two other trials in between. The algorithm was able to detect this mood of the person, switching to *confident mode*, where elastic net model directly found the desired node. As a result, we were able to make the person realize both two desired intentions sequentially (see 'Success' status on the right-handed snapshot in part (d)), and made him read the book he took from the library on the work table.

95

These two experiments, with an obstinate and initially highly suspicious person and less suspicious person comparing with the first one, showed the accuracy of our methodology realizing all of our system requirements and expectations mentioned previously. We successfully made the experimenters confident with our robots, made them approach to the robots confidently (gaining their trusts), by exhaustively following found way points from the intention feature space, which has previously learned trajectories of human subjects gathered from training experiments. Then, for the second phase, we were able to reshape their intentions into what we desired again by following a learned trajectory leading to that desired intention, proving and finalizing our concept of thesis work.
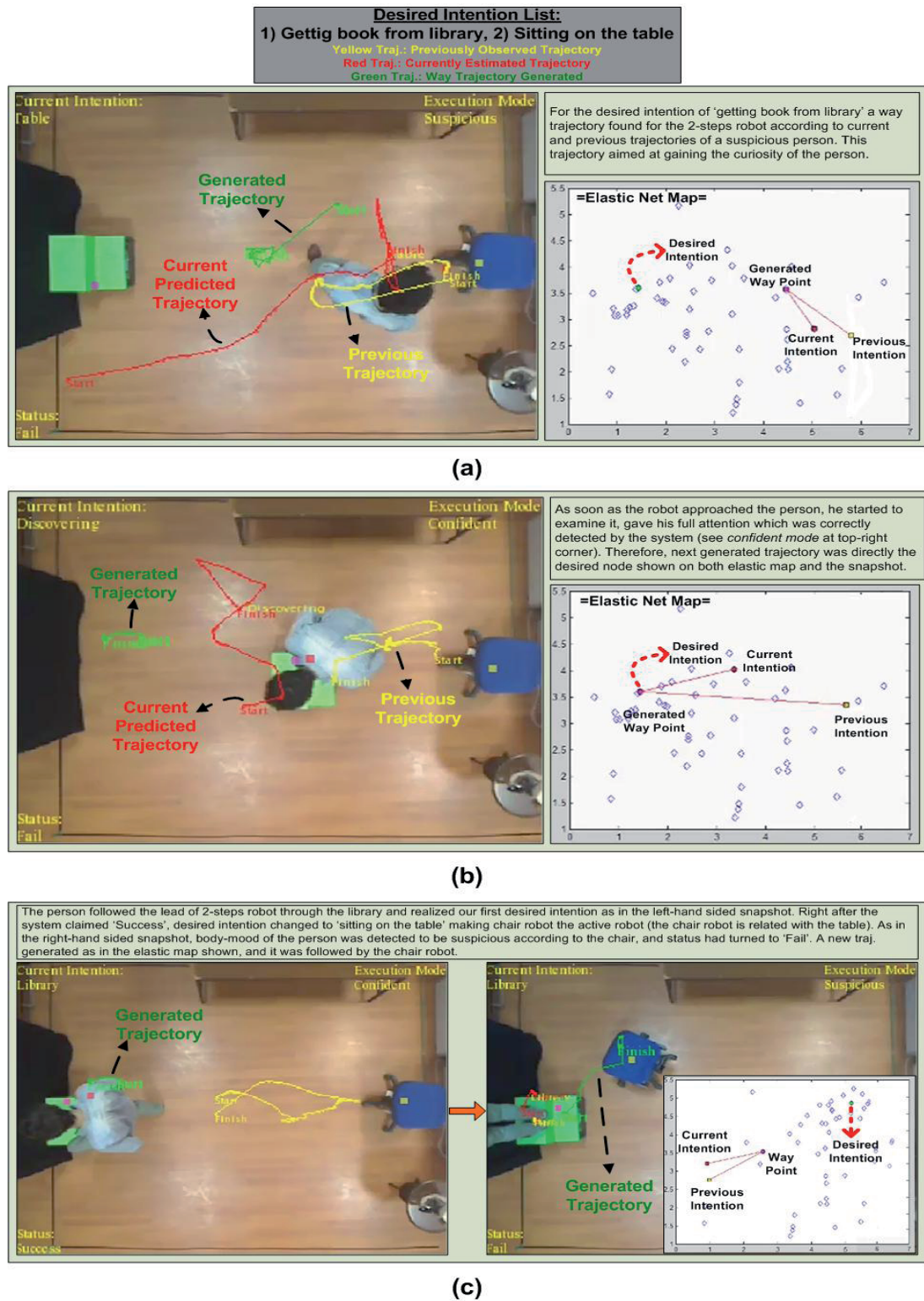
Figure 4.3 In this experiment, we assigned two sequential desired intentions 'getting book from the library' and 'sitting on the table'. After the person realized the first desired intention related with the library, 'sitting on the table' becomes the new one. Reactions of the person and details are drawn and explained on each part.

After two more trials, the person approached to our chair robot and started to examine it as in the left-hand sided snapshot, and the system detected *confident mood*. Since the desired intention is about the table, elastic network found that intention as shown with the green trajectory being a point in front of the table. Then, the chair realized this trajectory and the person followed it as we expected. In the right-handed snapshot a correct estimation of the current intention being 'Table' resulted in the change of the system status to 'Success'.

(d)

Figure 4.3 Continued

4.2. Analysis of Elastic Network Performance Dependency to Parameter Changes: Sensitivity Analysis of Generation of Intention Transients

In elastic network model proposed in (Richard Durbin & Willshaw, 1987), the authors tested system performance according to parameters of the force equation given with Eq.(3.11) and Eq.(3.12), which are α, β and K, with 100 nodes on a map of unit square. The authors stated that parameters change depending on the size of the map constructed due to the fact that the energy function is calculated according to the distances between nodes (see Section 3.4.5). For comparative purposes, we also conduct a sensitivity analysis on these parameters within the context of new intention generation for sequential reshaping. We also analyzed the energy of the system to check whether the optimality is ensured by our modification of elastic networks and lastly we conduct tests to analyze the performance of the system towards satisfying requirements for path planning once an intention transient is optimally generated.

98

4.2.1. Sensitivity Analyses

We began our tests by using the same parameters of Durbin et al. (Richard Durbin & Willshaw, 1987) where α is 0,2, β has the value of 1, K is initially 0,1, and the reduction ratio of K is 1% in every 10 iterations, with the total amount of 7000 iterations (reduced to 2500 after energy function analysis, detailed in Section 4.2.2) in our force equation with randomly but homogenously placed 100 nodes. We use 100 nodes because our intention feature space has 100 intention features (trajectories) as trained in Section 3.1.3 and scaled to 2-D for elastic nets in Section 3.4.4. We use random nodes only for testing purposes to create an elastic net model generic to any node distributions. Elastic network map has the size of [8x6] decreased proportional to our actual ROI of the camera deployed in the experimental room being [320x240] in order to decrease computational cost and increase CPU performance. An exemplary view of the map of the elastic net is given in Figure 4.4. In addition, as a reminder, we give our energy (Eq. (4.1)) and force (change rule of the position of dynamic node with Eq. (4.2)) functions of elastic nets below:

$$E = -\alpha\left(K\sum_i^n e^{\frac{-\|x_i-y\|^2}{2K^2}} + K_{desired}e^{\frac{-\|x_{desired}-y\|^2}{2K_{desired}^2}}\right) + \beta(\|y_k - y\|^2 + \|y_{k-1} - y\|^2)$$

(4.1)

$$\Delta y = \alpha\left(\sum_i^n e^{\frac{-\|x_i-y\|^2}{2K^2}}(x_i - y) + e^{\frac{-\|x_{desired}-y\|^2}{2K_{desired}^2}}(x_{desired} - y)\right)$$

$$+ \beta K(y_k - 2y + y_{k-1}) \qquad (4.2)$$

where initially $K_{desired} = \|x_{desired} - y\|$ being the distance between desired and dynamic node at the initialization (see Section 3.4.5). As for the introduction of our simulated maps, Figure 4.4 gives the state trajectory of the dynamic node as the elastic net solves for a way point (transient intention) in the intention space. We should note that, all of the exemplary simulations given in this section were executed

in *suspicious mode.* That is, reduction ratio for $K_{desired}$ in Eq. (4.1) and Eq. (4.2) in such a mode is 0,5 which is five times bigger than the one of K, decreasing the effect of desired node sharply as iteration advances. That way, the dynamic node is not strongly pulled by the desired node and ends up on a closer node to the current intention node as seen in Figure 4.4(b) (see more details for the way point generation in *suspicious* and *confident mode* in Section 3.4.5). For the other mode, *confident mode,* the reduction ratio of $K_{desired}$ is 0,1 which is the same as for K while the other parameters are kept constant. This time the attraction of the desired node, which is initially the biggest, remains stronger than the other nodes resulting in the dynamic node ends up on the desired intention node (see Section 3.4.5). Illustrations of these scenarios for both modes are given in performance analysis in this section. Blue diamond shapes represent the intentions that occurred in one of our experiments where the dotted circle, which is inflated in the right corner of Figure 4.4(b), includes the previous and the current intention along with the dynamic node. In addition, the desired final intention is highlighted with green diamond shape. In each of the snapshots given in Figure 4.4 and in other simulations included in this section, the trajectory of the dynamic node searching for the way point (an intention transient) is represented by a series of black dots. Oscillations in the iterative solutions yield branching of the trajectory. As the iterations advances, the spacing between black dots diminishes showing the occurrence of convergence and the trajectory becomes a bolder black colored line. The final solution is the way point represented by a large purple square. The *critical iteration* that was introduced in Section 3.4.5 is given explicitly in the inflated region in Figure 4.4(b). This is the iteration when the oscillations of the dynamic node end having this node to converge to the closed intention node in the intention space. In other words, this iteration occurs when the value of K parameter becomes small enough that the second force becomes negligibly small comparing with the first one in Eq. (4.2) which results in the first force taking effect. Moreover, since the force applied by a node increases as dynamic node approaches to it (increase in specificity), dynamic node ends up on the closest one eventually. It should be noted that, *critical iteration* depends on the distance between the nodes and the dynamic one in the intention map; therefore, it changes

100

depending on the scenario or the distribution of the dynamic nodes and the location of current, previous and desired nodes.

Comparative analysis of the effects of the α and β parameters: These parameters are for the relative strengths of the two forces introduced in Section 3.4.5 with Eq. (3.17). The first parameter α, strengthens the first force which the steady nodes apply on the dynamic point whereas β is for the second force of current and previous nodes on dynamic point. The first force pulls the dynamic point towards the candidate way points on the dense areas with desired node over weighted, while the second one favors the closest nodes by pulling this dynamic point back to its initial location between the current and previous nodes.

The response of our model in *suspicious mode* with the parameter values given at the beginning of this section ($\alpha = 0,2$, $\beta = 1$, K is initially 0,1, the reduction ratio of K is 1%, $K_{desired}$ is initially the distance between the location of desired node and the dynamic node with the reduction ratio of 5% (*suspicious mode*)) is demonstrated in Figure 4.4(a). This figure demonstrates that, the force applied by the steady nodes (the first force of Eq. (4.2)) is not big enough to break strong attraction between dynamic node and the current and previous nodes. Therefore, the dynamic node could not reach further locations on the map. This yields the coverage of a small area encircled in the figure. The black dots in the scanned area are showing the previous locations of the dynamic node which gets bigger as the iteration proceeds. In order to break stronger effect of the forces of current and previous nodes, we need to increase the α value. Interchanging the values of these parameters, by having $\beta = 0.2$ and $\alpha = 1$, we can then make the dynamic node move to further locations covering larger area and converges to a node which is shown in Figure 4.4(b). It is important to note that, when the *critical iteration* is reached or when K parameter is lowered enough, the force of static nodes in first term in Eq. (4.2) takes effect, the oscillations ends and therefore the dynamic node is dragged towards the closest node on the map grid. Therefore, our aim must be to make the dynamic node move towards further locations increasing the coverage area of oscillations prior to convergence.
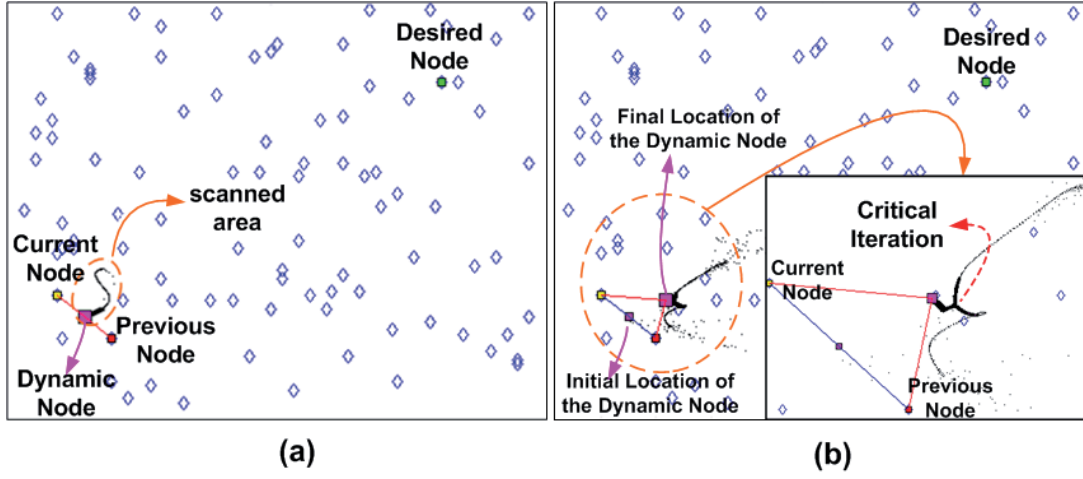
Figure 4.4 Analysis on α and β parameters. (a) System response with $\alpha = 0.2$ and $\beta = 1$ with initial K is 0,1 with the reduction ratio of 1% and initial $K_{desired}$ is the distance between the desired and the dynamic nodes with the reduction ratio of 5%. Scanned area represents the locations dynamic node was able to reach. It is clear that forces of current and previous nodes are strong enough to pull dynamic node back to its initial location; (b) System response with $\alpha = 1$ and $\beta = 0.2$. Steady nodes were strong enough to pull the dynamic node towards dense areas, having it ending up on the closest node in the closest dense area as K value decreases. These parameter values fully satisfied our requirements.

As for the analysis of the β parameter, any further increase on its value (Figure 4.4(b)) results in the dynamic node to be strongly pulled back to its initial location without possible enlargement of the convergence area. If we increased both parameters for testing purposes within the range of 1 to 2, we had a dynamic node oscillating out of our map, exceeding its size. We conclude that, any further increase of the values given in Figure 4.4(b) has the dynamic node exceeds the map or decreasing these values makes the dynamic node to be pulled strongly back to its initial location.

Analysis of the effects from changes of the K parameter: This parameter is the most crucial part of the created elastic network model. The decision of the initial value of K should be made according to the map. As clearly stated in the formulations, K

102

parameter both sets the effect of the second force and the maximum distance between a steady and dynamic node. The amount of change in the initial value of K decides the oscillation range of the dynamic node. If we have a bigger value for K parameter, the response of the system (the forces) is much stronger as in Figure 4.5(b). However, since we are lowering K as the iteration advances, these higher force values will gradually be lighter and again the dynamic node will end up oscillating in an intensive region. At a critical value of K, when the *critical iteration* is reached, again the system drags the dynamic node towards the closest node. Briefly, we may increase the K parameter to a value which is low enough to see its critical value. A comparative example with $K = 0.1$ and $K = 0.5$ is demonstrated in Figure 4.5 with the same α and β parameters on the same map as in Figure 4.4(b). The higher the K value, the stronger are the forces applied on the dynamic node causing further oscillations of this node. As K value decreases, the oscillations taper off to settle within a closer and the mostly visited region of the initial location of the dynamic node. In the inflated region at the bottom-right corner of Figure 4.5(b), it is shown that the *critical iteration* of K parameter can be reached before the iteration ends. As stated in Section 3.4.5, once the critical value can be reached in the generation process, the second force term in Eq.(4.2) pulling the dynamic node towards its initial location becomes negligibly small resulting in the dynamic node ending up on the same steady node as in Figure 4.4(b).
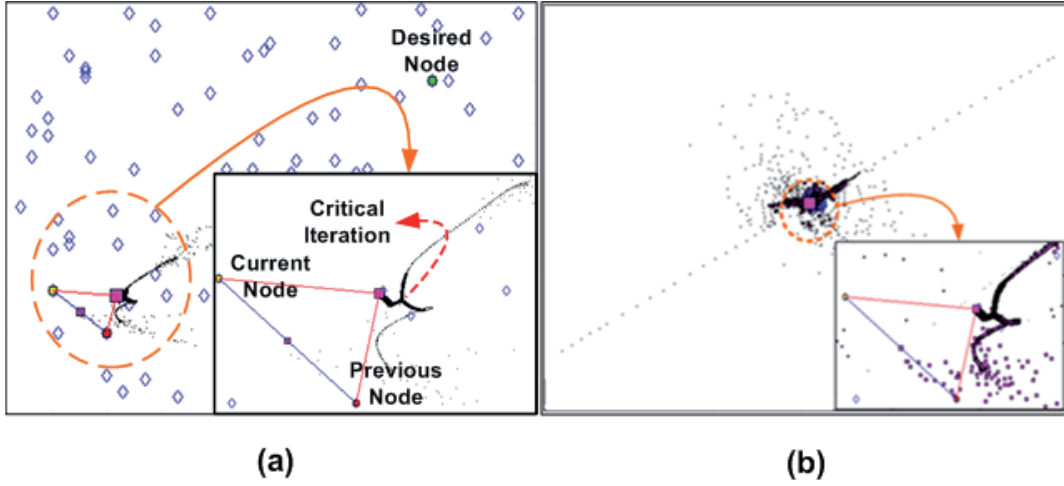
Figure 4.5 (a) Keeping all of the parameters except *K* the same as in Figure 4.4(b), the system response with the initial *K* being 0,1; (b) The response of the system when *K*=0,5. Dynamic node oscillates exceeding the size of the map due to the strong forces initially applied with higher K value. As the iteration advances, K value decreases settling down the system response and resulting in the dynamic node to oscillate within closer intensive regions to its initial location. In the inflated region at the bottom-right corner, it is shown that the critical iteration of K parameter could be reached before iteration ends. As a result, dynamic node ended up on the same steady node as in Figure 4.4(b).

By looking at Eq.(4.2), *K* value decides how big the forces will be and how far the oscillations of the dynamic node occur away from the initial location (detailed in Section 3.4.5). As for the lower limit of the *K* parameter, we should have it big enough to make dynamic node oscillate under the effect of stronger forces (as seen in Eq.(4.2), both force terms are proportional to *K* value) reaching to further locations from its initial location. For the reduction ratio of *K*, the upper limit should be small enough not to have the second force in Eq.(4.2) vanish before that the dynamic node finds the dense convergence. That is to say, if the *K* parameter reaches the critical point fast, whenever the dynamic node is still far from the current and previous nodes, the dynamic node approaches a steady node far away from the current intention. However, we want the dynamic node to approach a node closer to the current intention node in *suspicious mode.* On the other hand, the lower limit of the

reduction ratio should be big enough to satisfy the critical value for K, that is, K should lowered down to its critical value before iteration ends (*critical iteration*).

4.2.2. Energy Analyses of the Elastic Network Model

Energy function of our system in Eq.(4.1) is the work done by the system. The spring forces applied to the dynamic node by all other nodes do a negative work on it. That is, increase in the forces results in a work which decreases the speed of the dynamic node, meaning that kinetic energy of it decreases. Therefore, it is clear that the energy equation calculated belongs to work of the system done on the dynamic node and negative of this work gives kinetic energy of our node. We used this kinetic energy in optimality analysis of our system, or to check if we could reach the global minimum which is the optimal solution as explained in Section 3.4.2. Kinetic energy of dynamic node in the scenario given with Figure 4.4(b) is demonstrated in Figure 4.6(a). The oscillations occurred at the first iterations, which is inflated at the top-right corner of Figure 4.6(a), are due to the higher K values resulting in stronger spring forces holding the dynamic node (second force term in Eq.(4.2)). In other words, while the springs between the dynamic and steady nodes are pulling the node (first force of Eq.(4.2)), the other springs of the steady current and steady previous nodes holding the dynamic one are stretched (see details in Section 3.4.5). At one point, this stretch becomes so high that the dynamic node will be pulled back strongly towards its original locations this time increasing the forces of steady nodes. These oscillations get smaller as the spring forces decrease with decreasing of K. However, they exist until the critical value of the K parameter is reached, which is shown on Figure 4.6(a), making the second spring force negligibly small. An example to better understand the analogy between oscillations and K parameter is given in Figure 4.6(b) where K remained constant. An analogy can be made between this response of the system in part (b) and an inverted pendulum in a frictionless environment. Both of the systems don't have any external force or effect which makes the moving node settle down eventually.
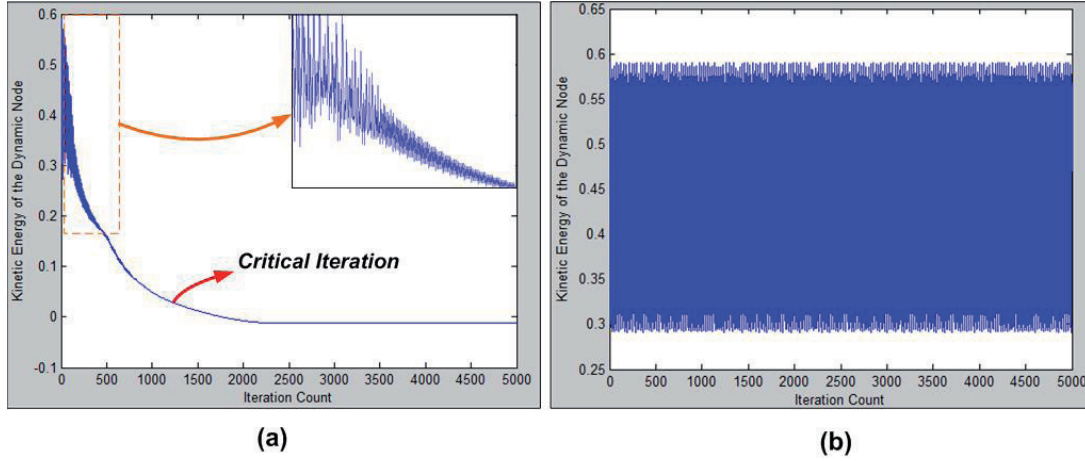
Figure 4.6 (a) Kinetic energy of the dynamic node in the scenario given in Figure 4.4(b) vs. iteration count is demonstrated. Oscillations were high at the beginning of the system due to higher K value resulting in stronger spring forces on the dynamic node. As the iteration advances, system moves towards the global minimum where the energy of the dynamic node becomes nearly zero; (b) An exemplary system response to constant K parameter is given. This condition resulted in constant oscillations. Since there is no external effect (friction) in the system, dynamic node behaves like an inverted pendulum in a frictionless environment.

After the *critical iteration* is reached, energy of the dynamic node converges to zero. By looking at Figure 4.6(a), it is clear that global minimum of the function happened between counts 2000 and 2500, then the system settled down, energy of the node became nearly zero meaning that dynamic node converged to its found closest node in Figure 4.4(b). An inference can be made that, in our applications 2500 iteration will be enough to reach the global minimum. More iteration will result in unnecessary computation and time delay on the system. To conclude, we verified that any change according to Eq.(4.2) in $y,$ which is the dynamic point, results in a reduction of kinetic energy of the dynamic node reaching the global minimum with the help of decreasing K parameter as in *simulated annealing method* (see the details in Section 3.4.1.2 and Section 3.4.5).

4.2.3. Performance Analyses

Our performance norm was the ability to generate a way point closer to the current intention in the dense area of intentions in *suspicious mode.* To evaluate this performance of the system, we made comparative analysis on a scenario, by changing location of the nodes and re-initiating the system. To start with, in Figure 4.7(a) and (b) represents two different scenarios having current, previous, desired and all other nodes at the same locations, except one steady one. In Figure 4.7(a), found way point can be clearly seen with the nodes placed. We claimed that this way point should be the closest point to the current and previous node in a dense area of intentions. To test the related performance, we represented a case, where just after way point is found in Figure 4.7(a) in a prior trial, the human without being influenced by any robot move, location of a node (node in red circle) in the intention map has changed due to a wrong estimation of this intention node (as represented by dragging the node to the location a red arrow points out in Figure 4.7(a)). The elastic net reinitiating its iteration, replacing the dynamic point to its initial location shown in Figure 4.7(a), re-plans a trajectory for the dynamic node as in Figure 4.7(b) and the new way point was found to be near the red circle, which highlights the newly dragged node. It is clear that, this time a highly dense area of nodes is formed very close to the current node in Figure 4.7(b) and our elastic network was able to find the closest node within this area. This is proving the methodology of way points for our robots to gain curiosity by generating a way point being both close to the current intention of the person and in a familiar place to the human subject where human trajectories are observed more densely.

As for another illustrative example in Figure 4.8, we used the same initialization of the map given in Figure 4.7(a). This time, Figure 4.8(a) demonstrates two very close new intentions of the human generated (case of confusion in estimation or human hesitation) before the robot moves executing the way point found in previous iterations. Again assuming that we are in *suspicious mode*, elastic net recalculates a new way point by first putting the initial dynamic point back to its original location in the middle of current and previous intention nodes in Figure 4.8(b). In part (b) of

the figure, the newly generated dynamic point trajectory with newly added intentions circled in red demonstrates a lot of oscillations (clouds of black dots) that converge into a branching and ends at *critical iteration.* This convergence ended near a new way point (purple square) close to newly added intentions which formed a much denser area than the one in Figure 4.8(a).
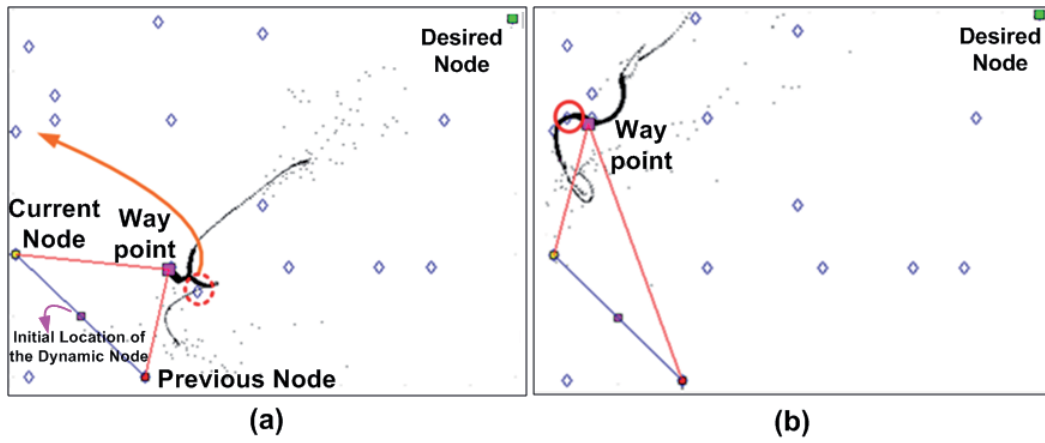
Figure 4.7 Way point planning performance of our system. (a) This is the first performance with the nodes replaced as shown; (b) The second performance is realized on the same map but dragging one node in (a) to the location highlighted with a red circle. Now the way point found is in the newly created dense area close to the current intention.
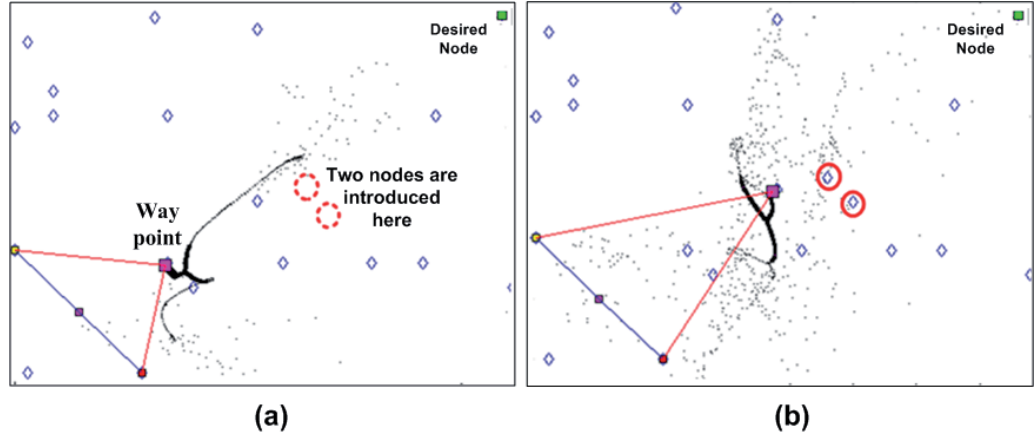
Figure 4.8 (a) With the same map given in Figure 4.7(a), the same solution was found; (b) The new system performance is given in response to the introduction of two new nodes in the map given in (a). It is clear that, there was created a new denser area by the introduction of these nodes and our system was able to locate a new node within this denser area.

For the last but not the least important performance of our system, we checked if the system is in *confident mode,* whether we can locate the desired intention and/or roam around it. For the purpose, we again used the same scenario, that is, the same replacements of all nodes introduced in Figure 4.7(a). However, this time the aim was to get close to the desired node since we had already gained the curiosity of the person. As clearly shown in Figure 4.9, a reinitiated elastic net with dynamic node replaced to its initial location directly got close to the desired node and found a node very close to the desired one in this case. It is clear that, a node that close to the desired one is actually a trajectory nearly same as the one of desired node (see Section 3.4.4). Therefore, if one of our robots realizes this trajectory, the robot will end up on the location of desired intention as we expected.
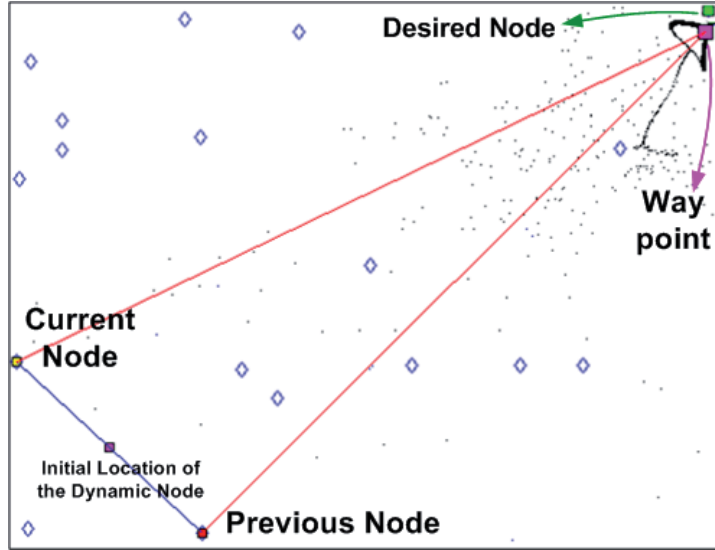
Figure 4.9 The same scenario given in Figure 4.7(a) is realized assuming that the system is in *confident mode.* As it is expected, our elastic net model found the desired node. Then, our robot will realize this trajectory guiding the person to the location of the desired intention.

If we are to discuss these resultant simulations of our elastic network model created, we conclude that this method is very suitable for our purpose of motion planning to the robots making the human change his/her intention to a desired one. We are trying to reshape intentions of the people with not just random movements towards the desired intention, but we first check if the person is *suspicious* or *confident* with our robots, then plan trajectories either aiming to gain the curiosity or directly showing the desired intention location. Especially in *suspicious mode*, there are smooth transitions towards the desired intention which elastic network model fitted very well. More examples with real-time scenarios are given in the next section where the benefits of elastic networks discussed.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

In this novel study, we designed a robotic system which can socially interact with people and reshape previously recognized intentions by autonomous robot moves in real-time scenarios. Robot motions are planned by generating intention transients through Elastic Networks according to the different moods of the person. In other words, as long as human body-mood is detected to be *suspicious,* our contextual robots (a chair robot and a 2-steps robot) plan their own trajectories aiming to break the obstinance of the person on what s/he was doing and to gain the curiosity and the trust of the person relying on proxemics behaviors. After a successful switching to *confident mood,* the people begin to easily focus on the robot moves this time directly showing the readiness to accept new intention reshaped into. These reshaping moves are planned by intention trajectories formed by sequences of intention transients generated by Elastic Networks using previously learnt trajectories of the human subjects. Experimental results with humans given in Chapter 4 showed the success of our methodology in real-time. It was proved that, inducing positive emotional mood to a person increases the comfort of the person to interact with our robots in an unfamiliar environment, resulting in the chance of making the person follow the lead of the robot.

This ability of social cognition and interaction makes the system usable as a human assistant in industry or even in our daily lives. For example, in emergency situations where people need guidance, our system can classify humans as confident and suspicious and lead them to safer places. In addition, our sociable robots can have commercial usage catching the attention of the humans and guiding them towards intended shops. Furthermore, the system can be used as assistants for the people in need for help (elderly people, people with Alzheimer's etc.) understanding their

needs and guiding them accordingly. Similarly, the system can be an educator for children or a therapy robot for people with autism, with its ability of estimating and inducing intentions and emotions. Finally, in our daily lives within the concept of smart home, our robots can be used as a service robot without requiring any manual input from its human companion by recognizing their needs and serving autonomously.

Satisfying the possible usages mentioned above, we are planning in the future to enhance our experimental room setup to include more variations on intentions. Therefore, we are going to adapt the system for more complex environments equipped with advanced manipulators having higher sensing capabilities and reaction speeds (as in the case of smart homes). In addition, our system only covers the cases for neutral body-mood of humans. If we can extend the work to detect neutral or aggressive body-mood of the people, it can be used as a detector of criminal or foul-minded people, and can even manipulate their deviant intentions thereafter. Finally, our system focuses on non-verbal interactions, which limits its possible usage and interaction capabilities. In the future, we are planning to add speech cognition to our system increasing its social ability and making it more like a human-being.

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672. doi:10.1038/372669a0

Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior*. (J. Kuhl & J. Beckmann, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-69746-3

Ali, N., Khan, N., & Imran, A. (2007). Controlling Mouse through Eyes. In *International Conference on Emerging Technologies, ICET 2007*. (pp. 179–183). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4516339

Arai, K., & Mardiyanto, R. (2011). Eye Based HCI with Moving Keyboard for Reducing Fatigue Effects. *2011 Eighth International Conference on Information Technology: New Generations (ITNG)*, 417–422. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5945272

Bahar, I., & Rader, a J. (2005). Coarse-grained normal mode analysis in structural biology. *Current opinion in structural biology*, *15*(5), 586–92. doi:10.1016/j.sbi.2005.08.007

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70(3)*, 614–636. doi:10.1037/0022-3514.70.3.614

Barakova, E. I., & Lourens, T. (2010). Expressing and interpreting emotional movements in social games with robots. *Personal and Ubiquitous Computing*, *14*(5), 457–467. doi:10.1007/s00779-009-0263-2

Baum, L., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics*, *41*(1), 164–171. doi:10.1214/aoms/1177697196

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2013). Featural processing in recognition of emotional facial expressions. *Cognition & emotion*, *28*(3), 416–432. doi:10.1080/02699931.2013.833500

Boeres, M., de Carvalho, L. A. ., & Barbosa, V. C. (1992). A faster elastic-net algorithm for the traveling salesman problem. *IJCNN, International Joint Conference on Neural Networks, 1992*, *2*, 215–220. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=227005

Bratman, M. (1999). *Intentions, Plans, and Practical Reason*. Cambridge University Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=RhgnY0-

6BmMC&oi=fnd&pg=PR9&dq=intentions+in+communication&ots=66yAtkukEd&sig=iQKmlx2ibFbmbYa-_hVFU08HYsc

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1-2), 119–155. doi:10.1016/S1071-5819(03)00018-1

Burkhardt, F., & Sendlmeier, W. (2000). Verification of acoustical correlates of emotional speech using formant-synthesis. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. Retrieved from http://www.isca-speech.org/archive_open/speech_emotion/spem_151.html

Butler, J. T., & Agah, A. (2001). Psychological Effects of Behavior Patterns of a Mobile Personal Robot. *Autoınomous Robots*, *10*(2), 185–202.

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, *8*(1), F13–F20. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-7687.2004.00385.x/full

Charniak, E., & Goldman, R. (1993). A Bayesian model of plan recognition. *Artificial Intelligence*, *64*(1), 53–79.

Chen, T.-H., Lin, Y.-F., & Chen, T.-Y. (2007). Intelligent Vehicle Counting Method Based on Blob Analysis in Traffic Surveillance. In *Second International Conference on Innovative Computing, Information and Control (ICICIC 2007)* (p. 238). IEEE. doi:10.1109/ICICIC.2007.362

Chennubhotla, C., Rader, a J., Yang, L.-W., & Bahar, I. (2005). Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Physical biology*, *2*(4), S173–80. doi:10.1088/1478-3975/2/4/S12

Chouchourelou, A., & Matsuka, T. (2006). The visual analysis of emotional actions. *Social Neuroscience*, *1*(1), 63–74. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/17470910600630599

Christensen, H., Pacchierotti, E., & Hgskolan, K. T. (2005). Embodied social interaction for robots. In *Proceedings of the 2005 Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB-05), Hertfordshire* (pp. 40–45).

Clarke, T. J., Bradshaw, M. F., Fieldô, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception*, *34*(10), 1171–1180. doi:10.1068/p5203

Cowie, R., & Douglas-Cowie, E. (1995). Speakers and hearers are people: reflections on speech deterioration as a consequence of acquired deafness. In *Profound Deafness and Speech Communication* (pp. 510–527).

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. . (2001). Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, (January), 32–80. doi:10.1109/79.911197

Cunningham, M. R. (1988a). What do you do when you're happy or blue? Mood, expectancies, and behavioral interest. *Motivation and Emotion*, *12*(4), 309–331. doi:10.1007/BF00992357

Cunningham, M. R. (1988b). Does Happiness Mean Friendliness?: Induced Mood and Heterosexual Self-Disclosure. *Personality and Social Psychology Bulletin*, *14*(2), 283–297. doi:10.1177/0146167288142007

Cutting, J., & Kozlowski, L. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, *9(5)*, 353–356. Retrieved from http://people.psych.cornell.edu/~jec7/pubs/friends.pdf

Daprati, E., Wriessnegger, S., & Lacquaniti, F. (2007). Kinematic cues and recognition of self-generated actions. *Experimental brain research*, *177*(1), 31–44. doi:10.1007/s00221-006-0646-9

Dario, P., Guglielmelli, E., & Laschi, C. (2001). Humanoids and personal robots: Design and experiments. *Journal of robotic systems*. Retrieved from http://citeseer.uark.edu:8080/citeseerx/showciting;jsessionid=4942263C2202CB506C89EE4F4CEBBD6F?cid=261580

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: Murray. (Reprinted, Chicago: University of Chicago Press, 1965.).

Dautenhahn, K, & Werry, I. (2000). Issues of robot-human interaction dynamics in the rehabilitation of children with autism. *Proc. From animals to animats*, *6*, 519–528. Retrieved from http://cognet.mit.edu/library/books/mitpress/0262632004/cache/chap54.pdf

Dautenhahn, Kerstin. (1999). Robots as social actors: aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference* (p. 374). San Francisco.

Dautenhahn, Kerstin, & Billard, A. (1999). Bringing up robots or—the psychology of socially intelligent robots: From theory to implementation. In *AGENTS '99 Proceedings of the third annual conference on Autonomous Agents* (pp. 366–367). doi:10.1145/301136.301237

De Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current biology : CB*, *18*(6), 454–7. doi:10.1016/j.cub.2008.02.057

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*(1), 1–38.

Dennett, D. (1989). *The Intentional Stance*. MIT Press.

Dielmann, A., & Renals, S. (2004). Dynamic bayesian networks for meeting structuring. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)* (pp. 629–632).

Dietterich, T. (2002). Machine learning for sequential data: A review. *Structural, syntactic, and statistical pattern recognition*, 1–15. Retrieved from http://link.springer.com/chapter/10.1007/3-540-70659-3_2

Dittrich, W., Troscianko, T., Lea, S., & D, M. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738.

Durbin, R. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. *Selected Works of Terry Speed*. Cambridge University Press.

Durbin, Richard, Szeliski, R., & Yuille, A. (1989). An analysis of the elastic net approach to the traveling salesman problem. *Neural Computation*, *1*(3), 348–358. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/neco.1989.1.3.348

Durbin, Richard, & Willshaw, D. (1987). An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, *326*, 689–691. Retrieved from http://comptop.stanford.edu/u/references/dw.pdf

Durdu, A. (2012). *Robotic system design for reshaping estimated human intention in human-robot interactions*. Middle East Technical University.

Durdu, A., Erkmen, I., Erkmen, A. M., & Yilmaz, A. (2011). Morphing Estimated Human Intention via Human-Robot Interactions. In *Proceedings of the World Congress on Engineering and Computer Science* (Vol. I). San Francisco.

Durdu, A., Erkmen, I., Erkmen, A. M., & Yilmaz, A. (2012). Robotic Hardware and Software Integration for Changing Human Intentions. In T. Sobh & X. Xiong (Eds.), *Prototyping of Robotic Systems: Applications of Design and Implementation* (pp. 380–406). IGI Global Publisher USA.

Ekman, P., & Friesen, W. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, *29(3)*, 288–298. doi:10.1037/h0036006

Erden, M. S., & Tomiyama, T. (2010). Human-Intent Detection and Physically Interactive Control of a Robot Without Force Sensors. *IEEE Transactions on Robotics*, *26*(2), 370–382. doi:10.1109/TRO.2010.2040202

Fogassi, L., Ferrari, P., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*, 662–667. doi:10.1126/science.1106138

Fogg, B. (1999). Persuasive technologies. *Communications of the ACM*, *42*(5), 26–29. Retrieved from http://dl.acm.org/citation.cfm?id=301396

Fogg, B. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, 89–120. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Computers+as+Persuasive+Social+Actors#2

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3-4), 143–166. doi:10.1016/S0921-8890(02)00372-X

Fredrickson, B. (2003). The value of positive emotions: The emerging science of positive psychology is coming to understand why it's good to feel good. *American scientist*, *91*, 330–335. Retrieved from http://www.jstor.org/stable/27858244

Geng, X., Chen, Z., Yang, W., Shi, D., & Zhao, K. (2011). Solving the traveling salesman problem based on an adaptive simulated annealing algorithm with greedy search. *Applied Soft Computing*, *11*(4), 3680–3689. Retrieved from http://www.sciencedirect.com/science/article/pii/S1568494611000573

Grol, M., Koster, E., Bruyneel, L., & Raedt, R. De. (2013). Effects of positive mood on attention broadening for self-related information. *Psychological research*. doi:10.1007/s00426-013-0508-6

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience*. Retrieved from http://www.nature.com/neuro/journal/v5/n4/abs/nn827.html

Haliloglu, T., Bahar, I., & Erman, B. (1997). Gaussian Dynamics of Folded Proteins. *Physical Review Letters*, *79*(16), 3090–3093. doi:10.1103/PhysRevLett.79.3090

Hall, E., Birdwhistell, R., & Bock, B. (1968). Proxemics. *Current Anthropology*, *9*, 83–108. Retrieved from http://www.jstor.org/stable/10.2307/2740724

Hart, P., Nilsson, N., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4082128

Hatfield, E., Cacioppo, J., & Rapson, R. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*, 96–99. Retrieved from http://www.jstor.org/stable/10.2307/20182211

Heinze, C. (2003). *Modelling intention recognition for intelligent agent systems*. the University of Melbourne, Australia. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA430005

Hopfield, J., & Tank, D. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics*, *52*(3), 141–152. Retrieved from http://www.springerlink.com/index/m62385p2j7132844.pdf

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion*, *3(2)*, 150–166. doi:10.1037/1528-3542.3.2.150

Huettenrauch, H., Eklundh, K., Green, A., & Topp, E. (2006). Investigating Spatial Relationships in Human-Robot Interaction. In *2006 IEEE/RSJ International*

Conference on Intelligent Robots and Systems (pp. 5052–5059). doi:10.1109/IROS.2006.282535

Jenkins, O. C., Serrano, G. G., & Loper, M. M. (2007). Interactive human pose and action recognition using dynamical motion primitives. *International Journal of Humanoid Robotics*, *04*(02), 365–385. doi:10.1142/S0219843607001060

Jessen, S., & Kotz, S. a. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage*, *58*(2), 665–74. doi:10.1016/j.neuroimage.2011.06.035

Jones, J. (1924). On the determination of molecular fields. II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, …*, *106*(738), 463–477. Retrieved from http://www.jstor.org/stable/10.2307/94265

Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, *129(5)*, 770–814. Retrieved from http://doi.apa.org/index.cfm?doi=10.1037/0033-2909.129.5.770

Kautz, H. A., & Allen, J. F. (1986). Generalized plan recognition. *AAAI*, *86*, 32–37.

Kelley, R., Tavakkoli, A., King, C., Nicolescu, M., Nicolescu, M., & Bebis, G. (2008). Understanding human intentions via hidden markov models in autonomous mobile robots. In *Proceedings of the 3rd international conference on Human robot interaction - HRI '08* (pp. 367–374). New York, New York, USA: ACM Press. doi:10.1145/1349822.1349870

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (1993). Facial expression of emotion. In *Handbook of affective sciences* (pp. 415–432). New York,US: Oxford University Press. Retrieved from http://psycnet.apa.org/journals/amp/48/4/384/

Kirkpatrick, S., Jr, C. G., & Vecchi, M. (1983). Optimization by simulated annealing. *science*, *220*(4598), 671–680. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Optimization+by+Simulated+Annealing#0

Kittel, C., & McEuen, P. (1996). *Introduction to solid state physics* (8th ed.). New York: John Wiley & Sons, Inc. Retrieved from http://tocs.ulb.tu-darmstadt.de/125610068.pdf

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, *27(2)*, 456–465. doi:10.1037/0096-1523.27.2.456

Kohler, E., Keysers, C., Umilta, M., Fogassi, L., Vittorio, G., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, *297*, 846–848. doi:10.1126/science.1070311

Koo, S., & Kwon, D. (2009). Recognizing Human Intentional Actions from the Relative Movements between Human and Robot. In *The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009. RO-MAN 2009.* (pp. 939–944).

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care: a case-study in autism therapy. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 341–346. doi:10.1109/ROMAN.2005.1513802

Lee, K. K., & Xu, Y. (2004). Modeling human actions from learning. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, *3*, 2787–2792. doi:10.1109/IROS.2004.1389831

Lee, S., & Son, Y. (2008). Integrated human decision making model under belief-desire-intention framework for crowd simulation. *Simulation Conference, 2008. WSC 2008. Winter*, (Norling 2004), 886–894. doi:10.1109/WSC.2008.4736153

Lewin, K. (1952). *Field theory in social science: Selected theoretical papers*. (D. Cartwright, Ed.). Retrieved from http://library.wur.nl/WebQuery/clc/388286

Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: a stimulus set of point-light communicative interactions. *Behavior research methods*, *42*(1), 168–78. doi:10.3758/BRM.42.1.168

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., … De Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC,2011* (Vol. 2011, pp. 4515–8). doi:10.1109/IEMBS.2011.6091119

Mead, R., & Matarić, M. (2011). *An experimental design for studying proxemic behavior in human-robot interaction*. Technical Report CRES-11-001, USC Interaction Lab, Los Angeles.

Meltzoff, A. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology*, *31(5)*, 838–850. Retrieved from http://psycnet.apa.org/journals/dev/31/5/838/

Mitchell, T. M. (1997). *Machine learning* (p. 432). McGraw-Hill Science/Engineering/Math.

Mori, T., Segawa, Y., Shimosaka, M., & Sato, T. (2004). Hierarchical recognition of daily human actions based on continuous Hidden Markov Models. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.* (pp. 779–784). doi:10.1109/AFGR.2004.1301629

Mori, Taketoshi, Shimosaka, M., Harada, T., & Sato, T. (2005). Time-Series Human Motion Analysis with Kernels Derived from Learned Switching Linear Dynamics. *Transactions of the Japanese Society for Artificial Intelligence*, *20*, 197–208. doi:10.1527/tjsai.20.197

Nahar, S., Sahni, S., & Shragowitz, E. (1986). Simulated annealing and combinatorial optimization. In *DAC '86 Proceedings of the 23rd ACM/IEEE Design Automation Conference* (pp. 293–299). Retrieved from http://dl.acm.org/citation.cfm?id=318059

Pacchierotti, E., Christensen, H. I., & Jensfelt, P. (2005). Human-robot embodied interaction in hallway settings: a pilot user study. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (pp. 164–171). doi:10.1109/ROMAN.2005.1513774

Parlangeli, O., Guidi, S., & Caratozzolo, M. C. (2013). A mind in a disk: the attribution of mental states to technological systems. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *41*, 1118–1123.

Paulos, E., & Canny, J. (1998). Designing personal tele-embodiment. In *IEEE Internetional Conference on Robotics and Automation, 1998* (pp. 3173–3178). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=680913

Qi, Y., Wang, Z., & Huang, Y. (2007). A non-contact eye-gaze tracking system for human computer interaction. *Wavelet Analysis and Pattern ….* Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4420638

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. doi:10.1109/5.18626

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between Emotion and Facial Expression: Evidence from Laboratory Experiments. *Emotion Review*, *5*(1), 16–23. doi:10.1177/1754073912457228

Robins, B., Dautenhahn, K., Boekhorst, R. Te, & Billard, a. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, *4*(2), 105–120. doi:10.1007/s10209-005-0116-3

Roether, C., Omlor, L., Christensen, A., & Giese, M. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*. Retrieved from http://jov.highwire.org/content/9/6/15.short

Roy, S., Sarma, S., Soumyadip, C., & Maity, S. (2013). A comparative study of various methods of ANN for solving TSP problem. *Internation Journal of Computers & Technology*, *4(1)*, 19–28. Retrieved from http://ijssronline.com/cir/index.php/ijct/article/view/721

Sato, E., Yamaguchi, T., & Harashima, F. (2007). Natural interface using pointing behavior for human–robot gestural interaction. *IEEE Transactions on Industrial Electronics*, *54*(2), 1105–1112. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4140637

Sauter, D., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97–113. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.2044-835X.2012.02081.x/full

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech communication*, *40*(1-2), 227–256. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167639302000845

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). The Utility of Affect Expression in Natural Language Interactions in Joint Human-Robot Tasks. In *HRI '06 Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 226–233).

Schmidt, C., Sridharan, N., & Goodson, J. (1978). The plan recognition problem: an intersection of psychology and artificial intelligence. *Artificial Intelligence*, *11*(1-2), 45–83.

Schmidt, S., & Färber, B. (2009). Pedestrians at the kerb–Recognising the action intentions of humans. *Transportation research part F: traffic psychology and Behaviour*, *12*(4), 300–310. Retrieved from http://www.sciencedirect.com/science/article/pii/S1369847809000102

Sedikides, C. (1992). Mood as a determinant of attentional focus. *Cognition & Emotion*, *6*(2), 129–148. doi:10.1080/02699939208411063

Sevdalis, V, & Keller, P. (2009). Self-recognition in the perception of actions performed in synchrony with music. *Annals of the New York Academy of Sciences*, *1169*, 499–502. doi:10.1111/j.1749-6632.2009.04773.x

Sevdalis, Vassilis, & Keller, P. E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Consciousness and cognition*, *19*(2), 617–626. doi:10.1016/j.concog.2010.03.017

Shams, S. (1996). Neural network optimization for multi-target multi-sensor passive tracking. *Proceedings of the IEEE*, *84*(10), 1442–1457. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=537110

Shimosaka, M., Mori, T., Harada, T., & Sato, T. (2005). Marginalized Bags of Vectors Kernels on Switching Linear Dynamics for Online Action Recognition. In *International Conference on Robotics and Automaiton* (pp. 72–77).

Soille, P. (2003). *Morphological image analysis: principles and applications*. Springer-Verlag New York, Inc. Retrieved from http://dl.acm.org/citation.cfm?id=773286

Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 246–252. doi:10.1109/CVPR.1999.784637

Suzuki, K., Camurri, A., Ferrentino, P., & Hashimoto, S. (1998). Intelligent agent system for human-robot interaction through artificial emotion. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)* (Vol. 2, pp. 1055–1060). doi:10.1109/ICSMC.1998.727828

Tahboub, K. (2005). Compliant human-robot cooperation based on intention recognition. In *International Symposium on Intelligent Control. Proceedings of the 2005 IEEE* (pp. 1417–1422). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1467222

Tahboub, K. A. (2006). Intelligent Human – Machine Interaction Based on Dynamic Bayesian Networks Probabilistic Intention Recognition. *Journal of Intelligent and Robotic Systems*, *45*(1), 31–52. doi:10.1007/s10846-005-9018-0

Takayama, L., & Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5495–5502). doi:10.1109/IROS.2009.5354145

Terada, K., & Ito, A. (2010). Can a robot deceive humans? *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 191–192. doi:10.1109/HRI.2010.5453201

Terada, K., Shamoto, T., Mei, H., & Ito, A. (2007). Reactive Movements of Non-humanoid Robots Cause Intention Attribution in Humans. In *IEEE/RSJ International Conference on Intelligent Robots & Systems* (pp. 3715–3720). doi:10.1109/IROS.2007.4399429

Tirion, M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters*, *77*(9), 1905–1908. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10063201

Vakhutinsky, A., & Golden, B. (1994). Solving vehicle routing problems using elastic nets. *Neural Networks, 1994. IEEE World Congress on Computational Intelligence*, *7*(2), 4535–4540. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=375004

Vanderborght, B., Simut, R., & Saldien, J. (2012). Using the social robot probo as a social story telling agent for children with ASD. *Interaction Studies*, *13*, 348–372. Retrieved from http://www.ingentaconnect.com/content/jbp/is/2012/00000013/00000003/art00002

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269. doi:10.1109/TIT.1967.1054010

Vlasenko, B., & Wendemuth, A. (2009). Heading toward to the natural way of human-machine interaction: the NIMITEK project. In *IEEE International Conference on Multimedia and Expo, 2009. ICME 2009.* (pp. 950–953).

Wada, K., Shibata, T., Saito, T., & Tanie, K. (2004). Effects of robot-assisted activity for elderly people and nurses at a day service center. In *Proceedings of the IEEE* (Vol. 92, pp. 1780–1788). doi:10.1109/JPROC.2004.835378

Wadlinger, H., & Isaacowitz, D. (2006). Positive mood broadens visual attention to positive stimuli. *Motivation and Emotion*, *30*, 89–101. Retrieved from http://link.springer.com/article/10.1007/s11031-006-9021-1

Wallbott, H. G. (1998). Bodily expression of emotion. *Europian journal of social psychology*, *28*(6), 879–896.

Walters, M. L., Oskoei, M. a., Syrdal, D. S., & Dautenhahn, K. (2011). A long-term Human-Robot Proxemic study. In *IEEE, RO-MAN 2011* (pp. 137–142). doi:10.1109/ROMAN.2011.6005274

Wang, E., Lignos, C., Vatsal, A., & Scassellati, B. (2006). Effects of head movement on perceptions of humanoid robot behavior. In *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06* (p. 180). New York, New York, USA: ACM Press. doi:10.1145/1121241.1121273

Wang, S., Liu, Z., Lv, S., Lv, Y., & Wu, G. (2010). A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Biometrics Compendium*, *12*(7), 682–691. doi:10.1109/TMM.2010.2060716

Wang, Z., Mülling, K., Deisenroth, M. P., Amor, H. Ben, Vogt, D., Schölkopf, B., & Peters, J. (2013). Probabilistic Movement Modeling for Intention Inference in Human-Robot Interaction. *The International Journal of Robotics Research*. doi:10.1177/0278364913478447

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98(2)*, 219–235. Retrieved from http://psycnet.apa.org/journals/bul/98/2/219/

Webb, T., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological bulletin*, *132*, 249–268. doi:10.1037/0033-2909.132.2.249

Wehrle, T., & Kaiser, S. (2000). Emotion and facial expression. *Affective interactions*, *1814*, 49–63. Retrieved from http://link.springer.com/chapter/10.1007/10720296_5

Welch, G., & Bishop, G. (1995). *An Introduction to the Kalman Filter by* (pp. 1–16).

Werry, I., Dautenhahn, K., & Harwin, W. (2001). Investigating a robot as a therapy partner for children with autism. *Procs AAATE 2001*, *6*. Retrieved from http://uhra.herts.ac.uk/handle/2299/1905

Werry, I., Dautenhahn, K., Ogden, B., & Harwin, W. (2001). Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy. *Computer Technology: Instruments of Mind*, *2117*, 57–74. Retrieved from http://link.springer.com/chapter/10.1007/3-540-44617-6_6

Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning*. Retrieved from http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=5673187

Wood, J. V, Saltzberg, J. A., & Goldsamt, L. A. (1990). Does affect induce self-focused attention? *Journal of personality and social psychology*, *58*(5), 899–908.

Wurtz, R., & Albano, J. (1980). Visual-motor function of the primate superior colliculus. *Annual review of neuroscience*, *3*, 189–226.

Yamato, J., Ohya, J., & Ishii, K. (1992). Recognizing human action in time-sequential images using hidden Markov model. In *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 379–385). IEEE Comput. Soc. Press. doi:10.1109/CVPR.1992.223161

Yokoyama, A., & Omori, T. (2010). Modeling of human intention estimation process in social interaction scene. *International Conference on Fuzzy Systems*, 1–6. doi:10.1109/FUZZY.2010.5584042

Zhang, D., Gatica-Perez, D., Bengio, S., & McCowan, I. (2006). Modeling individual and group actions in meetings with layered HMMs. In *IEEE Transactions on Multimedia* (pp. 509–520).