

Estimating and reshaping human intention via human–robot interaction

Akif DURDU^{1,*}, İsmet ERKMEN², Aydan Müşerref ERKMEN²

¹Department of Electrical and Electronics Engineering, Faculty of Engineering, Selçuk University, Konya, Turkey

²Department of Electrical and Electronics Engineering, Faculty of Engineering, Middle East Technical University, Ankara, Turkey

Received: 10.06.2013

Accepted/Published Online: 02.10.2013

Final Version: 01.01.2016

Abstract: Human–robot interaction (HRI) is studied in two important research areas, intention estimation and intention reshaping. Although there are many studies in the literature that define human intention, new research examines the reshaping of human intentions by using robots in HRI. In this paper, 2 different robot movements are tested in a real environment in order to reshape current human intention. The hidden Markov model (HMM) is used to estimate human intention in our intelligent robotic system. The algorithmic design of the system comprises 2 parts: the first part tracks the moving objects in the environment, and the second part estimates human intention and reshapes the estimated current human intention by using intelligent robots. In the first part, a feature vector consisting of the headings of the human posture and the locations of the humans and robots is created by using video processing techniques. The second part is related to estimating the current intention of a human participant via HMM models and to reshaping the current intention into another intention. The system is tested in a real experimental environment including humans and robots, and the results in the recorded videos are given at the end of the paper.

Key words: Human–robot interaction, intention estimation, intention reshaping, hidden Markov models

1. Introduction

In this paper, interactions between humans and robots are computer vision-based and without communication, which could be gesture-based and verbal. Although our method only builds upon the initial estimates of human intentions, in order to change them voluntarily by robot-triggered interactions, we will first review the problem of intention estimation in the literature. Furthermore, we will discuss the problem of intention estimation and our design of the experimental environment, which contains particular equipment. We will then discuss the sparse works on changing one’s intentions, whether by robots or humans.

Two intelligent agents that interact biomimetically are required to predict each other’s intention and either morph their own actions to the other’s intention or strategically change the other’s intention to achieve the desired change. This strategic change is termed “intention reshaping” [1]. Recently, several works have emerged on changing intention. The literature abounds with works on plan/goal recognition and intention estimation. For example, [2] used a facial expression changing in time in order to estimate the human intention. Moreover, gesture recognition and action recognition are studied for human intention estimation [3–6].

The next section overviews intention recognition and estimation. This problem is characterized by prototypical phases of problem definition and classification issues. For instance, if the interpretation of human

*Correspondence: durdu.1@selcuk.edu.tr

activity in a video is problematic, the classification uses a learning algorithm to solve this problem according to the training database of the human. The training database is constructed with computer vision techniques such as background subtraction and object detection.

Recognition of human intention generally requires detecting a characteristic posture of intention feature and then classifies it against known features for recognition. In other instances, human motion heading is characteristic of its intention when the heading target is identified, similar to heading to a cafe. In our computer vision-based system, a camera captures a view of the human–multirobot environment that includes hardware components, such as mobile robots, and a human agent. For each image captured, the software component processes the image to generate a feature vector. In that case, the vision-based inference system (VIS) maps the feature vector to a low-dimensional one that represents the intention characteristics of the human. This reduced space provides a simple yet powerful representation.

In this work, the process of estimating current human intentions and reshaping them voluntarily by robot motion decisions is a continuous and complex process that requires robots to be equipped with decision-making models, based on interpretations of human-to-human interactions that have been taught or modeled as rules or by any other closed-form mathematical ways. Considering human-to-human conversations as a form of human-to-human interaction, humans try to predict the context-dependent reaction of each other within a dialogue, by estimating the intention and the next direction that the conversation will take [7].

The first portion of this paper pertains to prototyping issues, both hardware and software, in intention recognition and estimation, as found in the literature. Reshaping human intention by robotic interaction is a relatively new perspective, and very few works exist on modeling, reshaping, and generating intentions. Although [8] introduced the reshaping of human intention via robots, the general purpose of this study is to examine the psychological statuses of humans with human–robot interaction (HRI). These psychological statuses, which were mentioned in [9], are physical, design, and intentional stance. In this paper, we demonstrate how a computer vision-based approach is a benchmarking approach to estimating human intention by using real environment examples. However, the benchmarking perspective in our work recognizes that reshaping the current human intention can be achieved by robots carrying out decisions on how, where, and when to move in certain directions in order to reshape the current intention of the person in the environment in a desired manner. We will provide a brief overview of the limited works in the literature as well as introduce in detail and discuss our novel yet benchmarking perspective, both in hardware setup and in software.

An overview of the current literature on the problem of intention estimation and intention reshaping is presented in the following section. Afterwards, the real experimental environment and methodology used in our algorithmic approach will be explained. At the end of the paper we will discuss the results of our approach for reshaping intentions by robot motions and for future studies.

2. Literature review

For many years, philosophers have considered human intention as inferred from human activities and their effects on the environment [9–12]. The literature on robotics contains many examples recognizing human intentions in a robotic environment. The reshaping of human intention, however, has not been addressed in the literature on HRI. Therefore, in this paper we seek an answer to the following question: Why does intention-reshaping need to be studied? We believe that an example can illustrate the importance of this problem. Let us assume that the environment is a dangerous area and requires certain risky tasks, and that we put robots in the environment to eliminate dangers to humans by avoiding dangerous areas. Although textual descriptions in the form of notes may keep humans away from dangerous areas, not everybody pays equal attention to textual information; in

fact, the person may be illiterate. In this case, the motion of the robots may attract the attention of the person and reduce the risk of getting hurt. Such systems should, however, be rapid and reliable.

Intention estimation is usually realized by control commands of the human in a game or simulation setting [10,13–15]. One of the initial studies on plan recognition was developed for understanding stories told within a natural language story [16]. Kautz and Allan [17] extended this work to a general model for the recognition of general plans. The improved system of this study is based on a set of observations and action taxonomy with simplicity constraints in the environment. Charniak and Goldman [18] approached the inference estimation by measuring uncertainty during plan recognition. In their work, decreasing the number of top-level plans provided them with computational advantage, whereas limiting its application to real-world applications inherently included uncertainty in both the perception and actuation. In addition, they performed Bayesian updating to select the most likely interpretation by using the set of observed actions. A related work [19] examined the Bayesian networks for traffic monitoring problems. This work was extended to examining the intention of pedestrians on a curb, in order to develop a cognitive driver assistance system to warn the driver according to the pedestrians' intentions [20]. They developed a driver assistance system that captures and analyzes the traffic situation in order to alert the driver in case a pedestrian is a potential danger. The system recognizes the action intentions of the pedestrian by using the video database taken from natural traffic scenes. In a related work, Tahboub [13] introduced a compliant human-machine interaction architecture using a dynamic Bayesian network for probabilistic intention inference. Using that system, a human can remotely command a mobile robot. The study used the hidden Markov model (HMM) for the action recognition problem. Schrempf et al. [21] presented a different approach to creating a Bayesian network for estimating intentions without a learning framework. Two other studies [22,23] also used HMM to recognize human behavior via an online probabilistic algorithm, while others [24,25] designed an experimental room for detection of human intention. They used an ID4-based leaning system to recognize the human intention, complete with sensors for the estimation of human position and sounds in the environment. Their studies, although constrained in their capacity, are more capable of intention recognition than the system developed by Koo and Kwon [26]. In this paper, unlike in the above examples, all given data are obtained from a real environment using 2 cameras that extract human as well as robot intentions. The capability of the estimation is highly dependent upon (and directly proportional to) the feature vectors.

In the literature, examples of reshaping intention are found in psychology. Webb and Sheeran [27] studied whether changing behavioral intention causes behavior change. Forty-seven experimental tests of intention-behavior relations were used to achieve this goal. In those experiments, participants were assigned randomly to a process that increased the power of their own intentions relating to a control situation, and the differences of the forward behavior were compared. Metaanalysis was used to predict the affectless ranges of the intention-behavior relation. After the impact of the interventions on intention and behavior was quantified in metaanalysis, the results indicated that the interventions could change intention and also reshape behavior.

Meltzoff [28] investigated whether children would change their intention towards what an adult intended to do or shape their intentions based on what an adult was currently doing and prepared 2 experiments for this purpose. In Experiment 1, children were confronted with an adult who tried to perform obvious target actions but did not succeed. Experiment 2 matched the reaction of the children to an adult towards an inanimate item. Taken together, 18-month-old children have adopted a key element of folk psychology: although inanimate items are not understood, adults are understood within a framework that includes intentions and aims. Terada et al. [8] reviewed the effect of reactive movements accomplished by a nonhumanoid robot, such as a chair and cube,

on human intention attribution. In their study, humans tried to understand the behaviors of an artifact with respect to its goal, which depends on how humans attribute intention to the artifact. The aims of this study were to observe how humans behave after a robot movement and whether or not there are differences caused by the appearance of the robot in human intention attribution. Their study was related to the investigation of human intention attribution toward the movements of robots. The novel contribution of this paper is an introduction to intention reshaping in the area of HRI research. Therefore, the actual aim of this study is to realize change in current intention via intentional robot movement within a real scenario.

In this paper, our novel perspective is to develop a hierarchical system by decoupling the low-level object detecting-tracking algorithms from the high-level intention estimation and reshaping. The system also makes decisions about the changing human intention using robot movements. The low-level detection and tracking algorithm focuses on a multirobot interacting with a human. High-level capability provides a solution to the human-machine intention problem by using learning algorithms such as HMM and principle component analysis (PCA).

3. System design and methodology

3.1. Experimental setup

We designed an experimental room that contains humans and robots to realize our intention reshaping problem and we proposed HMM to recognize human intention. The robots, namely chair and stair robots, move to reshape human intention. For example, since the human does not know the purpose of the robots in the environment, s/he may change intention according to the movement of the robot. The real-time system has to make fast decisions and show high efficiency, since our real-time application primarily aims to recognize the intention and then reshape the recognized intention. Therefore, we designed an HMM based on human location and human posture, which were taken from the video frames. As image/video processing was not used in any of the above-mentioned robotic applications, our method may be also referred to in HRI applications. An advantage of the usage of computer vision techniques is that the parameter of the intention recognition may be increased. For example, we could use both the location and the posture of the human, whereas Koo and Kwon [26] could only use the location, since they simply used the sensor. On the other hand, the advantage of sole usage of the sensor is faster decision-making, according to our system. However, our system performs more accurate intention recognition.

A designed experimental room (Figure 1) monitors the intention reshaping of the human participant by motion of the robotic chair and shallow stair. The room, which includes a bookshelf, worktable, mobile chair/stair robots, and a coffee machine, is observed with cameras. The cameras give us a top and side view of the room. While the locations of the robots and human are provided by the top-view camera, the heading of the human posture is obtained by the side-view camera. Chair and stair robots, which are computer-controlled, perform particular activities in order to reshape the intention of the human participant entering the experimental room (Figure 1). The chair and stair robots move with 2 wheels driven by a DC motor (12 V). The experimenter/computer remotely controls the robots by using the radio frequency communication system in a room different from the experimental room. According to our scenario, human participants that have information about our application act independently in the experimental area.

3.2. Methodology

In this paper, the goal of our design is to develop a hierarchical system by decoupling the low-level object detecting-tracking algorithms from the high-level intention estimation. The system also makes decisions about

the changing human intention with the movement of the robots. The low-level detecting and tracking algorithm focuses on the multiple robots and a human. High-level capability results in a solution to the human-machine intention problem by using learning algorithms such as HMM and PCA.

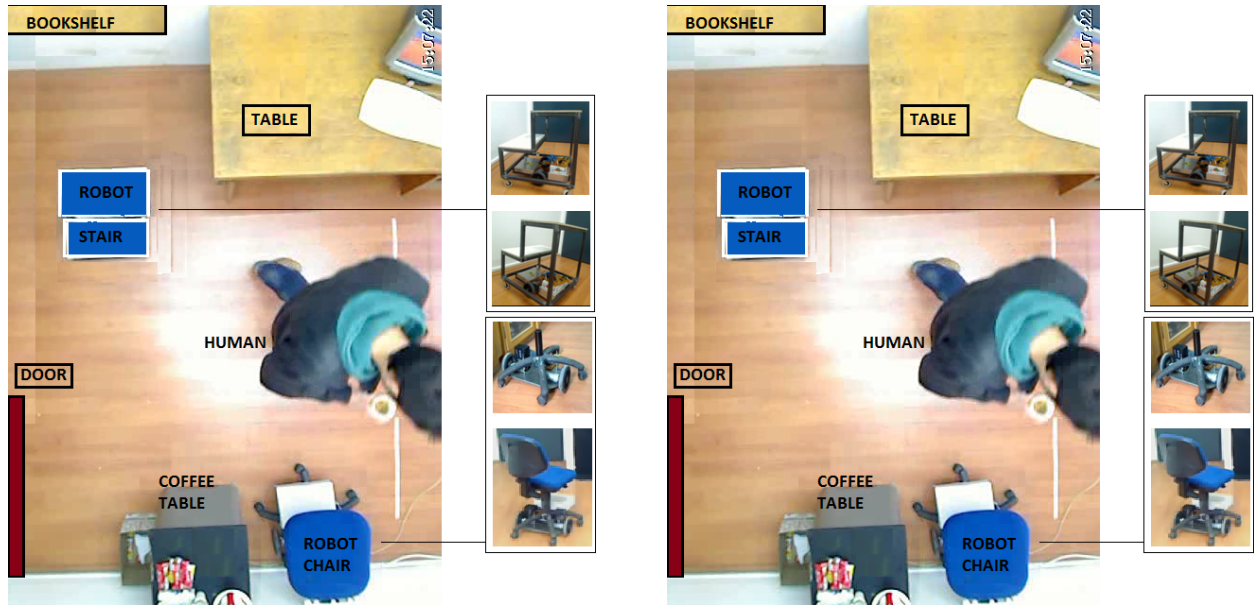


Figure 1. Experimental environment.

Figure 2 demonstrates a flow chart for the entire system. According to this flow chart, the algorithm of our system, which reshapes the current human intention via the movements of the mobile robots, is described in the following text.

In the first step, background subtraction is used to construct the feature vector, which includes the locations of the movable objects (human and robots) and the heading information of the human posture. In a stationary background, the following process explains how the movable objects are detected in video frame sequences: 1) separating the pixels that represent the movable objects from the pixels that represent the background, 2) grouping together the pixels that represent the individual human and robots and calculating the appropriate bounding box for both, 3) matching human and robots in the current frame with those in the previous frame by comparing the bounding boxes between frames, and 4) using a PCA method that classifies the heading of the human posture at each frame.

The second step includes high-level stochastic learning by using low-level information. In this step, the observable human action states are the instantaneous movements of the person with the intention. Although intention estimation is not possible from observed human action only, one way of estimating it is by observing the human action sequences. We assume that a plan or action has an intention until it is carried out. In this section, we explore the use of probabilistic methods to learn and estimate human intention. We propose an approach to learning the structure of intention estimation from sequences of human actions through the application of HMM with heading of the human posture and location-dependent observation model.

The third step examines whether the robots change human intention or not. In this step, the system compares 2 predicted human intentions: one is estimated human intention, when the robot starts to move, and the other is human intention after the movement of the robot.

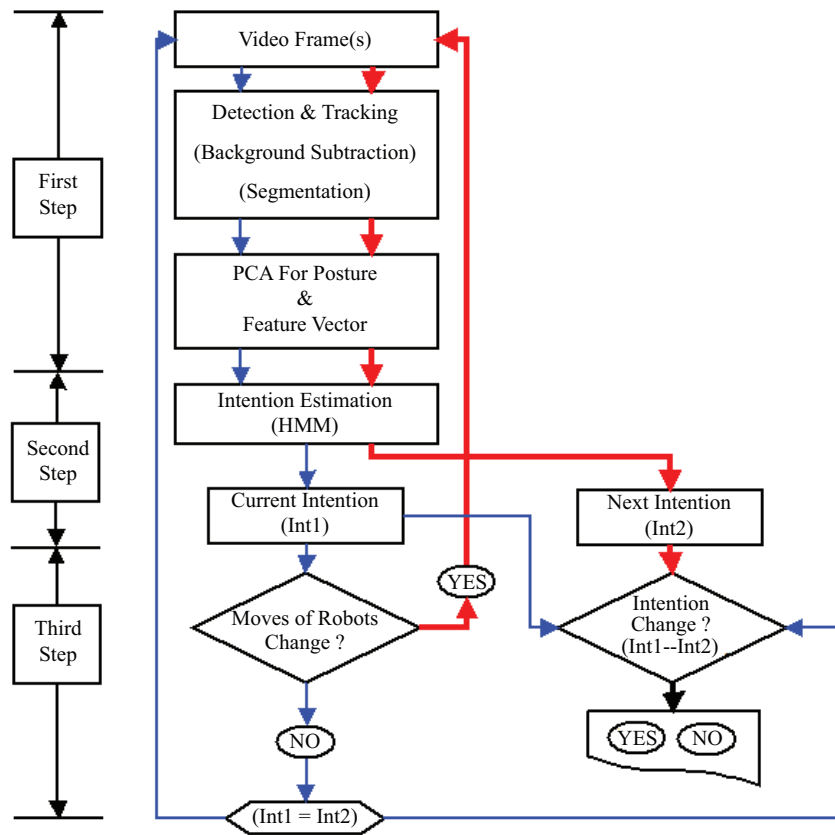


Figure 2. Flow chart of the total system.

3.2.1. Feature vector construction

The first requirement for feature extraction of human motions is the ability to track the human in question and generate his motion trajectory in order to relate it to a contextual meaning for intention estimation that will be introduced in Section 3.2.2. The characteristic features of human actions in the room, used to estimate and reshape current intention, consist of the heading of the human posture and the location of the human and robots. Therefore, the following sections explain how to find the location of the objects and how to determine the posture of the human, respectively. The final section gives the total information about locations and headings of the posture that is gathered in the feature vector.

3.2.1.1. Detection of moving objects

The detection of moving objects (robots and human) in a video frame is an important part of this application. For this purpose, background subtraction is realized after estimating the background image by using the first few frames of the video.

We used the autothresholding method in order to determine which pixels correspond to the moving objects in the scene. This method uses the difference in pixel values between the normalized input image and the background image. In the detection step, the morphological closing merges the object pixels that are close to each other to create blobs. For instance, pixels that represent a portion of the human’s body are grouped together. Next, we calculated the bounding boxes and the centers of these blobs. Finally, in the tracking step, we determined the locations of the specific robots and the human from one frame to another. The system

compares the predicted locations of the bounding boxes to the detected locations. This enables the system to assign a unique color to each robot and to the human (see Figure 3 for detection and Figure 4 for tracking of the detected points). Figure 4 shows the trajectories taken from a human participant and robots in the same environment.

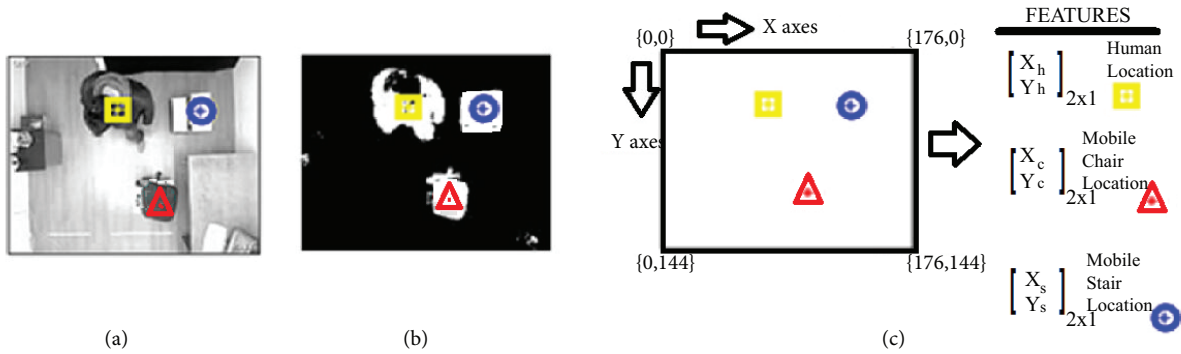


Figure 3. Detected location of moving objects (human and robots): (a) the centers of the moving objects in a video frame are shown as colored points, (b) locations after background subtraction, and (c) locations on the simulated image.

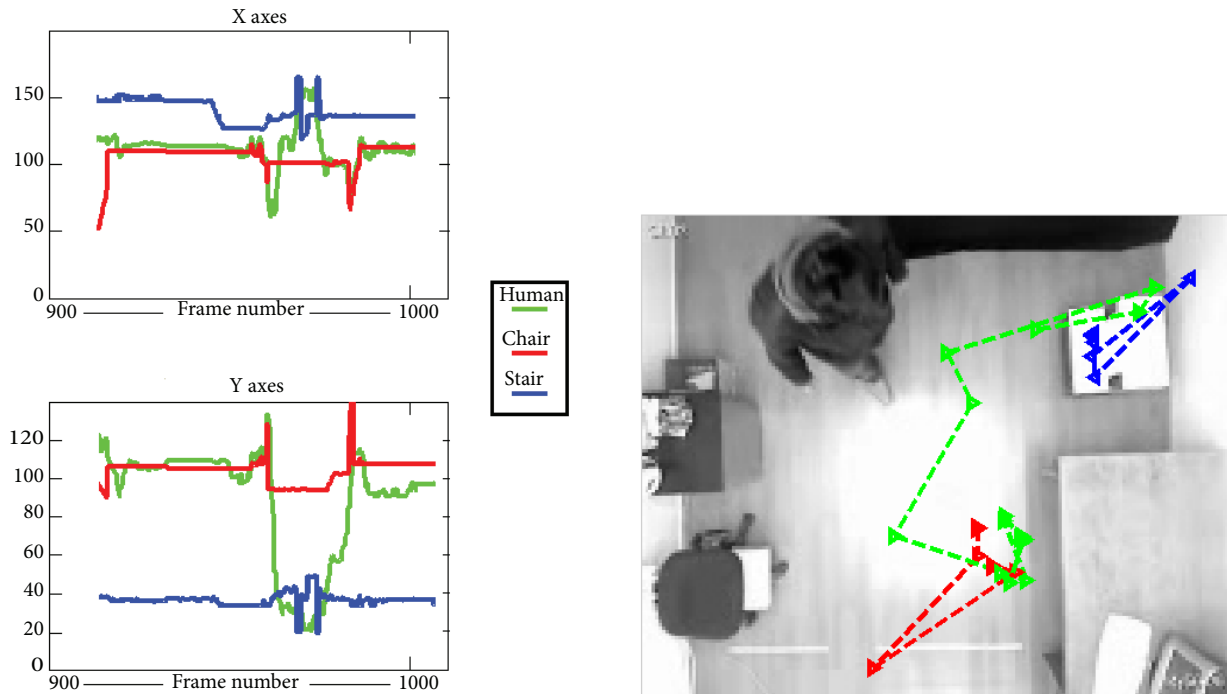


Figure 4. Example frame trajectories for the location of moving objects (human and robots) in a video made by a participant in the environment (between 900 and 1000 frames).

The first stage in the presented algorithm is to construct the feature vector. The feature vector, comprising sequences O , can be formed by using the output trajectories of previous system simulations (similar to Figure 4). The observation model in our approach is labeled by the grid cells from 144×176 to 24×25 . Each grid cell that substitutes the location of the human participant is also known as an observation output. Figure

5 illustrates a set of labeled cells in our setup. Given a trajectory represented by sequences of length T , by evaluating the heading of human postures we can label them as intentions with respect to actions (Figure 6). This is explained in detail in Section 3.2.2.

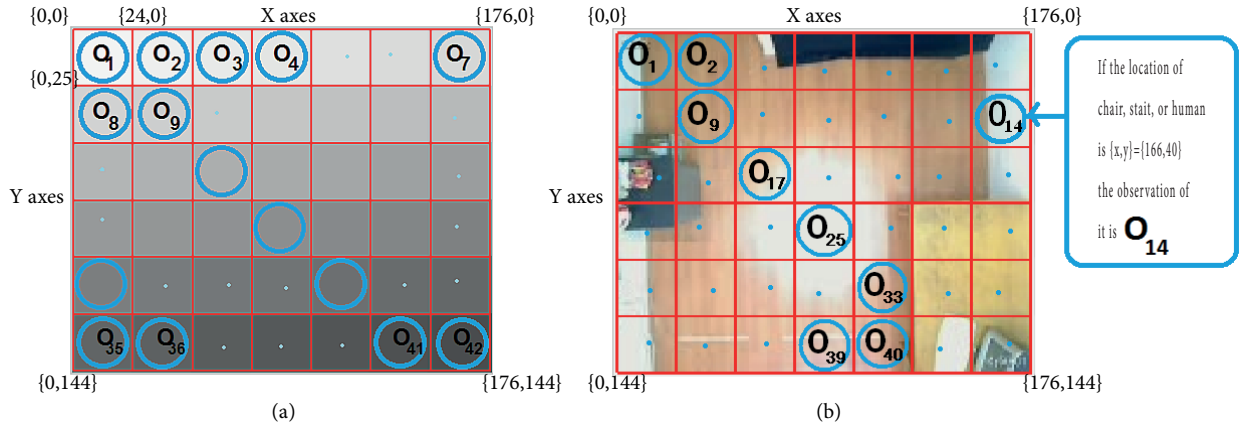


Figure 5. Labeling the location trajectories of human and robots.

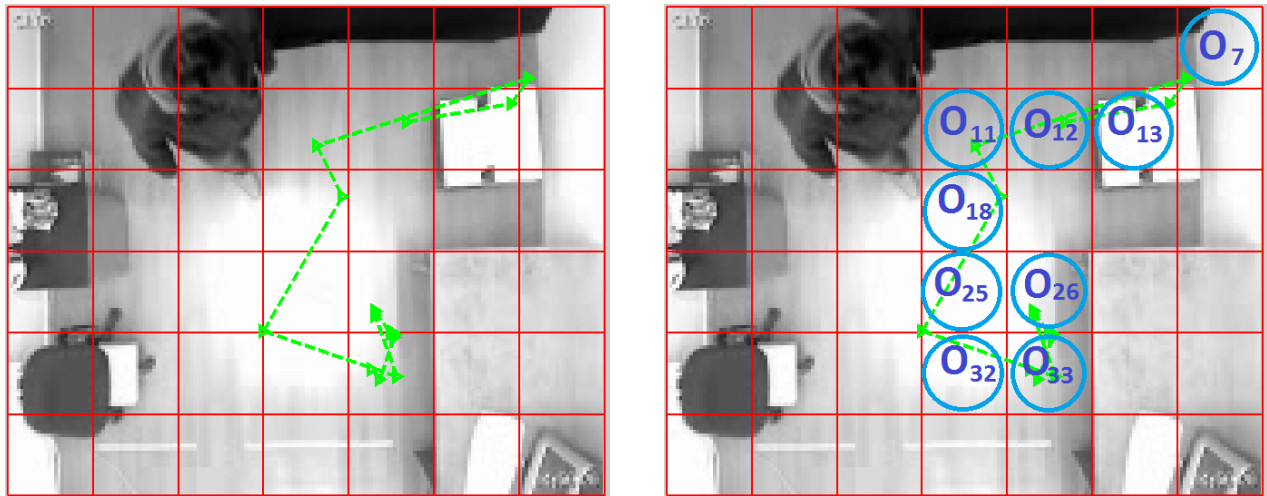


Figure 6. An example of an observation cell trajectory. This example of training trajectory data is for “going to the worktable”.

3.2.1.2. Heading information from human posture via PCA

In this section we explain the extraction of the heading information from human posture. In the literature, researchers have studied several techniques, such as self-organizing maps, fuzzy C-means, K-means [29], and correlation filter classifier [30], for the analysis and recognition of human postures. As an alternative to these, PCA for feature mapping is another popular approach [31]. Identification of principle components, also known as orthogonal axes, is the goal of the PCA [32,34]. In this paper, we use the PCA-based posture analysis approach, which identifies and classifies postures into one of the following heading classes: “heading left”, “heading right”, and “keeping central heading” [1]. These 3 different headings are chosen according to different intentional regions of the environment. For example, if we have a frame taken from the side camera, we find

the “heading right” as “human heading to the bookshelf”; however, if the direction of the human changes to the opposite side in the next frames as “heading left”, then we conclude that the human is “heading to the coffee machine”. Therefore, the heading of the human posture is important for estimating action-based human intention.

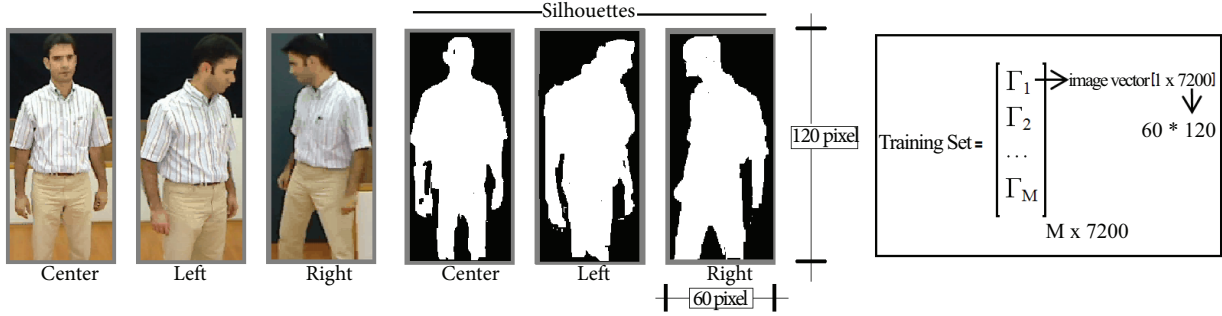


Figure 7. The training set example for “heading left”, “keeping central heading”, and “heading right” postures of the human.

The PCA algorithm used for generating heading information from human posture is first trained from a set of the human heading vectors $\Gamma_1^T, \Gamma_2^T, \dots, \Gamma_M^T$, where T stands for transpose, each of which is a 7200-dimensional row matrix generated by concatenating the columns of a 60×120 array (Figure 7). These vectors are mean-normalized to generate $\Phi_j = \Gamma_j^T - \Psi$, where Ψ is the mean vector. Given these normalized vectors, a low-dimensional human heading space is constructed by selecting a set of eigenvectors with the largest eigenvalues. This eigenspace is referred to as the human heading space, spanned by those eigenvectors that are used to test incoming posture by computing the Euclidian distance ε_m between the input images projected into the human heading space and the image of the mean-valued “keeping central heading” training data:

$$\varepsilon_m = \|\Omega - \Omega_m\|, \quad (1)$$

where $\Omega^T = [e_1, e_2, \dots, e_L]$ explains the contribution of each eigenvector related to the input image. We choose Ω_m as the mean of the eigenspace representations of several training images belonging to the same class of posture. In our application, since we partition our test environment into 3 directional areas, we define 3 headings from human postures; thus, we calculate 3 different Ω_{mi} values $i = 1, 2, 3$ from training data, as shown in Table 1. When the minimum ε_{mi} is below a certain threshold, we conjecture the detection of a heading. Otherwise, the heading of the human posture is classified as having undetermined heading, and the procedure continues with new state data.

Table 1. The calculated Euclidian distances for posture recognition.

Euclidian distance with posture	
Frame number = 340	
	ε_{m_i}
Heading left	0.2221 (1)
Heading right	0.4764 (0)
Keeping central heading	0.5292 (0)
Posture result (PCA)	“Heading left”

The complete information about moving human and robot agents in the scene generates the feature vector. For a classified heading of human posture, we use a binary coding and state the detected feature as 0 and 1, corresponding to true or false. For instance, if the determined posture is “heading left”, then vector V_h of the heading of human posture becomes $[1\ 0\ 0]^T$ (Table 1). The constructed feature vector is shown in Figure 8.

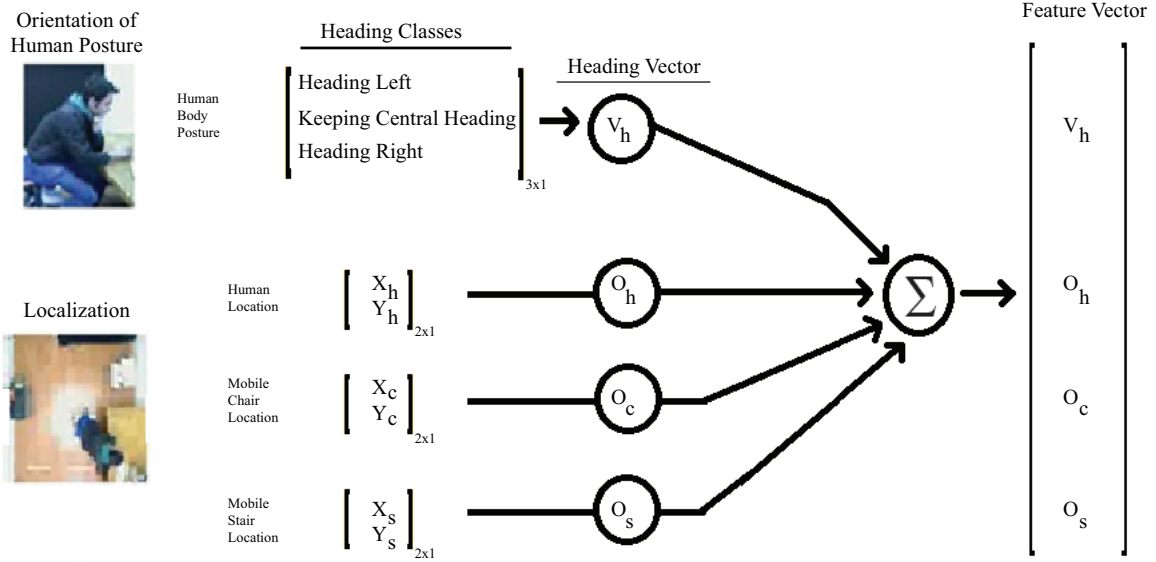


Figure 8. Feature vector used for intention estimation.

3.2.2. Intention estimation with HMM

We use the HMM, which is employed for estimating intentions, to model the motion patterns of the human. Before we detail how the estimation is achieved, we give a formal explanation of HMM. A discrete HMM, $\lambda = (AB\pi)$ can be described as N hidden states, $H_1 H_2, \dots H_N$; M observations (outputs/emissions), $O_1 O_2, \dots O_M$; transition matrix $A = \{a_{ij}$ that gives probability of transition from H_1 to H_j : $a_{ij} = p(H_j|H_i)$, observation matrix that encodes probability of observing O_j at H_i : $p(O_j|H_i)$; and, finally, prior information $\pi_1, \pi_2, \dots, \pi_N$, which is the starting probability at state H_i : $p(H_i)$. These components of the HMM are subjected to common probability rules, including $\sum_{i=1}^N a_{ij} = 1$, $\sum_{i=1}^N b_{ij} = 1$, $\sum_{i=1}^N \pi_{ij} = 1$, where $a_{ij} \geq 0$, $b_{ij} \geq 0$ and $\pi_i \geq 0$ [33,34].

Table 2. Intention list.

Number	Intention
H1	Getting a book from the bookshelf
H2	Resting on the worktable
H3	Getting a coffee
H4	Exploring the robots

In our experiment, ‘going to the coffee table’, ‘going to the worktable’, ‘discovering the environment’, and ‘going to the bookshelf’ are defined as observable human actions. Each action, consisting of consecutive image sequences of the human, is labeled with a human intention in Table 2. The HMM model in Figure 9 has 4 hidden states that symbolize human intentions. In our discrete HMM, we have 4 hidden states, $H_1H_2H_3H_4$; output symbols that are 42 observation cells $O_1O_2, \dots O_{42}$; and one heading information, V_h , at the last frame of a sequence. According to our method, a sequence of consecutive output symbols constructs human actions (mentioned earlier in this paragraph), and each human action replaces a hidden state that is one of the human intentions in Table 2 (Figure 9). Output symbols symbolize observation cells, which are forty-two 24×25 -sized grid cells obtained from the division of 144×176 -sized image frames. The feature vector, described in Section 3.2.1, consists of the human location and heading. In this section, we use the feature vector to calculate the learning parameter of our HMM model. Training sequences with observation cells are constituted by using each location of the human in the training data, and the human heading at the end of the sequence (in the last frame) is also added to this training sequence for each training person. For example, if human location in an image frame $(x_1, y_1) = (11, 12)$, then this location gives observation cell O_1 , and if human location in the next image frame $(x_2, y_2) = (28, 12)$, then the related observation cell will be O_2 . In this case, if we generate a sequence with these 2 samples, our sequence will be $[V_h O_2 O_1]$, where V_h is the heading of the human posture in the last frame.

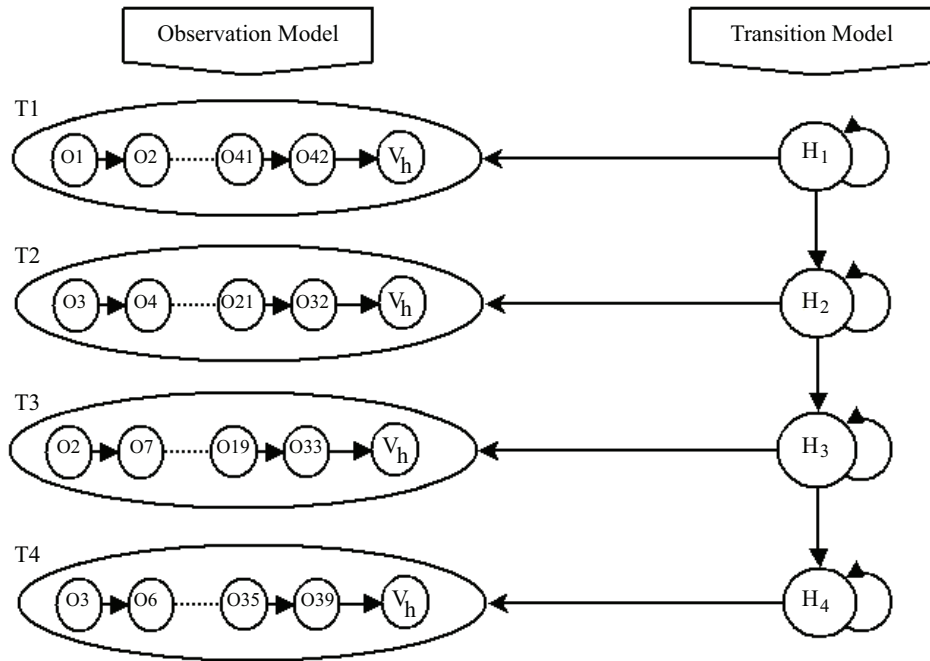


Figure 9. The hidden Markov model.

In our HMM model ($\lambda = (AB\pi)$), the probabilistic observation matrix is $B = \{b_{ij}$, which encodes the probability of observing O_j at H_i : $b_{ij} = p(O_j|H_i)$. The observation matrix (B) is constituted as seen in Figure 10. In this figure, an example sequence from the training data is used. As shown in Figures 10 and 11, we counted human trajectories in each observation cell for all training sequence data and calculated emission probabilities (b_{ij}). For the calculation of the transition probabilities, which are $a_{ij} = p(H_j|H_i)$ in

the transition matrix that is $A = \{a_{ij}\}$, we used consecutive state sequences in the training data. For example, $a_{12} = p(H_2|H_1)$ indicates the probability of transitions from H_1 to H_2 in the total training data. Each hidden state has probabilistic weight $p(H_i)$ assigned from training trajectories.

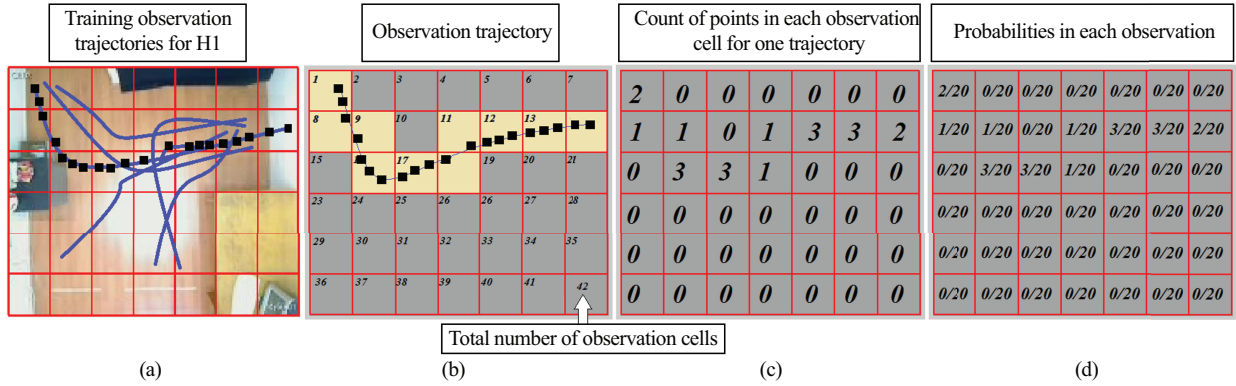


Figure 10. Example construction of the probabilistic observations. For H1 in Table 2, training trajectories in (a) are used. (b) shows one of the trajectories in (a), (c) shows the counted frames in each observation cell, and (d) indicates the probability distribution for H1 in each observation cell.



Figure 11. Trajectory examples of training data: (a) “resting on the worktable” (H2 in Table 2), (b) “getting a coffee” (H3 in Table 2).

The training data set in the learning phase of the algorithm and comprising sequences is formed using the trajectories from the previous simulation. Given a set of human trajectories represented by the sequences of length T ($T = 7$ for the example in Figure 12) and human headings at the last frame, we can relate these trajectories and human headings to intentions (see Figure 12). In Figures 12a and 12b, the intention estimation example includes a trajectory given by a heading of human posture.

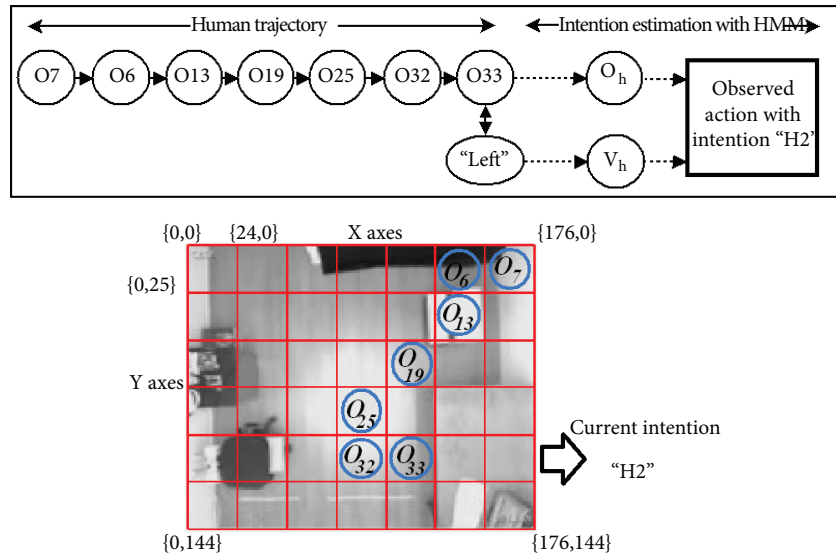


Figure 12. Example for Intention 2 (from Table 2): (a) trajectory data for testing, (b) action observation.

4. Experimental results

We tested the scenario designed to change human intention with 15 participants. In Figure 13, the results demonstrate that the mobile robotic agents reshaped human intention in the experimental environment. While the human participant is interested in the computer in Frame 270, our system detects the current intention as the second intention in Table 2.

After that frame, the stair robot changes its location in the environment. The human participant initially tries to understand what happened. Next, after Frame 310, s/he gets up from the chair and moves to get a

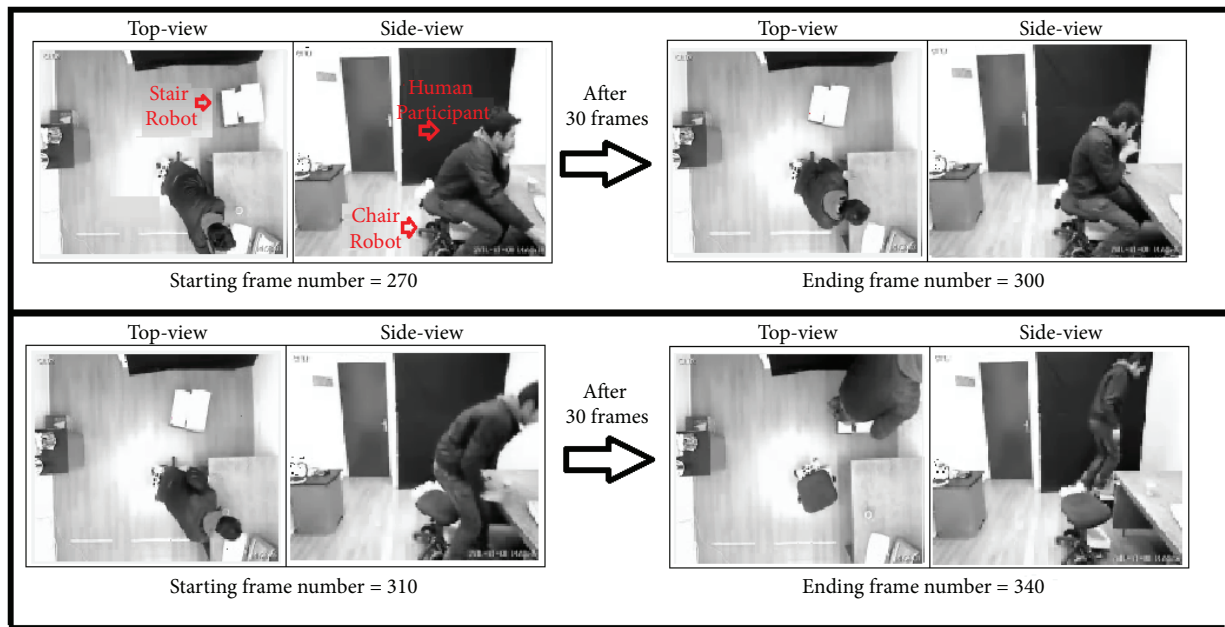


Figure 13. Results for the human participant. a) Frame captures of the experiment. b) (x,y) positions of the human participant, mobile stairs, and intention reshaping.

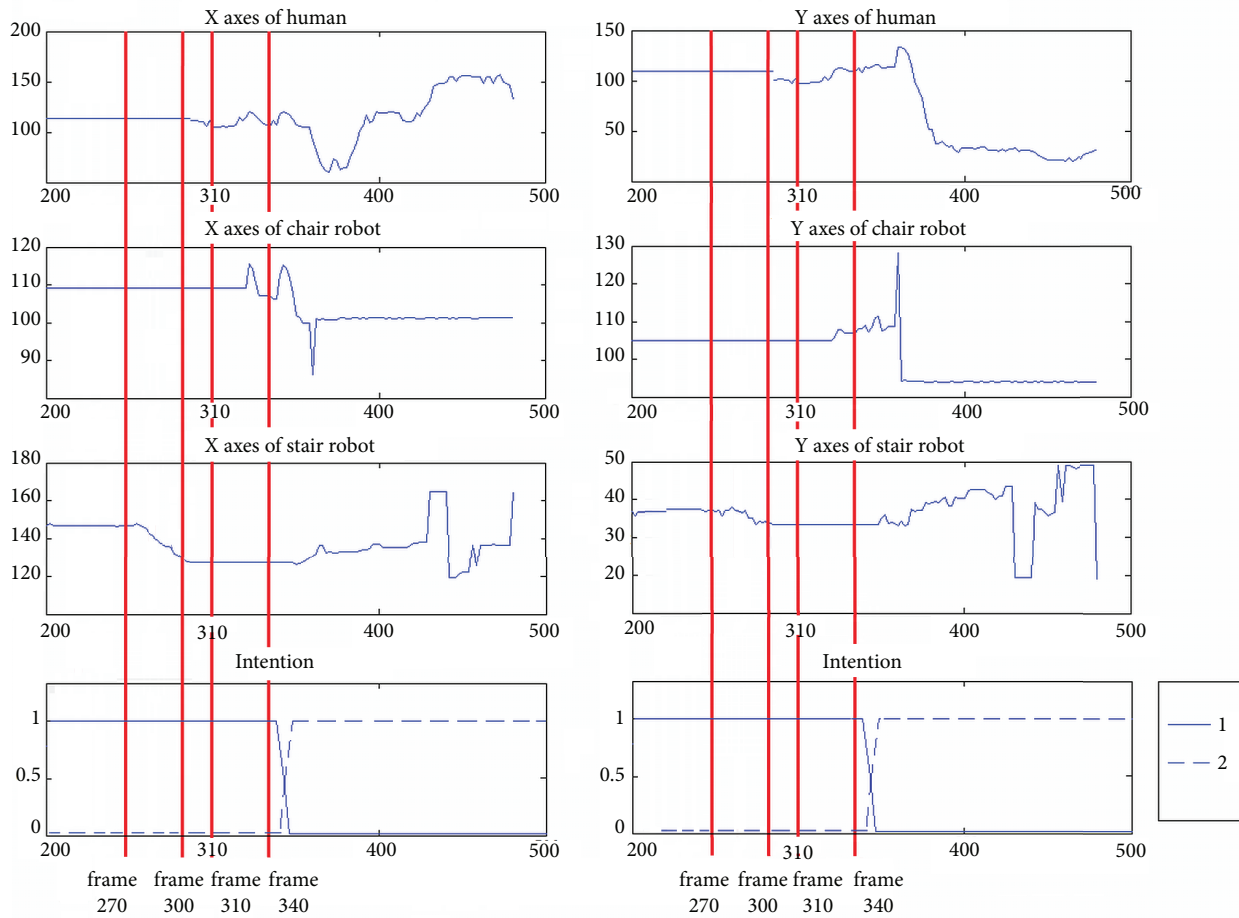


Figure 13. Continued.

book from the bookshelf (the first intention in Table 2), since the stair robot in front of the bookshelf has the movement (this situation shows that the desired intention is the first intention in Table 2). If the human goes to get coffee from the coffee machine instead of getting a book from the bookshelf, reshaping the intention will not be successful due to the difference between the current and desired intention. Some HMM results of the testing sequences (Figure 13) used to estimate the current intention are shown in Table 3.

Table 3. Results for a person.

Action observation with HMM (log-likelihood)		
Trajectory	Between frames	Between frames
	270 and 300	310 and 340
'Going to bookshelf' (Action 1)	-0.507	-0.363
'Going to work-table' (Action 2)	-0.385	-0.424
'Going to coffee machine' (Action 3)	-0.465	-0.534
'Discovering the environment' (Action 4)	-0.563	-0.519
Action result (HMM)	Action 2	Action 1
Intention		
Intention (from Table 2)	H2	H1

The experimental results of this example suggest that the current human intention will be successfully reshaped by the movements of the robotic agents. This success, however, was not true for all participants. Considering that the participants did not have knowledge about the purpose of the experiment and the room, the first movement of the robots attracted their attention. In particular, we used this attraction to show that changing the human intention was possible. Figure 13 shows the intention-reshaping result for an individual. While the person goes to take a book from the bookshelf after the movement of the stair robot in Figure 13, the other participants may go to take a coffee or discover the robots in the environment. For example, some participants did not respond to the robot movement. After a certain amount of time, during which the participants understood the robots, we observed that most participants did not easily change their intention with the movements of the robots.

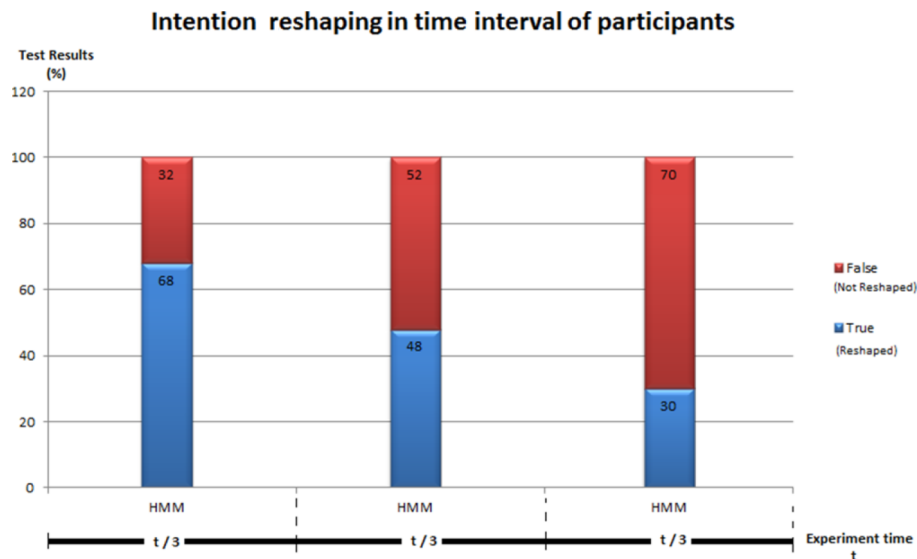


Figure 14. Intention reshaping of participants in the experiment time interval.

Considering that the participants did not have knowledge about the purpose of the experiment and the room, the first movement of the robots attracted their attention. In particular, this attraction is used to show that changing human intention is possible. Figure 14 shows the efficiency of our scenario, which is designed to reshape human intention, in the time interval of each participant. After a certain amount of time, during which the participants understood the robots, we observed that most participants did not easily change their intention with the movements of the robots. Before the experiment, the participants did not have any information about what was tested in this area. After the experiment, the experimental video of each participant was watched with him/her and the experimenter. The real intentions of the participants were checked by our intention estimations. Figure 14 also reflects this evaluation survey.

5. Conclusion

In this paper, we demonstrated that reshaping human intention can be observed in a human–robot interactive experimental environment. A sequence of video frames is used in our approach and these are processed according to the scenario at hand. Our future work is to realize this scenario in real-time frames, instead of video frames, for intelligent HRI systems. To the best of our knowledge, this study with humans and robots is the first work on reshaping human intention; therefore, we expect that our different approach towards reshaping human intention will be useful for future research on HRI.

We will continue our efforts to realize an online system that is reliable and fast. We conjecture that such a system can be applied to security systems, information services, assisted services, etc. Particularly, attracting the attention of elderly people suffering from Alzheimer disease or of infants interacting in a daycare setting may eliminate potential dangers by changing their intention, which would otherwise result in getting hurt.

At the present time, the number of terrorist attacks is high. The security of certain locations, such as airports, railway stations, and large stadiums, is important because many people use these public spaces. Particularly, after the September 11 attack airport security has improved. Sometimes the police require time to defuse terrorism. In this case, robots that are modeled with different shapes and purposes may be added to security measures by using intention reshaping.

On the other hand, intention reshaping via robots can be used for training animals or chastening inmates in prison. More examples can be given for this new research area. In the future, if intention estimation is fast, reliable, and robust, intention controllers will be a commercially available off-the-shelf component in stores. We hope that our novel contribution to science will be the starting point of something useful for humanity in manufacturing.

References

- [1] Durdu A, Erkmen I, Erkmen AM, Yilmaz A. Morphing estimated human intention via human–robot interactions. In: Proceedings of the World Congress on Engineering and Computer Science; 19–21 October 2011; San Francisco, CA, USA. Hong Kong: IAENG. pp. 354–359.
- [2] Miyake T, Matsumoto T, Imamura T, Zhang, ZE. Estimation of facial expression from its change in time. *ICIC Exp Lett* 2011; 2: 641–645.
- [3] Daprati E, Wriessnegger S, Lacquaniti F. Kinematic cues and recognition of self-generated actions. *Exp Brain Res* 2007; 177: 31–44.
- [4] Loula F, Prasad S, Harber K, Shiffrar M. Recognizing people from their movement. *J Exp Psychol Human* 2005; 31: 210–220.
- [5] Sevdalis V, Keller PE. Self-recognition in the perception of actions performed in synchrony with music. *Ann NY Acad Sci* 2009; 1169: 499–502.
- [6] Sevdalis V, Keller PE. Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Conscious Cogn* 2010; 19: 617–626.
- [7] Clarke TJ, Bradshaw MF, Field DT, Hampson SE, Rose D. The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception* 2005; 34: 1171–1180.
- [8] Terada K, Shamoto T, Mei H, Ito A. Reactive movements of non-humanoid robots cause intention attribution in humans. In: *IEEE 2007 Conference on Intelligent Robots and Systems*; 29 October–2 November 2007; San Diego, CA, USA. New York, NY, USA: IEEE. pp. 3715–3720.
- [9] Dennett DC. *The Intentional Stance*. Cambridge, MA, USA: MIT Press, 1987.
- [10] Tahboub KA. Intelligent human-machine interaction based on dynamic Bayesian networks probabilistic intention recognition. *J Intell Robot Syst* 2006; 45: 31–52.
- [11] Bratman M. *Intention, Plans, and Practical Reason*. 2nd ed. Cambridge, MA, USA: Harvard University Press, 1999.
- [12] Bratman M. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge, UK: Cambridge University Press, 1999.
- [13] Tahboub KA. Compliant human–robot cooperation based on intention recognition. In: *IEEE 2005 International Symposium on Intelligent Control*; 27–29 June 2005; Limassol, Cyprus. New York, NY, USA: IEEE. pp. 1417–1422.

- [14] Iba S, Paredis CJJ, Khosla PK. Intention aware interactive multi-modal robot programming. In: IEEE 2003 International Conference on Intelligent Robots and Systems; 27–31 October 2003; Las Vegas, NV, USA. New York, NY, USA: IEEE. pp. 3479–3484.
- [15] Yokoyama A, Omori T. Modeling of human intention estimation process in social interaction scene. In: IEEE 2010 International Conference on Fuzzy Systems; 18–23 July 2010; Barcelona, Spain. New York, NY, USA: IEEE. pp. 1–6.
- [16] Wilensky R. Planning and Understanding: A Computational Approach to Human Reasoning. Reading, MA, USA: Addison-Wesley, 1983.
- [17] Kautz HA, Allan J. Generalized plan recognition. In: AAAI 1986 National Conference on Artificial Intelligence; 11–15 August 1986; Philadelphia, PA, USA. Palo Alto, CA, USA: AAAI. pp. 32–38.
- [18] Charniak E, Goldman R. A Bayesian model of plan recognition. *Artif Intell* 1993; 64: 53–79.
- [19] Pynadath D. Probabilistic grammars for plan recognition. PhD, University of Michigan, Ann Arbor, MI, USA, 1999.
- [20] Schmidt S, Färber B. Pedestrians at the kerb—recognising the action intentions of humans. *Transport Res F-Part* 2009; 12: 300–310.
- [21] Schrempf OC, Albrecht D, Hanebeck UD. Tractable probabilistic models for intention recognition based on expert knowledge. In: IEEE 2007 Conference on Intelligent Robots and Systems; 29 October–2 November 2007; San Diego, CA, USA. New York, NY, USA: IEEE. pp. 1429–1434.
- [22] Bui H, Venkatesh S, West G. Policy recognition in the abstract hidden Markov model. *J Artif Intell* 2002; 1: 451–499.
- [23] Aarno D, Kragic D. Layered HMM for motion intention recognition. In: IEEE 2006 Conference on Intelligent Robots and Systems; 9–15 October 2006; Beijing, China. New York, NY, USA: IEEE. pp. 5130–5135.
- [24] Nakauchi Y, Noguchi K, Somwong P, Matsubara T, Namatame A. Vivid room: human intention detection and activity support environment for ubiquitous autonomy. In: IEEE 2003 Conference on Intelligent Robots and Systems; 27 October–1 November 2003; Las Vegas, NV, USA. New York, NY, USA: IEEE. pp. 773–778.
- [25] Noguchi K, Somwong P, Matsubara T, Nakauchi Y. Human intention detection and activity support system for ubiquitous autonomy. In: IEEE 2003 International Symposium on Computational Intelligence in Robotics and Automation; 16–20 July 2003; Kobe, Japan. New York, NY, USA: IEEE. pp. 906–911.
- [26] Koo S, Kwon DS. Recognizing human intentional actions from the relative movements between human and robot. In: IEEE 2009 International Symposium on Robot and Human Interactive Communication; 27 September–2 October 2009; Toyama, Japan. New York, NY, USA: IEEE. pp. 939–944.
- [27] Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A metaanalysis of the experimental evidence. *Psychol Bull* 2006; 132: 249–268.
- [28] Meltzoff, AN. Understanding the intentions of others: reenactment of intended acts by 18-month-old children. *Dev Psychol* 1995; 31: 1–16.
- [29] Kiran M, Kin LW, Ali KKH. Clustering techniques for human posture recognition: K-means, FCM and SOM. In: Proceedings of the 9th WSEAS International Conference on Signal, Speech and Image Processing, and 9th WSEAS International Conference on Multimedia, Internet and Video Technologies; 3–5 September 2009; Budapest, Hungary. Stevenson, WI, USA: WSEAS. pp. 63–67.
- [30] Tahir NM, Hussain A, Samad SA, Husain H. Posture recognition using correlation filter classifier. *J Theor Appl Inf Technol* 2008; 4: 767–773.
- [31] Jolliffe IT. *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.
- [32] Lee KK, Xu Y. Modeling human actions from learning. In: IEEE 2004 Conference on Intelligent Robots and Systems; 27 September–2 October 2004; Sendai, Japan. New York, NY, USA: IEEE. pp. 2787–2792.
- [33] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *P IEEE* 1989; 77: 257–286.
- [34] Durdu A, Erkmén I, Erkmén AM. İnsan-robot etkileşimi yoluyla insan niyetlerinin tahmin edilmesi ve değiştirilmesi. In: ELECO 2012 Symposium on Electric, Electronics, and Computer Engineering; 29 November–1 December 2012; Bursa, Turkey (in Turkish).