

# ACTIVE INTENTION INFERENCE FOR ROBOT-HUMAN COLLABORATION

HSIEN-I LIN, XUAN-ANH NGUYEN & WEI-KAI CHEN

Graduate Institute of Automation Technology National Taipei University of Technology, Taipei, Taiwan.

## ABSTRACT

Understanding human intention is an important ability for an intelligent robot to collaborate with a human to accomplish various tasks. During collaboration, a robot with such ability can predict the successive actions that a human partner intends to perform, provide necessary assistance and support, and remind for the missing and failure actions from the human to achieve a desired task purpose. This paper presents a framework that allows a robot to automatically recognize and infer the action intention of a human partner based on visualization, in which an inverse-reinforcement learning (IRL) system is learnt based on the observed human demonstration and used to infer the human successive actions. Compared to other systems based on reinforcement learning, the reward of a Markov-Decision process (MDP) is directly learned from the demonstration. In our experiment, we provide some examples of the proposed framework which yields promising results with coffee-making and pick-and-place tasks. Regarding to the human-intention model based on IRL, the coffee-making experiment indicates that the action is globally predicted because the action of putting down the water pot is selected instead of pouring water when the cup is already filled with water.

*Keywords:* Human gesture recognition; human-robot collaboration; Markov decision process.

## 1 INTRODUCTION AND MOTIVATION

Nowadays along with the rapid development of intelligent robots, robots are expected to become a partner of elders, a friend of children, or a co-worker in manufacturing factories. However, these needs are still different from what the robots are programmed manually as usual. Thus, developing a robot which can interact with human partners is one of the major challenges in recent robotic research. Among a variety of problems in human-robot interaction, the importance of human-intention prediction is clear because if a robot understands human intention, it can provide necessary support and assistance at an appropriate moment. Similarly in our society, we observe each other, ask directly or based on the observations to understand and infer correctly the intentions of the others, and then we can work together to accomplish a task successfully. The above implies that endowing a robot with an ability to understand human intention helps the robot become a real helper to humans.

Many works for message exchange between humans and robots have done so far. In these methods, human action plays a key means to deliver information from humans to robots. For example, Bascetta *et al.* [1] used human intention to secure human safety in a coexistence system. On the other hand, Kanno *et al.* [2] categorized the operational definition of intention recognition into three types: keyhole recognition, intended recognition, and obstructed recognition. Intended recognition for human-robot collaboration (HRC) indicates that a human and a robot are both aware of cooperating in the recognition. Thus, this paper focuses the mode of intended recognition, meaning that the robot attempts to recognize human intention during HRC.

The recognition models of human intention have been developed by genetic algorithm (GA), Graphical model, Hidden Markov model (HMM), and Dynamic Bayesian network (DBN), neural-fuzzy network, ontology. Zhou *et al.* [3] used the genetic and ant algorithms and driving data to predict a driver's intention such as lane changing and car following. Wu *et al.* [4] proposed a graphical model to recognize attack intention which was represented by

a plan or sequence of actions and adopted D-S evidence theory to infer intention by probability. Jin *et al.* [5] used a HMM to predict a driver's intention for assisting lane changing. Gehrig *et al.* [6] used 24 motion units and HMMs to form five complex motion sequences in a real-world kitchen tasks. Tahboub [7] used a DBN inferring user intention to reduce the latency in teleoperating a mobile robot. Jeon *et al.* [8] used ontology-based rules to do user intention inference. In other words, user intentions were described by hierarchical rules and sensor data. Huang *et al.* [9] applied human intention to control a robot walking. The intention was predicted by an adaptive network-based fuzzy inference system (ANFIS) using the force sensor data collected from the handle of the robot. Usually, human explicit actions are implicitly controlled by intention. However, the human-intention models such as HMM, DBN and neural-fuzzy networks lack the control mechanism. On the contrary, A MDP involves the control to infer human intention. In the work of [10] McGhan *et al.* introduced a MDP-based human task-level intent system in the context of a collaborative task. Their system used the human subject experiment data to feed the task model, a robot was hypothesized to predict companion intents by observing and identifying actions as a part of a motion sequence. The approach was based on the fulfillment of the human recurrent targets to find the reward function. In a similar way, our method adopts MDPs to infer human intention in HRC, but we apply a more optimal method to get the reward function.

In a human-intention recognition system, hand gestures are commonly used to predict human intention. For example, Song *et al.* [11] presented a probabilistic graphical method to estimate human intention in grasping tasks. A Kinect camera was used to monitor grasp processes and collect image data. Two major components of this framework are modeling grasping tasks and generative soft vector quantization using self-organizing maps (SOMs) and Gaussian mixture models (GMMs). Similarly, the approach in Ref. [12] employed a Bayesian classifier to recognize human intention based on tracking hands. The results were used in a remote mouse. Alternatively, Ref. [13] introduced continuous dynamic programming (CDP) as a solution for recognizing gesture spotting from human demonstration in video streams. However, most of the previous research mainly considers classifying human hand actions, while human intention should be regarded as an optimization problem of task planning from human demonstration. In this paper, we propose to use IRL [14] to find the optimal reward of a MDP from human demonstration.

The novelty of our research is to develop the human-intention recognition framework in HRC scenarios. Based on the marriage of MDP and IRL, the proposed framework is innovative in human-intention recognition. The MDP complements the probabilistic graphical models with the control power and the IRL provides an optimization process for a MDP from human demonstration. To obtain a correct type of hand gestures, we use a GMM as the robust skin colour model and convolutional neural networks (CNNs) [15] as the hand gesture classifiers. Without loss of generality, the image pre-process *e.g.* poses calibration is implemented before gesture classifiers. For the MDP, how to specify the reward is still a challenging task. However, with the demonstration, the IRL derives the optimal reward function of the MDP and accelerates the learning time for obtaining the MDP policy. All in all, the advantages of our approach include: first, the environmental uncertainty and decision making in the process of human intention are modeled by a MDP. Second, the reward function of a MDP is optimally found from human demonstration and the learning time is significantly reduced.

The rest of this paper is organized as follows: Section 2 briefly describes the problem statement. Section 3 presents the proposed human-intention framework, where the motion layer for recognizing hand gestures and the human-intention model are introduced. Section 4 presents

the experiments that validate the feasibility of the proposed framework. Finally, concluding remarks and the future work are presented in Section 5.

## 2 PROBLEM STATEMENT AND FORMULATION

This paper considers a typical human-robot collaboration scenario as illustrated in Fig. 1 in which a robot and a human execute a daily task together. Some of the work is carried out by the human and the rest of the work is done by the robot. When the human cannot achieve the work by himself, the robot should be capable of detecting human intention for the task and providing appropriate assistance with the approval of the human. Figure 2 shows the collaboration principle with the proposed human-intention system. In this case, the robot is not only a collaborator but also a helper. Consequently, the collaboration performance will also be improved significantly.

As we know, human intention is explicitly regarded as a plan of sequential actions based on a human's desire. In other words, human intention is a process in which a human exploits his knowledge to make decisions about selecting suitable actions at different moments in order to accomplish a task. Thus, the goal of this paper is that a robot has to correctly recognize human intention with some uncertain situations during collaboration. Accordingly, a human-intention model that can tolerate noises and infer a reasonable plan of actions is the solution to this problem. Since a MDP has been known as a great probabilistic framework for

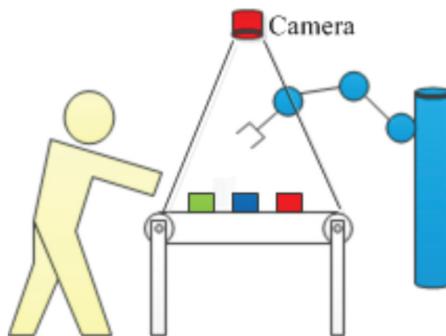


Figure 1: Human-robot collaboration scenario.

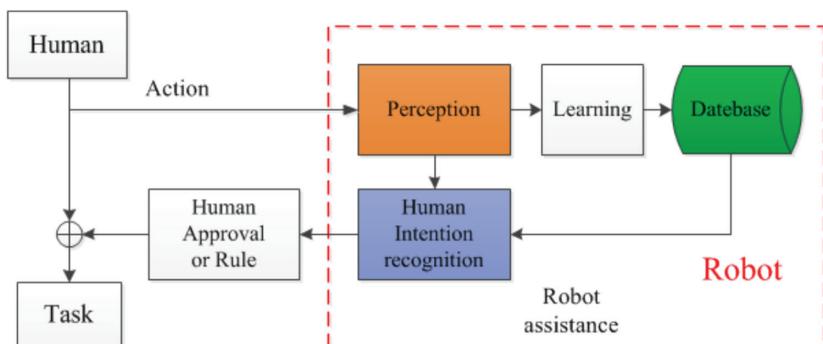


Figure 2: Collaboration principle in the proposed human-intention system.

decision-making tasks, we adopt a MDP to model human intention and then recognizing human intention is to find the policy of the MDP. A MDP  $(S, A, P, R, \gamma)$  consists of a set of states  $S$ , a set of actions  $A$ , a transitional probability  $P$ , and the reward function  $R$  that gives a reward for transition from the current state  $s$  to the next state  $s_0$  using action  $a$ . A discount factor  $\gamma \in [0;1]$  to avoid an infinite sum over a potentially infinite horizon. Most importantly, the reward function influences the policy of the MDP significantly but it is manually defined for a task. In this paper, we aim at exploring the reward function by human demonstration. Thus, we propose IRL as a suitable approach because it considers the entire task from scratch to the final step, in order to find a globally optimal policy. That is a unique and most favorable policy. For the proposed system, a human first demonstrates how to manipulate with objects in a certain order to achieve an intended task purpose, and then a robot learns the human intention to provide appropriate assistance.

### 3 THE HUMAN-INTENTION LEARNING SYSTEM

Figure 3 shows the flowchart of our framework, which consists of two phases: human gesture recognition and human-intention model. In the first phase, when a human and a robot start to collaborate, the captured images of the human gestures and manipulative objects are fed to the motion and object layers. The motion layer is in charge of recognizing human gestures by CNNs and the object layer provides the object attributes such as position and colour. In the second phase, the results from the first phase are encoded as the states in a MDP that serves as a human-intention model. The state and action spaces of the MDP are defined by users. Finally, the IRL is then applied to find the optimal reward function for the MDP using a collection of human demonstrations.

#### 3.1 Hand gesture recognition

In the learning phase, the hand images of human demonstrations for a specific task are collected to train a hand gesture recognition system. For instance, five gesture types for a coffee-making task are shown in Fig. 4. During the execution phase, human actions will be

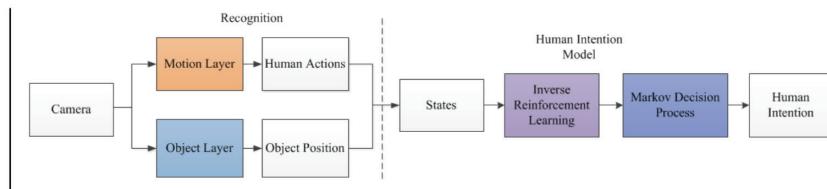


Figure 3: Flowchart of the proposed human-intention learning system.



Figure 4: Five gesture types to be recognized.

captured and recognized by the classifier above. In order to overcome the changing lighting condition, we adopt a GMM and the YCbCr color space to as the skin model for the images. The details of the implementation were presented in [16]. Using the skin model, the images are binarized into skin and non-skin areas. After image binarization, the images are filtered to remove unexpected noises and sharpen [16].

In addition, since human actions are continuous, the position and orientation of the hand are progressively changing, which makes it difficult to correctly recognize gestures. Thus, a pose calibration is applied to the binarized images to obtain better recognition results. The center position and principal axis of the hand are found by the gravity calculation method and then used to translate and rotate the image to the neutral pose. The pose calibration produces significantly better recognition results. Figure 5 (a) and (b) show an example of original and calibrated images respectively.

We now turn to the problem of recognizing hand gesture types. Recently, convolutional neural networks have been known as a successful tool of multi-class classification. This success is largely attributed to the use of local filtering, max pooling and weight sharing in the CNN architecture, which allows us to save memory size and improve efficiency. Additionally, a CNN uses less pre-processing than conventional neural networks, linear regression and support vector machines (SVMs). From the above mentioned advantages, we adopt CNNs as the approach to classify hand gesture types. However, our CNN is slightly modified to get a better result than the traditional one [17]. Let take coffee-making task as an example, we consider five gesture types, so our CNN output layer here has five units.

Each CNN for a gesture type has nine layers, denoted as I1, C2, S3, C4, S5, C6, S7, H8 and O9 in sequence. In particular, I1 is the input layer, which consists of  $50 \times 50$  pixel calibrated images. C2, C4 and C6 are convolutional layers with  $3 \times 3$ ,  $5 \times 5$  and  $5 \times 5$  masks, respectively. The masks in C2 layer are designed for sharpen, edge detection, horizontal edge detection, vertical edge detection, horizontal gradient detection, vertical gradient detection, horizontal sobel detection and vertical sobel detection. S3, S5 and S7 are sub-sampling layers. The feature maps in S3 and S5 layers are taken by  $2 \times 2$  pixel sub-sampling but  $3 \times 3$  in layer S7. The outputs of S7 are expanded and concatenated in terms of a  $108 \times 1$  vector. Finally, the hidden layer of the neural network H8 is also fully connected to the five neurons of the output layer O9. Stochastic back-propagation is used to update weights in the CNN. This modified CNN architecture demonstrates good results. Apparently, the modified CNN outperforms the traditional one. For continuous hand motions, transitive hand gestures are also taken into consideration



(a)



(b)

Figure 5: Pose calibration.

the proposed system. The transitive gestures are assigned to the gesture type in the closest preceding or successive frame [16].

### 3.2 Human-intention model

The human-intention model is very useful to a robot to understand the hidden information behind human motions. The hidden information could be a policy of actions which is a realization of the mental process. To find such a policy for our collaborative tasks, we adopt a MDP as a human-intention model and IRL as the method to find the optimal reward function of the policy.

After the previous stage of hand gesture recognition, the classified hand gestures and the object attributes are encoded as the states of the MDP represented as a 5-tuple  $\{S, A, P, R, \gamma\}$ . In our framework, let  $S_0$  be the set of possible attributes of the object layer, we consider object features like colour, position, status, etc, while  $S_m$  be the set of gesture status of the motion layer, for instance gestures are with or without object.  $S$  is the set of all possible states in the collaborative task. Consequently,  $|S|$  of the MDP is  $|S_0| \times |S_m|$ . We also define  $A = \{a_1, a_2, \dots, a_n\}$  as a set of all possible actions to accomplish desired tasks where  $n$  is the number of necessary actions. The action set is designed to ensure that the desired task can be completed by it and the set size should be kept to a minimum. In addition, the discount factor  $\gamma \in [0, 1]$  is selected during the optimization process and the square matrix  $P(s, a, s')$  is the transition probability matrix that is used to describe all the probabilities over all the state transitions. In order to obtain the transition matrix  $P$ , we implement each task 100 times, represent each task by states  $S$ , and analyze the transitions among them. A policy  $\pi : S \rightarrow A$  is evaluated by a value function (so-called Bellman equation). The Bellman equation is expressed as follows:

$$V^\pi = \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma V^\pi(s')] \quad (1)$$

For each policy  $\pi$ , a corresponding value  $V_\pi$  from Equation (1) is found. Furthermore, the goal of the learning agent is to find the optimal policy  $\pi^*$ , meaning that  $V_\pi$  should be the maximum value  $V^*$ :

$$V^* = \max_a \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma V^*(s')] \quad (2)$$

In Equation (1), the reward  $R$  plays a crucial role in finding the optimal policy  $\pi^*$ . In other words, the optimal criterion depends on the definition of the reward function [14]. In fact, the reward function is usually handcrafted, which is non-convincing and can lead us to a non-optimal policy. For example, in our previous research, we have applied the frequency-based method to design the reward function [18], but in some cases, it is not optimal according to the demonstration data. In this paper, we attempt to find a convincing and efficient approach to obtain the reward function. Thus, we apply IRL to the human-intention model to derive the reward function directly by the demonstration data. The recognized states and actions of the task by the hand gesture recognition system are used to compute the reward function. The procedure of IRL attempts to maximize the margin from the optimal value function to others and is written as:

$$P_{a^*} V^\pi \geq P_a V^\pi \quad \forall a \in A \setminus a^* \quad (3)$$

where:  $a^*$  is the demonstrated action, and  $P_{a^*}$  and  $P_a$  are the transition matrices of action  $a^*$  and  $a$ , respectively. The details of the derivations are referred to [14]. In particular, the reward function  $R$  can be calculated as a linear programming problem, defined as:

$$\begin{aligned} & \max \sum_{i=1}^N \min_{a \in A \setminus a^*} \left( (P_{a^*} - P_a)(I - \gamma P_{a^*})^{-1} R \right) - \lambda \|R\|_1 \\ & \text{s.t. } (P_{a^*} - P_a)(I - \gamma P_{a^*})^{-1} R \geq 0 \quad \forall a \in A \setminus a^* \\ & |R_i| \leq R_{\max}, i = 1, \dots, N \end{aligned} \quad (4)$$

In our tasks,  $\lambda$  is set to 0.65. Usually, robots are pre-programmed by default policies. To meet users' requirements, the policies need to be updated over time. However, in some cases such that gesture recognition error and insufficient amount of demonstration data, it is ineffective to update the human-intention model using demonstration data. As a consequence, in the experiment, when the number of the demonstrated trials is larger than a threshold, the IRL is launched to update the model. In the experiment section, we discuss more about the superiority of IRL with a specific task.

#### 4 EXPERIMENTAL WORK

The feasibility and the efficiency of the proposed learning framework were experimentally evaluated with a coffee-making task and a pick-and-place task. The robot work is to monitor the task process, predict human intention and remind human what action s/he should take by display a text on a screen. In the coffee-making task, the experiments demonstrated that the different reward functions *e.g.* frequency-based and IRL affect the human-intention model. The frequency-based method is a local approach which weights  $R_{ss}^a$  of the Bellman equation by the times of the demonstration. However, the IRL obtains  $R_{ss}^a$  in a sense of globally optimizing the Bellman equation. In the pick-and-place task, the experiment shows that the sequence of human actions was correctly predicted based on the predefined plan learned by the presented learning framework. When the blue, green, and red cubes were randomly grasped and placed, the system suggested a successive action to go back to the learned plan. By doing so, a robot could help the human continue the task with the human's approval.

##### 4.1 Coffee-making task

In the first task, the workspace is illustrated in Fig. 6, which consists of an inside and an outside area. A camera was mounted right above the work-space to monitor the environment. A pot, spoon, cup, and coffee can were placed initially in the outside area, while a human will manipulate objects in inside area. In the experiment, we implemented three different tasks to learn the reward functions. In task 1, we took a pot and pour water in a cup; in task 2 we took a spoon and spooned up coffee in the cup. Finally, task 3 was performed and illustrated in Fig. 8. Table 3 shows the optimal policies that we obtained from the experiments.

For these tasks, the actions were referred to the five classified human gestures in Fig. 4. Table 1 shows the action meaning of each hand gesture, while Table 2 shows the comparison of the recognition rates between the modified and traditional CNN architecture. Figure 7 depicts the cup attributes in the object layer. For example, when the cup was placed in the outside area, it meant that the work had not been started and there was only one 'empty' state for the cup. Similarly, when the cup was located in the inside area, there were three possible

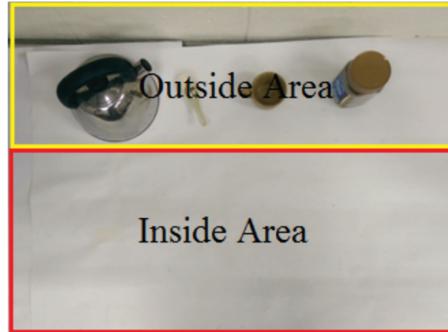


Figure 6: Workspace of the pick-and-place task.

Table 1: Action set in the coffee-making task.

	Gesture	Action
1	Empty	Grasp pot, stop
2	Grip	Place cup
3	GripTrans	Place pot, pour water
4	Spoon	Spoon up and down coffee powder
5	Hold	Hold cup

Table 2: Prediction results of actions performed.

Total (14 Testers)	Empty	Grip	Grip trans	Spoon	Spoon trans	Hold	Average
Modified CNN - Success	1400	1395	1354	1315	1351	1354	
Modified CNN - Rate (%)	100	99.64	96.71	93.93	96.50	96.71	97.24
CNN - Success	1395	1385	1341	1049	1246	1308	
CNN - Rate (%)	99.64	98.93	95.97	74.93	89.00	93.43	91.95

states (empty, with liquid, and with coffee powder). Statuses from the motion layer and the object attributes from the object layer were composed of the set of states  $S$  in the MDP of the human-intention model. Here, we have 10 states as illustrated in the set of actions in Fig. 8. Table 1 and object attributes were presented in a state-diagram to reflect the human intention for the task.

Figures 8 (a)–(g) depict the steps of the human demonstration for the coffee-making task 3. The demonstration was used to explore the optimal reward function of the MDP. Figure 9 shows the state diagram of the human-intention model based on the frequency-based method and IRL. It is clear to see that there were two differences between the two approaches at states 4 and 8. When the process was at state 7 with the spoon in the inside area and the empty cup, the action ‘spoon up coffee’ was selected by both approaches to go to state 8. However, at state 8, we had two different results. The frequency-based approach still stuck at the same

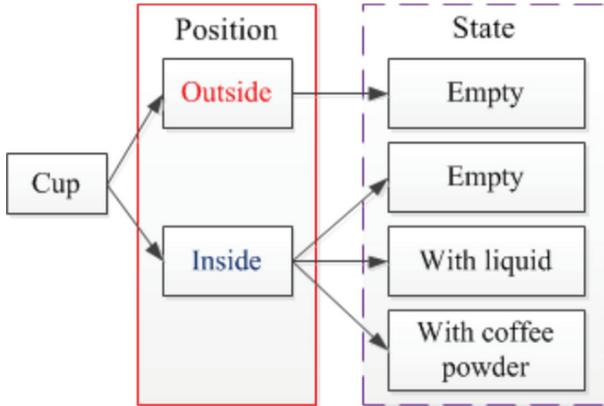


Figure 7: Cup attributes in the object layer.

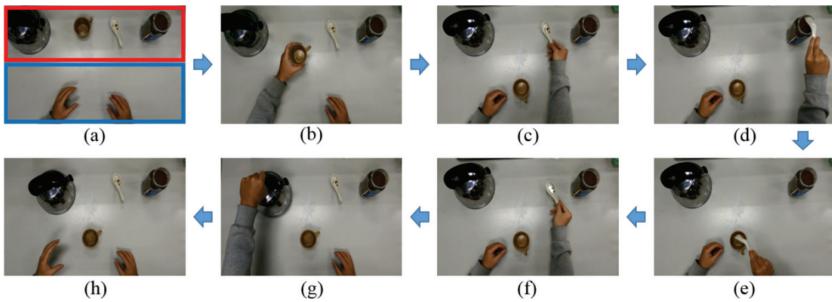


Figure 8: Human demonstration for the coffee-making task. (a) The human and objects were at initial status; (b) grasping and placing the cup; (c) grasping the spoon; (d) spooning up the coffee powder; (e) putting back the spoon; (f) grasping the pot and pouring water into the cup; (g) putting back the pot.

Table 3: Optimal policies.

Task 1		Task 2		Task 3	
Frequency	IRL	Frequency	IRL	Frequency	IRL
State 1	Grasp cup	Grasp cup	Grasp cup	Grasp cup	Grasp cup
State 2	Place cup	Place cup	Place cup	Place cup	Place cup
State 3	Grasp pot	Grasp pot	Grasp spoon	Grasp spoon	Grasp spoon
State 4	Pour water	Pour water	Pour water	Pour water	Place pot
State 5	Place pot	Place pot	Place pot	Pour water	Place pot
State 6	Stop	Stop	Stop	Stop	Stop
State 7	Place spoon	Place spoon	Spoon up coffee	Spoon up coffee	Spoon up coffee
State 8	Place spoon	Place spoon	Place spoon	Spoon up coffee	Place spoon
State 9	Grasp pot	Grasp pot	Grasp pot	Grasp pot	Grasp pot
State 10	Pour water	Pour water	Pour water	Pour water	Pour water

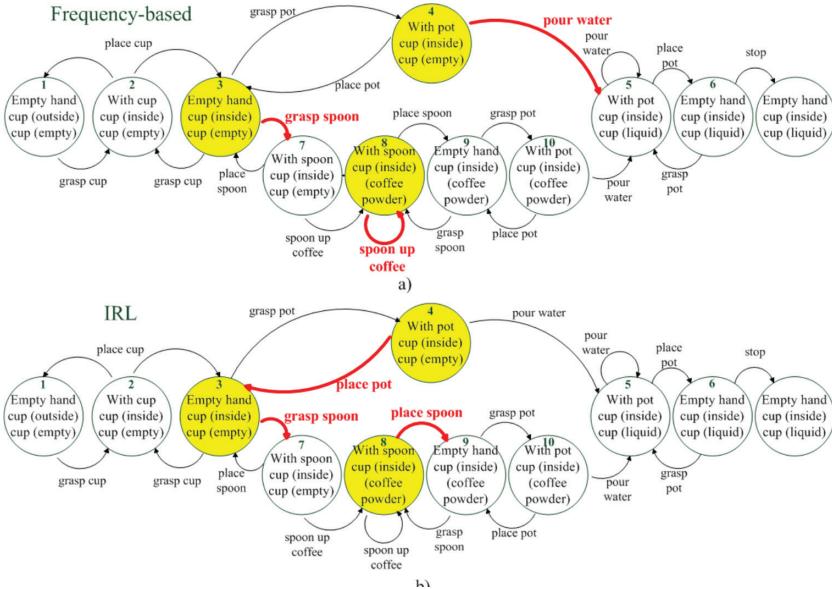


Figure 9: The reproduction results for the two methods. a) The result of frequency-based method. b) The optimization of the reward functions through IRL method.

state by the action ‘spoon up coffee’ while the IRL approach chose the action ‘place spoon’ to state 9. Even though the state 8 shows that the coffee powder was in the cup, the frequency-based approach kept doing the action ‘spoon up coffee’. Because of such an imperfection of human demonstration at state 8 for a few times, the action ‘spoon up coffee’ was repeated. However, the IRL approach took an optimal action ‘place spoon’ to accomplish the task.

Furthermore, another difference between two methods also appeared at state 4 when the action ‘grasp pot’ was chosen at state 3. Actually, in state 4, when the human already hold the pot, the reward function that we got from the frequency-based method suggested to perform the action ‘pour water’. However, before the action ‘spoon up coffee’ was performed, the action ‘pour water’ was unnecessary. The process based on the frequency-based approach tended to finish the task without coffee. On the contrary, the IRL approach chose the action ‘place pot’ before the coffee powder had not been put in the cup.

Evidently, in the coffee-making task, the frequency-based reward function provided a locally optimized solution because it regulated the reward based on the count of the excited action at each state. Alternatively, the IRL explored the optimal reward function from the demonstration and derived the action at each state to achieve the most advantageous performance for task accomplishment.

#### 4.2 Learning a pick-and-place task with a human

This task involved a pick-and-place task and illustrated how the robot can help the human. The goal was to stack three different colour cubes in the same order executed in the human demonstration. When the human stacked the cubes, the robot recognized the human actions and predicted the most possible successive action the human should do. In the task, the set of states was constructed from the object layer. The features including position, colour, and

Table 4: Action set in the pick-and-place task.

	Action	Description
1	Transport empty	Transport without object
2	Transport loaded	Transport with an object
3	Grasp	Grasp an object
4	Release	Release an object
5	Stop	Stop

order of the three cubes were extracted to create the set of states  $S_0$ . There were 16 states. In the further work, more actions could be considered to make the collaboration versatile. Table 4 shows all actions that were used in the task. The reward function and the optimal policy were explored as mentioned in the section 3. In this experimental scenario, to evaluate the obtained policy, some wrong action sequences was purposely performed by the human and we would like to check whether the human-intention model could predict the successive action that made the task being executed in the same action sequence of the demonstration.

The task was demonstrated as the following sequence: grasped the blue cube, placed the blue one in the inside area, grasped the green one on the blue one, pick the red one, and place it on the green one (*blue → red → green*). The robot watched the human operations and successfully predicted the next actions to implement the task. Figure 10 shows that when the human intentionally executed the wrong task sequence, the robot was able to detect it and suggest the human the correct one (the command in yellow). Figure 10a shows that at the beginning of the task, the robot suggested the human the blue cube. In Fig. 10b, when the human took the cube correctly, the robot continued to suggest the green one. Figure 10c shows that when the human made a mistake to take the red one instead of the green one, the robot suggested the human should remove the red cube. Figure 10d shows that when the human removed the red one, the robot suggested the green one again. Figure 10e shows that when the human took the green one, the red one was suggested in the next step; and finally the human took the red cube to finish the task as shown in the Fig. 10f. The details of the task are shown in the attached video.

The experimental work was the initial attempt to apply MDP and IRL to the problem of modeling collaborative tasks. While the robot was able to recognize human intention to

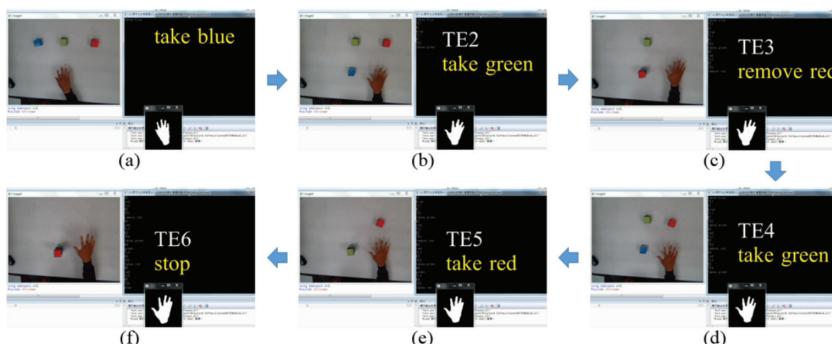


Figure 10: Predicted actions of the pick-and-place task by the human-intention model. (a)–(f) were the steps of the task.

perform the task, it could predict successive actions and suggest them to the human. The following task is about to propose a collaborative model between robot and human according to the obtained human intention. More experiments will be performed on a robot.

## 5 CONCLUSIONS AND FUTURE WORK

We have introduced a human-intention learning system for HRC. To achieve the learning from human demonstration, we leveraged superiority from the marriage between MDPs and IRL. The MDPs are used to explore decision-making rules and deal with uncertainties from human demonstration and environment. The IRL is used to obtain the optimal reward function based on demonstration and provide a global solution for a given task. Compared to the previous work, our framework involves control actions to infer human intention and provides suggested actions for task accomplishment. Our extensive experimental evaluation with the coffee-making and pick-and-place tasks indicated that the presented human-intention learning system is not only capable of performing collaborative task but also obtaining a globally optimal policy.

## ACKNOWLEDGEMENTS

This work was supported in part by the National Science Council under Grant NSC MOST 104-2221-E-027-077. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Council.

## REFERENCES

- [1] Bascetta, L., Ferretti, G., Rocco, P. & Ardo, H., *Towards safe human-robot interaction in robotic cells: An approach based on visual tracking and intention estimation*. IEEE/RSJ IROS, San Francisco, pp. 2971–2978, 2011.  
<https://doi.org/10.1109/iros.2011.6094642>
- [2] Kanno, T., Nakata, K. & Furuta, K., A method for team intention inference. *International Journal of Human-Computer Studies*, **58**(4), 393–413, 2003.  
[https://doi.org/10.1016/s1071-5819\(03\)00011-9](https://doi.org/10.1016/s1071-5819(03)00011-9)
- [3] Zhou, S. & Wu, C. H., A recognition method for drivers intention based on genetic algorithm and ant colony optimization. Conference on Natural Computation, Shanghai, pp. 1033–1037, 2011.  
<https://doi.org/10.1109/icnc.2011.6022185>
- [4] Wu, P., Wang, Z. & Chen, J. H., Research on attack intention recognition based on graphical model. *Proceedings Of International Conference on Information Assurance and Security*, Shanghai, **1**, pp. 360–363, 2009.  
<https://doi.org/10.1109/ias.2009.158>
- [5] Jin, L., Hou, H. & Jiang, Y., Driver intention recognition based on continuous hidden Markov model. *Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*, **1**, pp. 739–742, 2011.  
<https://doi.org/10.1109/tmee.2011.6199308>
- [6] Gehrig, D., Khne, H., Wrner, A. & Schultz, T., *HMM-based human motion recognition with optical flow data*. Proceedings of the 9th IEEE-RAS International Conference on Humanoid Robots, Paris, pp. 425–430, 2009.
- [7] Tahboub, K. A., Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition. *Journal of Intelligent and Robotic Systems*, **45**(1), 31–52, 2006.

- [8] Jeon, H., Kim, T. & Choi, J., *Ontology-based user intention recognition for proactive planning of intelligent robot behavior*. International Conference On Multimedia and Ubiquitous Engineering, Busan, pp. 244–248, 2008.
- [9] Huang, Y.C., Young, H.P., Ko, C. H. & Young, K.Y., Design and implementation of a robot control system with traded and shared control capability. *Proceeding Of Asian Control Conference*, pp. 311–316, 2011.
- [10] McGhan, C.L.R., Nasir, A. & Atkins, E.M., Human intention prediction using Markov Decision Processes. *Journal of Aerospace Information Systems*, **12**(5), pp. 393–397, 2015.  
<https://doi.org/10.2514/1.i010090>
- [11] Song, D., Kyriazis, N., Oikonomidis, I., Papazov, C., Argyros, A., Burschka, A. & Kragic, D., *Predicting human intention in visual observations of hand/object interactions*. IEEE International Conference Robotics and Automation, pp. 1608–1615, 2013.  
<https://doi.org/10.1109/icra.2013.6630785>
- [12] Keskinpala, H.K., Adams, J.A. & Kawamura, K., *PDA-based human robotic interface*. IEEE International Conference on SMC, Washington, DC, pp. 1310–1315, 2003.  
<https://doi.org/10.1109/icsmc.2003.1244502>
- [13] Breazeal, C., Emotion and social humanoid robots. *International Journal of Human-Computer Studies*, **59**, pp. 119–155, 2003.  
[https://doi.org/10.1016/s1071-5819\(03\)00018-1](https://doi.org/10.1016/s1071-5819(03)00018-1)
- [14] Ng, A. Y. & Russell, S., *Algorithms for inverse reinforcement learning*. Proceedings of ICML, pp. 663–670, 2000.
- [15] Nagi, J., Ducatelle, F., Caro, G., Ciresan, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J. & Gambardella, L.M., *Max-pooling convolutional neural networks for vision-based hand gesture recognition*. IEEE International Conference on signal and Image Processing Applications, pp. 342–347, 2011.
- [16] Lin, H.I. & Chiang, Y.P., Understanding human hand gestures for learning robot pick-and-place tasks. *International Journal of Advanced Robotic Systems*, **12**(5), p. 49, 2015.  
<https://doi.org/10.5772/60093>
- [17] Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P., Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**, pp. 2278–2324, 1998.  
<https://doi.org/10.1109/5.726791>
- [18] Lin, H.I. & Chen, W-K., *Human intention recognition using markov decision processes*. Proceedings of International Conference on Automatic Control (CACS), pp. 240–343, 2014.  
<https://doi.org/10.1109/cacs.2014.7097213>