Morphing Estimated Human Intention via Human-Robot Interactions

Akif DURDU, Ismet ERKMEN, Aydan M. ERKMEN, Alper YILMAZ

***Abstract*— Estimating and reshaping human intentions are topics of research in the field of human-robot interaction. Although works on estimating human intentions are quite well known research areas in the literature, reshaping intentions through interactions is a new significant branching in the field of human-robot interaction. In this paper, we research how the human intentions change based on his/her actions by moving the robots in a real human-robot environment. Our approach uses the Hidden Markov Model (HMM) tailored for the intelligent robotic systems. The algorithmic design consists of two phases: human tracking and the use of intelligent robots that aims to change intentions of individuals in the environment. In the former phase, postures and locations of the human are monitored by applying low-level video processing methods. The latter phase learned HMM models are used to reshape the estimated human intention. This two-phase system is tested on video frames taken from a real human-robot environment. The results obtained using the proposed approach is discussed according to performance towards the "degree" of reshaping the detected intentions.**

***Index Terms*— Intention Estimation; Human-robot Interaction; Hidden Markov Model (HMM); Principle Component Analysis (PCA).**

I. Introduction

HUMAN-ROBOT interaction, which is generally used for cooperation tasks, require reliable and full communication capabilities. In the case when the human and robot does not communicate, the cooperation between them can be established with estimating the intention of the human and/or the robot. If the human is able to express his intention clearly for a specific environment task then intention estimation reduces to both communication and the necessary budget for communication [1].

Manuscript received June 2, 2011; revised August 08, 2011. This work was supported in part by the Middle East Technical University under project BAP-2002K120510. This work was also supported in part by the TUBITAK (the scientific and technological research council of Turkey).

A. Durdu is with Department of Electrical and Electronics Engineering Department, Middle East Technical University, 06531, Ankara, Turkey (phone: 90-312-210 45 90; fax: 90-312-210 23 04; e-mail: dakif@metu.edu.tr).

I. Erkmen is with the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, 06531, Turkey (e-mail: erkmen@metu.edu.tr).

A. M. Erkmen is with the Electrical and Electronics Engineering Department, Middle East Technical University, Ankara, 06531, Turkey (e-mail: aydan@metu.edu.tr).

A. Yilmaz is with Photogrammetric Computer Vision Laboratory, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH, 43210, USA. (e-mail: yilmaz.15@osu.edu).

In this study, we consider human-robot interactions without direct verbal [1] or gesture based communication. While our method significantly differs from the conventional treatment of the problem, we introduce the intention estimation problem, its characteristics and requirements as covered in this conventional literature. We also discuss the prototyping environment for the intention estimation problem which includes the hardware components as well as the software tools.

Intention has been discussed by philosophers for many years [2],[3],[4]. Intention estimation is, in technical terms, defined as the problem of making inferences about a human's intention by means of his actions and the effects of these actions in the environment they occur [1]. Intention estimation, which is also referred to as the plan recognition, utilized in many human-robot applications is based on human's control commands in a simulation or a game environment [1],[5],[6]. Koo et al. present a real-time experimental setup for recognizing the intention of human action with robot that has IR sensors [9]. Aarno et al. give an example for intention recognition of robot-hand motion using the HMM model [8]. In this paper, we present computer vision-based human intention estimation in a real environment and demonstrate that a change in human intention can be observed by moving robots in specified directions.

In the following discussion, we detail how the robotic agent, systematically, recognizes the human intention through posture and position information before his next move. The discussion is organized by first providing an overview of the experimental environment by detailing its hardware components. We, then, introduce the software prototyping for recognizing human intention. This section provides an overview of algorithms developed in the field of machine vision which estimates human posture and tracks it. Last but not least we present our experimental results, and conclude in Section IV.

II. System Design And Methodology

A. Review

Vision based design and implementation of algorithms make inferences from a sequence of images acquired from a camera. By rephrasing Aristotle's description about vision as finding "what is where by looking", we emphasize two most essential problems that computer vision researchers in computer vision try to solve: recognition (identifying "what") and localization (identifying "where") [10].

Recognition, fundamentally, requires the classification of

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

a detected target image into one of many discrete classes. However, assuming that the target is known, its motion for accomplishing some intention in a video stream setting remains an unknown. In our setting, a human walks in a room which only includes a human and a set of robots. The position (coordinates) of the human is estimated by computer vision, as a target is another method for human-robot interaction which can be considered mapping from the space of images to continuous output which has been previously used as an assistive benchmarking technology for people with physical difficulties and communication [10]. The exploitation of the computer vision based mapping of image data to a continuous output, in our framework improves a general prototyping framework known as Vision-based Inference System (VIS), which is appearance-based approach that has learning and reasoning capabilities.

Figure 1 schematically illustrates our computer vision-based set-up: a webcam captures a view of the environment and converts it into a digital format, which is then processed by a feature extractor to generate a descriptor. The VIS then maps to a low-dimensional output vector generating the characteristics of human intention from the input image stream, which is a simple yet powerful representation.

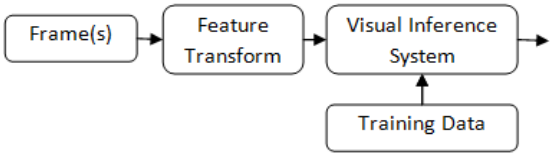

Fig. 1. The visual inference flow diagram of the general system. A digital image from video frames is processed by the feature transform to give a feature vector which the vision based inference system then translates to an output vector. With no initial model of how images are formed, the vision based inference machine mapping is defined by a training data containing example input–output pairs.

Another important aspect of our work, in addition to intention estimation, is the morphing of initial intentions of a human agent during human-robot interactions. In our setup, the hardware component is composed of robots poking human behaviors so as to reshape the initial intentions of the human entering the interaction scenario. The physical environment is designed to foster human-robot interactions. In this setup, the process of understanding current intentions and changing them is a continuous and complex and relates to the human to human interactions.

Let's first consider human-human interactions. In a human-human conversation, humans try to predict the future direction of the conversation and the reaction of the other person by estimating the intention during the conversation. Another example is two people walking together who mutually coordinate their steps and move by estimating the intention of each other that result in a walk pattern similar to each other. In these and other human-human interactions, the intention estimation is performed subconsciously, such that conversations unfold and subtle queues are reacted to without much conscious thought given to the underlying motives of the other.

The interactions between a human and a robot, however, do not direct communication tools or human subconscious to predict intention. It is commonly observed that to more intelligent human agent changes his intentions speedily. The following, we detail hardware and software prototypes for morphing intentions of human agent by robotic interaction. This is achieved by modeling, reshaping and generating intentions.

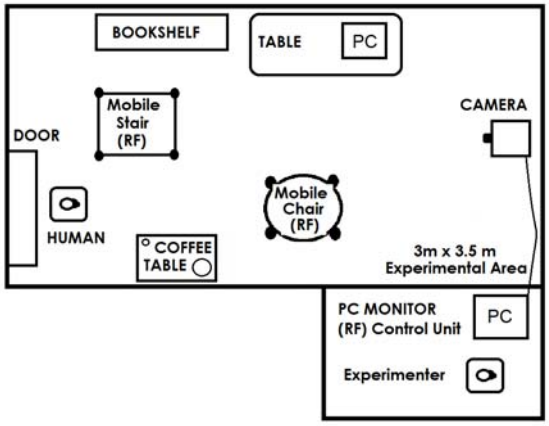

Fig. 2. (a) The prototype of the experimental environment

Fig. 2. (b.1) Top view of the experimental environment

Fig. 2. (b.2) Side view of the experimental environment

B. Experimental Setup

An experimental application room (Fig.2.a) is designed to observe the reshaping of human intention by targeted motion

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

of the robotic chair and shallow stair shown in Figure 2.b.3. The room, which includes worktable, bookshelf, coffee machine and mobile chair/stair robots, is monitored by two cameras mounted at an oblique and orthos angles.

The oblique view provides full posture of the agent, while the ortho view provides position of the agent relative to other scene components. A computer controlled stair and chair performs a series of activities for changing the intention of the human agent by interacting with the participant entering the experimental room (see Figure 2 b.2-b.3). The stair and chair robots are powered by two wheels driven by 12V DC motor and are moving in plane. A Radio Frequency (RF) communication system allows the experimenter/computer to remotely operate the robotic chair and stair. In our procedure, the human agents are permitted to behave freely within experimental area; however, damage to robots and the equipment are prohibited. Aside from these instructions no other instructions are given to the agent before the start of application.

Fifteen university students, whom ages ranging from 18 to 24 years, participated in the experiment. None of them had had were given prior knowledge about the experiment or had an experience in interacting with the mobile stair/chair robots in the experiment. First of all, the videos are collected to learn we recorded the video possible intentions of human agents which are used to predict and in which the experimenter controlled the robots according to a scenario regarding to changing the predicted intentions of human. This recorded video database is required to construct supervised learning based intelligent -overall- system in which the robots are controlled by computer automatically.

C. *Methodology*

The algorithmic approach to change the predicted human intention by using mobile robots is achieved using three step approaches.

First step is related to low level object detection and tracking by using background subtraction and matching techniques. In this step, people and robots are detected and tracked in a video sequence with a stationary background using the following process: 1) Using the first few frames of the video to learn the static background. 2) The moving pixels are then detected by comparing them against the learned background. 3) Moving pixels are grouped by connected component analysis, which are labeled as human or robots. 4) Finally, connected regions in the current frame are matched with those in the previous frame based on appearance and size information.

Second step deals with high-level information from low level cues based on Bayesian inference. In this step, the state of observed activities is defined as instantaneous movement of agents that executes his intention. A single state observed from the video, does not provide adequate information to predict the intention. We assume that the agent carries out a sequence of actions following a plan until the intention is achieved. Hence, we alternatively use a sequence of human action to predict the intention of the human agent.

Once the state is s constructed, we employ Bayesian inference to learn and estimate the human intentions. In particular, we adopt the Hidden Markov Model (HMM) that takes human posture and location information as its observation model. The proposed method classifies the posture of the human at each frame by using Principle Component Analysis (PCA), which provides us with the means to measure the Euclidian distance between projected data and nominal data.

The Last step in our approach is about inferring whether or not the robots change the human intention. In this step, the system compares the previous intention, which is predicted when the robot starts to move, with the current intention predicted after the robot completes its movement.

The flow chart of the whole system is shown in Fig.3 below.

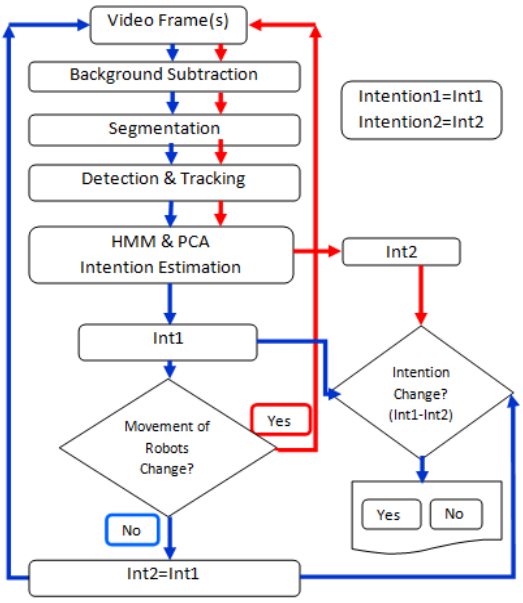

Fig. 3. Flow chart of the overall system

III. Formulation And Experimental Evaluation

A. *Low-level Object Detection and Tracking*

Detecting and tracking human and robots in a video sequence with a stationary background are an important part of this experiment. For this goal, background subtraction is performed by computing the likelihood of a pixel in the current image against the background model which is learned using a few images without and moving objects in it.

The likelihood values are subjected to auto-thresholding to determine which pixels belong to the moving objects in the scene. Detected pixels are then completed by the morphological closing which merges disconnected but proximal pixels to create blobs. Next, we calculate the bounding boxes and the centers of these blobs. In the final tracking step, to determine the locations of specific robots and human from one frame to another, the system compares the predicted locations of the bounding boxes

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

with the detected locations. This enables the system to assign a unique color to each robots and human (see Fig.4. for detection and Fig. 5. for tracking).

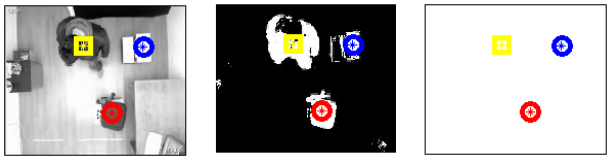

Fig. 4. Detected location of the human and robots (a) the colored points show the center of the robots and human on the one real video frame. (b) the locations on background subtracted frame (c) the locations on the empty image

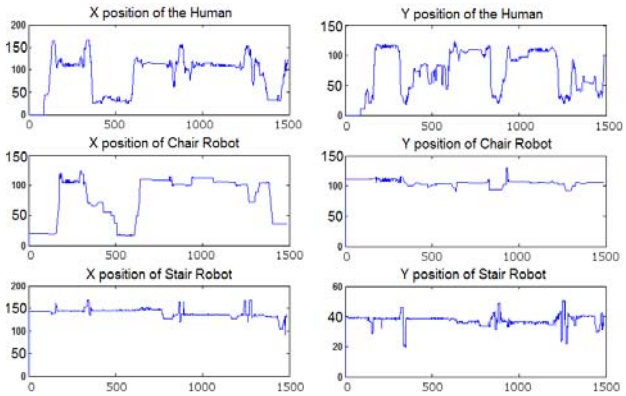

Fig. 5. Example frame trajectories as the location of human and robots (1500 frames)

B. High Level Learning

B.1. Principle Component Analysis (PCA) for Human Posture

PCA is a popular feature mapping approach for creating new descriptors that are linear combination of the original dimensions [11]. Geometrically, the purpose of the PCA is to identify a set of orthogonal axes called the principle components [7]. The new variables are formed by projected the original variables using their principle components. In this application, we use the projected variables to classify the posture of the human as "left position", "right position" and "center position". Figure 6 demonstrates the example training set and the coordinate system of the image sequence.

The algorithm based on the PCA for mapping the human posture is as follows:

1. Loading training images and preparing for PCA by subtracting mean,
2. Low dimension posture space construction (PCA),
3. Load test data frame and project on posture space,
4. Calculation of Euclidian distance (e_d) between the test frame image and neutral image which is the mean value of the "center position" training data,

$$e_d = \sqrt{(\text{test\_image}^T - \text{mean\_neutral})^T . (\text{test\_image}^T - \text{mean\_neutral})} \quad (1)$$

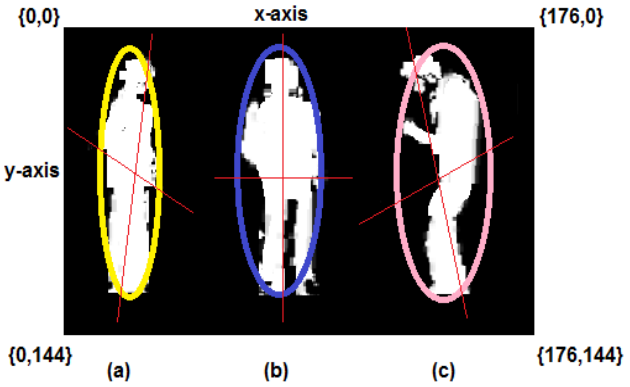

Fig. 6. The example training set and the coordinate system of the image sequence (a) "left position" of the human (b) "center position" of the human (c) "right position" of the human

B.2. Hidden Markov Model (HMM) for Intention Estimation

In our application, observable actions are defined as 'going to bookshelf', 'going to worktable', 'going to coffee-table' and 'exploring the environment'. Each of these actions, which consist of a sequence of image frames, is labeled with human intention as the hidden state of the HMM model. The labeled intention is tabulated in Table 1.

TABLE I
INTENTION LIST

Number	Intention
1	Taking a book from bookshelf
2	Sitting on the work-table
3	Taking a coffee
4	Discovering the robots

.

The HMM model used in our framework is shown in Fig. 7. In this figure, each observation module (O_t), which includes a sequences of locations and posture information of the human, demonstrates the observed human action. In the transition model, a conditional probability distribution matrix represents the probability of the current observed action sequences (S_t) transmitted from the previous observed action sequences (S_{t-1}).

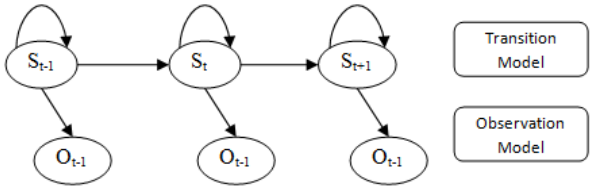

Fig. 7. The Hidden Markov Model (HMM)

The HMM model is shown in Fig. 7. In the figure, each observation module (O_t), which includes a sequences of location and posture information of the human, demonstrates observed human action. In the transition model, a conditional probability distribution matrix represents the probability of the current observed action sequences (S_t) transited from the previous observed action sequences (S_{t-1}).

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$P(S_t \mid S_{t-1}) \in \mathbb{R}^{4x4} \quad (2)$$

During the course of a human action, s/he keeps the intention in her/his mind so that transition from one intention to another one does not happen frequently. In other words, the diagonal terms of the transition matrix are near to the 1 and the other terms are relatively small value.

In the observation model, the probability of the observed human action (O) when human has a intention is defined as

$$P(O \mid S) \in \mathbb{R}^{4x9} \quad (3)$$

Human posture and position dependent Observation Model:

We discretize the observation model by labeling the grid cells from 144x176 to 24x25. Each cell symbolizes an observation part which depends on the position of the human agent. In Fig.8., we show a set of labeled cells in our framework. Figure 9 illustrates how the probabilistic observation matrix is constructed by using the training data. The probabilistic weight of each observation cell is assigned from five different trajectories defining each intention. Fig. 10. and 11. demonstrate respectively examples of two different training trajectories and one testing trajectory.

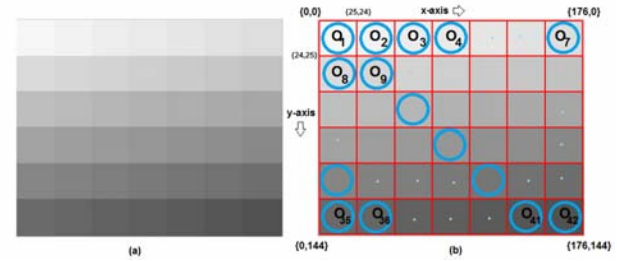

Fig. 8. Labeled observation cells.

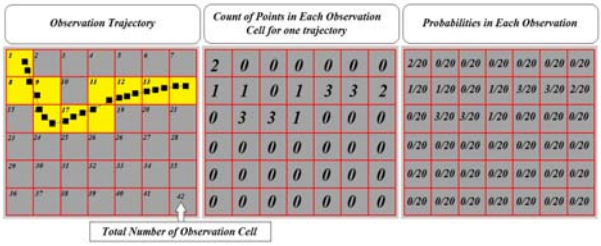

Fig. 9. Example construction of the probabilistic observation matrix

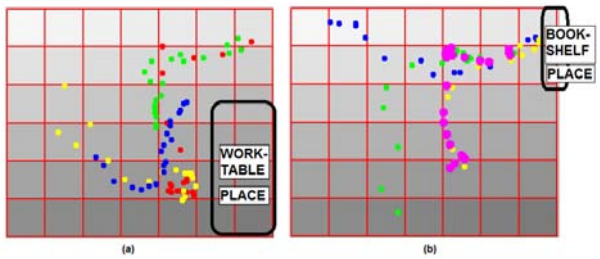

Fig. 10. Example training trajectory data's (a) "going to the work-table (b) "going to bookshelf"

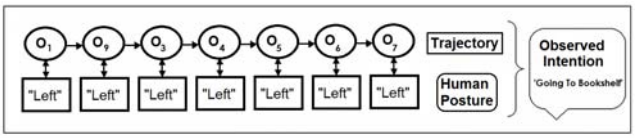

Fig. 11. Example testing trajectory data and observation of the action

IV. Results

In Fig. 12, the results show that the agent changes his intention in the experimental room due to the movement of robots. In frame 760, our system detects the intention of the agent as going and sitting at the desk which corresponds to the second intention in Table 1. After frame 770, the stair-robot starts moving to the different side of the room. At first the human agent tries to understand what was happened. Once the robot motion catches his attention, after frame 800, he stands up from the chair and moves in the direction of the bookshelf to take a book. This motion in our set up is recognized as intention 1. This and other experimental results suggest that we have successfully changed the intention of the agent. We present additional results from PCA based labeling and HMM based detection results respectively in Figure 12 and Table II.

TABLE II
Euclidian Distance with posture

	Frame= 760	**Frame= 800**	**Frame= 820**
Left	2221	2938	3887
Right	4764	3669	6319
Center	5292	4002	3503
Posture Result (PCA)	Left	Left	Center

Action observation with Expectation Maximization (log-likelihood) in HMM			
	Frame Interval= 750:760	**Frame Interval= 790:800**	**Frame Interval= 810:830**
Action 1 'going to bookshelf'	-854.265832	-949.938566	-795.833412
Action 2 'going to work-table'	-744.633492	-662.080792	-708.119380
Action 3 'going to coffee-machine'	-826.277215	-840.526317	-956.222300
Action 4 'discovering the environment'	-910.068189	-707.836509	-828.892867
Action Result (HMM)	**Action 2**	**Action 2**	**Action 1**

Intention			
Intention (from Table I)	**2**	**2**	**1**

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

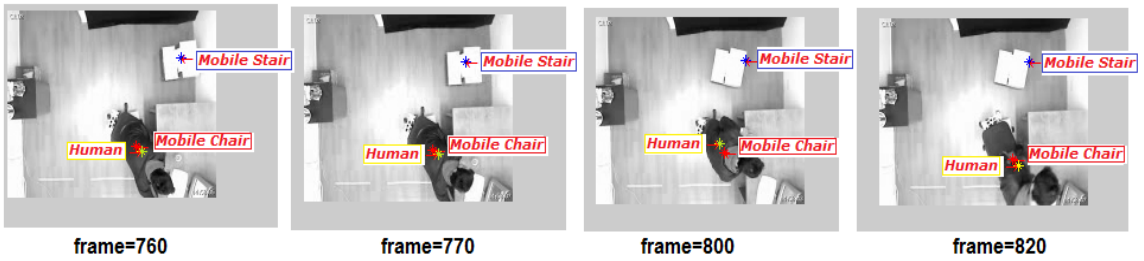

Fig. 12.a. Frame Captures of the experiment that show the changing the intention of the human by the movement of the mobile stair

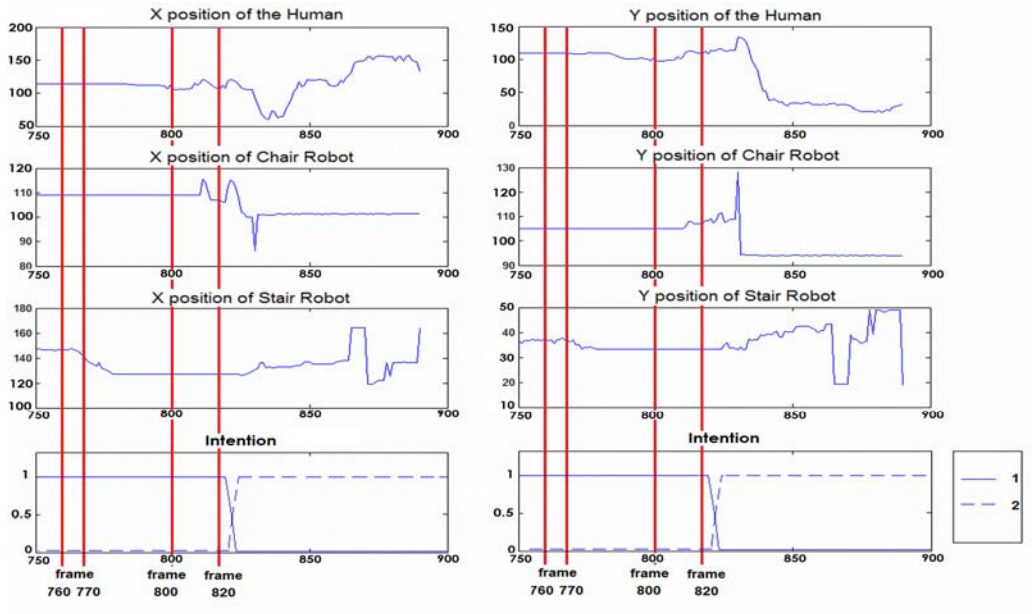

Fig. 12.b. (x,y) positions of the human, mobile chair and stair and the changing the intention

V. Conclusion

In this paper, we demonstrate that changing the intention of a human agent can be realized by interaction of a robot in our experimental room. Our approach uses a sequence of images and processes them according to the scenario at hand. Our future work is to realize this scenario in real-time instead of using prerecorded videos and manually maneuvered robots to generate a real-time intelligent human-robot interaction system. While this work does not completely analyze the activities to understand the intention, to the best of our knowledge, it is the only work aims at changing the intention of agents. This work is significant and we believe will foster future research on a different aspect of human-robot interaction.

References

[1] K. A. Tahboub, "Intelligent Human-Machine Interaction Based on Dynamic Bayesian Networks Probabilistic Intention Recognition," *Journal of Intelligent Robotics Systems*, vol. 45, no. 1, pp. 31–52, 2006.

[2] Bratman, M. Intention, plans, and practical reason, Harvard University Press.

[3] Bratman, M. Faces of Intention: Selected Essays on Intention and Agency, Cambridge University Press, (1999).

[4] Dennett, D. C. The Intentional Stance, MIT Press, (1987).

[5] Iba, S., C. J. J. Paredis, et al. Intention aware interactive multi-modal robot programming, (2003).

[6] Yokoyama, A. Omori, T. "Modeling of human intention estimation process in social interaction scene", Fuzzy Systems (FUZZ), 2010 IEEE International Conference on, (2010).

[7] K. K. Lee and Y. Xu, "Modeling human actions from learning," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04)*, vol. 3, pp. 2787- 2792, October 2004.

[8] D. Aarno and D. Kragic, Layered HMM for motion intention recognition, in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'06*, 2006.

[9] S. Koo, D. S. Kwon, Recognizing human intentional actions from the relative movements between human and robot, *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on* (2009), pp. 939-944.

[10] Oliver M. C. Williams "Bayesian Learning for Efficient Visual Inference", The dissertation for the degree of Doctor of Philosophy, 2005

[11] Jolliffe, I. T. *Principal Component Analysis*. Springer-Verlag. pp. 487, (1986).

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)