Human-Robot Collaboration by Intention recognition using Probabilistic State Machines

Muhammad AWAIS and Dominik HENRICH
Lehrstuhl für Angewandte Informatik III (Robotik und Eingebettete Systeme)
Universität Bayreuth, D-95440 Bayreuth, Germany
{Muhammad.Awais | Dominik.Henrich}@uni-bayreuth.de

***Abstract*—Combining the intelligent and situation dependent decision making capabilities of a human with the accuracy and power of a robot, performance of many tasks can be improved. The human-robot collaboration scenarios are increasing. Human-robot interaction is not only restricted to the humanoid robots interacting with the humans or to the mobile service robots providing different services but also industrial robots opens a wide range of human-robot collaboration set-ups. Intention recognition plays a key role in intuitive human-robot collaboration. In this paper we present a novel approach for recognizing the human intention using weighted probabilistic state machines. We categorize the recognition task into two categories namely *explicit* and *implicit* intention communication. We present a general intention recognition approach that can be applied to any human-robot cooperation situation. The algorithm is tested with an industrial robotic arm.**

***Index Terms*— Intuitive Human-Robot Collaboration, Intention Recognition, Probabilistic State Machines.**

I. INTRODUCTION

With the era of modern technologies, machines are becoming necessary part of human life. The goal is to provide the services to the humans. The provided services are appreciated if they are offered at the right time and need little input effort. Interaction characteristics make a machine more or less acceptable among the humans. The interface between the human and the robot describes the interacting capabilities of a robot, i.e., how much the robot is adaptive towards the behavior of the interacting human. If the interacting human needs to know prerequisites in order to interact with the robot then the level of interaction is less acceptable as compared to the one that does not demand any prerequisite for interaction. Therefore, a robot needs to intuitively recognize the intention of an interacting human. Recognizing the human intention, the robot can smoothly cooperate with the human. There are many working scenarios where the intelligence of a human and the efficiency of a robot can be combined to provide a better output.

Intention recognition has been performed using different probabilistic frame works and other approaches, e.g., Hidden Markov Model (HMM) [5] and Dynamic Bayesian Network (DBN) [3]. Hybrid DBN proposed by [4] describes the usage of discrete and continuous random variables in DBN. Other approaches involve the utility-based intention recognition [7], ontology based intention recognition [2] and graph based intention recognition [6]. Intention recognition proposed by [1] describes the use of proactive action selection if the recognized intention is ambiguous. Though the above mentioned intention recognition techniques differ but mostly the interacting agents actively engage each other. The actions performed by the agents are given more importance as compared to the scene information that also contributes to the intention recognition.

In this paper, the presented intention recognition algorithm uses probabilistic state machines. The paper focuses on the recognition of intention in two cases, i.e., the human is actively collaborating with the robot and the human is not collaborating but the robot decides to start proactive collaboration depending upon the scene information and the human actions. The input of the algorithm is the camera images of the scene that include the known objects and the human actions along with the list of human intentions. The output of the algorithm is the recognized human intention.

Human intentions can be categorized into two categories namely explicitly and implicitly communicated intentions. *Explicitly* communicated intentions are the intentions that are communicated by the human using different actions in different orders. The same type and number of actions in different organization may represent different explicitly communicated intentions. The main goal of human intention recognition is that the situation dependent human response and intelligence of human can be combined with the efficiency and power of the robot. The recognition of explicitly communicated intention in the industrial scenario showing the human-robot cooperation is illustrated by the following example. A human is examining the glass sheets (big building glass, windscreens) passing on a conveyer belt. It is difficult to find a defect like scratch or damaged corner by a vision system but a human can do it quite easily and picking and placing of heavy objects is easier for a robot than a human. In this situation, a human and a robot can cooperate to compensate each other's short comings.

Implicitly communicated intentions are those that are not directly communicated to the robot by human actions corresponding to the intention. The actions performed by the human representing the implicitly communicated intention do not directly engage the robot. The human seems to work on its own. The robot needs to know the probabilities of different

978-1-4244-6884-3/10/$26.00 ©2010 IEEE

human intentions by considering the human actions and the scene information. With the scene information involving the object on which the human is working and the human actions, the robot can estimate the human intention. The recognition of implicitly communicated intention helps the robot to collaborate intuitively with the human. Implicit intention communication can be explained with a general example, i.e., the robot is handing over the appropriate objects to the worker required by him to solve the problem. Concretely, the robot cooperating on an auto assembly line moves towards the vehicle to a specific location where the part carried by the robot is to be assembled according to the human intention.

The remainder of the paper is organized as follows. The intention recognition using the probabilistic state machines is described in Section II. Section III describes briefly the implementation, the results and the experiments. Section IV concludes the paper.

II. Probabilistic State Machines

It is fairly difficult to come up with a straight forward mathematical state prediction model that can predict the next state of human, i.e., next posture of the human body or part of the human body while performing a task, e.g., if the human approaches towards a table then it can not be mathematically predicted that he will pick the keys, the pen, a file, the glass, etc on the table. These are all hypotheses. If we consider these hypotheses as complete action sequences for performing different possible tasks then these sequences can be represented by different models that will represent different intentions of the human.

The action sequences considered as string will not be robust due to intolerant string matching, e.g., if *ABCD* is the target string and the experienced string is *ABCDE* then result will be negative. The *E* may be due to false recognition or unintentionally performed action.

If all the action sequences are considered as a composite state machine then the state transition will become very much complex. The composite state machine may require multiple start and end states due to distinct starting and ending action sequences. A state transition problem may occur if the human changes its action sequence (intention) without completing it, e.g., if the human performs actions *A* and *B* corresponding to an intention but switches its intention and performs an action *E*. If there is no state transition at the state (reached after action *B*) corresponding to action *E* then no transition will occur. Thus the changed human action sequence will not be recognized. There may be an action sequence that is completely irrelevant from other action sequences that will result in unconnected states in the composite state machine. Therefore each action sequence corresponding to a human intention is modelled by a distinct state machine. Different state machines are designed regarding different human intentions. Each state machine represents the flow of different human actions one after another.

A general state machine is shown in the Fig. 1. The formal description of the state machine is also given afterwards.

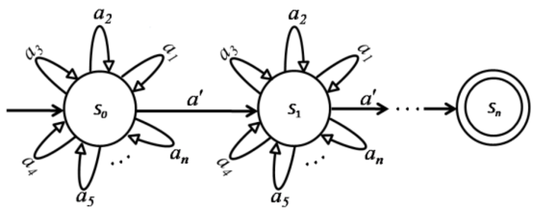


Figure 1. State machine with transitions.

$$Q = \{S_0, S_1, S_2, \ldots, S_n\}$$
$$\Sigma = \{a_1, a_2, a_3, \ldots a_n\}$$
$$\sum_{i=1}^{n} a_i + a' = 1$$
$$a' > a_i \mid a_i, a' \in \Sigma$$
$$q_0 = S_0$$
$$F = \{S_n\}$$

The set $\sum$ represents different actions at a state. Each action a_i represents a value at the state. The value corresponds to the likelihood of the a_i at that state. The action represented by a' at a state has the highest likelihood for that state and the state transition occurs if a' occurs at that state as shown in Fig. 1. The sum of all the values of all the actions at a state adds up to 1. δ is defined by the following state transition table.

TABLE I.
Transition Matrix of a State Machine

Current State	Next State			
	State$_{x-1}$	State$_x$	State$_{x+1}$	State$_{x+2}$
State$_{x-1}$	a_i	a'	NULL	NULL
State$_x$	NULL	a_i	a'	NULL
State$_{x+1}$	NULL	NULL	a_i	a'

The state transitions given in Table 1 describe that the state transition occurs from current state to the next state if a' occurs and to itself if any other action a_i occurs. There exists no transition from the current state to the previous state and to the non-immediate next states. Moreover each state machine carries a weight that describes how closely the state machine represents the actual human intention.

The state machines are termed as probabilistic state machine due to the weight assigned to them. The weight is a probabilistic value and it is updated at every new observation. The state machines can represent both kinds of human intentions, i.e., explicitly and implicitly communicated human intentions. The general flow of the algorithm for probabilistic intention recognition using state machines is shown in Fig. 2.

The following subsections describe the structure of the state machines and the intention recognition algorithm, given in the Fig. 2.

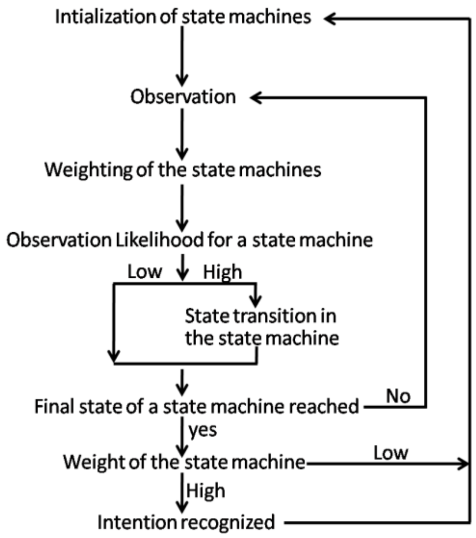
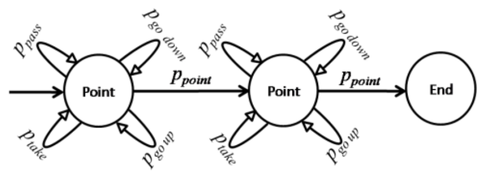


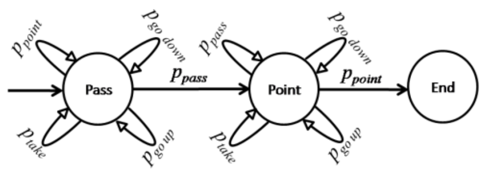
Figure 2. Flow of intention recognition algorithm.

A. Recognition of explicitly communicated intentions

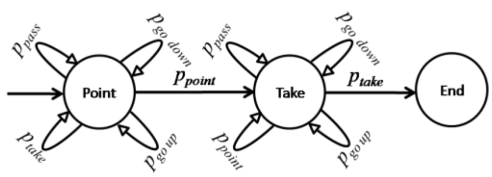
The different state transitions for a state in the state machines correspond to the different human actions. The sequence of the states in the intention state machine represents a unique human intention. Different state transitions have different values for a state in the state machine. The state transition that has high likelihood / not high likelihood / low likelihood for a state will have high / not high / low value for that state, e.g., the start state of state machine given below representing the *pickandplace* intention. The pointing action p_{point} has high value as compared to the open hand action p_{take} for taking an object, object in hand action p_{pass} for giving an object. The different state machines representing different explicitly communicated human intentions are shown in the Fig. 3.



The robot picks and place the object at intended place.



The robot takes and place an object at inteded place.



The robot picks and gives the intended object.

Figure 3. State machines for explicitly communicated human intentions.

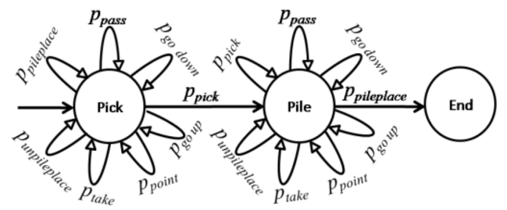
B. Recognition of implicitly communicated intentions

The states in the state machines representing the implicitly communicated human intentions correspond to the specific scene information and the human actions. The sequence of the states in an intention state machine represents specific changes in the scene along with the specific human actions. Different human actions and scene information have different values for a state in the state machine. Human actions and scene information correspond to the state transitions.

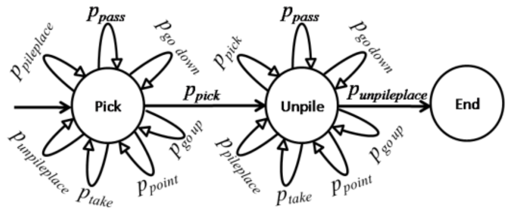
The state machines for implicitly communicated human intention are explained by an example scenario. There exist multiple known objects scattered in common working area. The human picks an object and places that object on another object in the observable area. The robot observes that the number of unpiled objects changes along with the human action of picking and placing of the object.

The state machine uses both the scene information and the human actions, as shown in Fig. 4.

The p_{pick} corresponds to the pick human action. The p_{pick} also corresponds to the no of objects in the scene that is picked up. The $p_{pileplace}$ corresponds to the place human action. This also corresponds to the decrease in the no of objects in the scene as the objects are piled. Similarly the $p_{unplieplace}$ corresponds to the place human action with the increase in the objects in the scene.



State machine for piling intention.



State machine for unpiling intention.

Figure 4. Pileup and unpile intention machines.

Two state machines are used to recognize the implicitly communicated intentions of *pileup* and *unpile*. The likelihood of pick action is the same for both of the state machines, i.e., *pileup* state machine and *unpile* sate machine. The scene information with increased number of unpiled objects has high likelihood for the *unpile* state of *unpile* state machine. Similarly the scene information with decreased numbers of unpiled objects has high likelihood for the *pile* state of *pileup* state machine.

C. Intention recognition algorithm

At the beginning each sate machine representing a unique explicitly / implicitly communicated human intention has the same weight, i.e., the probability of human intentions represented by the state machines are equal. An observation is made and the human actions along with the interesting scene information are extracted. The weights of the state machines

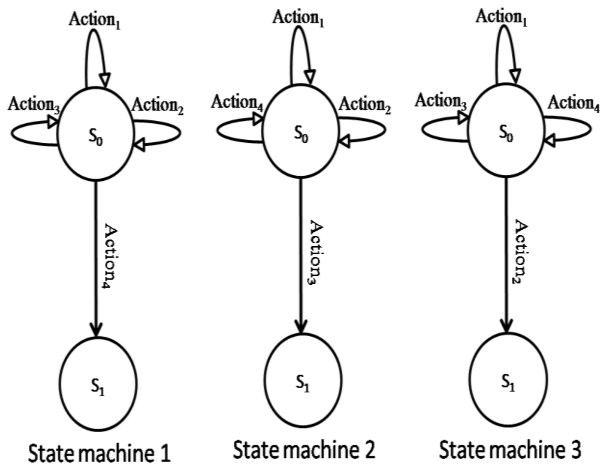


Figure 5. If $Action_4$ occurs then state transition will only occur in the state machine 1.

are updated and normalized so that they add up to 1. The probability of a state machine is directly related to the observation. The state machine for which the observation is most probable gets high weight as compared to the other state machines. If an observation is equally probable for some state machines then those state machines get the same normalized weight. After each observation, along with weight update the important data values necessary for human-robot cooperation are also determined, e.g., calculating the pointed object to be picked or to calculate the pointed place to place the object.

After an observation, state transition occurs in zero, one or more state machines. If there is no human action is observed or an irrelevant human action is observed then no state transition occur in any state machine. If a relevant human action is observed then it is checked for the currently active states of all the state machines. If the observation has the highest probability for the currently active state then the state transition will occur in that state machine. If the observation is highly probable for more than one state machine (currently active state) then the state transition will occur in more than one state machine. In other state machines no state transition will occur. It is explained by the figures given in Fig. 5.

The advantage of making transition in only most probable state machines is that if the human changes his intention in between then this situation can be easily handled, e.g., if the human has an intention and performs an action then the concerned state machine (intention) gets high weight and only in that state machine a transition occurs. If the human changes his intention then the new action sequence can be evaluated with the related state machine and the changed intention can be easily recognized.

The disadvantage may be if the sequence of actions performed concerns to an intention I_1 and before completing the sequence the human changes his intention to I_2. If an action is performed in the action sequence of I_2 that is required for I_1 then the false intention will only be recognized if I_1 reaches its final state and has the high weight.

If the end state of a state machine is reached and the state machine has highest weight then that intention is recognized and state machines are reinitialized. If the end state is reached but the weight is not the highest then the reinitialization is performed without intention recognition.

The defined intention recognition algorithm is given in the Fig. 6. As described earlier that the state machines work as the human intention hypotheses. This algorithm updates the intention hypotheses using the current observation.

At step 1 and 2 the state machines are initialized once with the equal weights. S^t represents the set at time t. It contains the pairs of state machine and weight. The weight of the state machine represents how closely the hypothesis represents the actual human intention. Step 5 describes how the weights of state machines are updated according to the observation probabilities. The observation probabilities correspond to the likelihood of different human actions for the current state of a state machine. Step 6 to 9 describe that state transitions occur

1- $S^{t=0} = \{(s_i^0, w_i^0) \mid i = 1,\dots,N\}$

2- $w_i = \frac{1}{N} \mid i = 1,\dots,N$

3- while (*Running*) do

4- for ($i = 1$ to N)

5- $w_i^{t+1} = w_i^t * p(z^t | s^t_{i,state_t})$

6- if $\left(p(z^t | s^t_{i,state_t}) == \underset{z}{\operatorname{argmax}} \langle p(z \mid s_{i,state_t}) \rangle \right)$

7- $s^{t+1}_{i,state_{t+1}} = s^t_{i,state_t+1}$

8- else

9- $s^{t+1}_{i,state_{t+1}} = s^t_{i,state_t}$

10- $S^{t+1} = S^{t+1} \cup \{(s_i^{t+1}, w_i^{t+1})\}$

11- endfor

12- for ($i = 1$ to N)

13- $w_i^{t+1} = \dfrac{w_i^t}{\sum_{i=1}^{N} w_i^t}$

14- endfor

15- for($i = 1$ to N)

16- if($s^{t+1}_{i,state_{t+1}} == s_{i,final}$)

17- if($w_i^{t+1} == \underset{w}{\operatorname{argmax}} \langle w_i^{t+1} \rangle$)

18- output Intention($s^{t+1}_{i,state_{t+1}}$)

19- endif

20- reintialize

21- endif

22- endfor

23- endwhile

Figure 6. Intention recognition using state machines

only in the state machines in which the current observation is highly likely for the currently active state. At step 10 the set S is updated. The weights of the machines are normalized at step 13. From step 15 to 21 it is checked if any state machine has reached its final state then it is checked that if the state machine has the highest weight then the concerned intention is output and the state machines are reinitialized. Otherwise the state machines are simply reinitialized.

III. Results and Discussion

The experiments have been performed with a robotic arm. The human and the robot interact in a human-robot interaction workspace shown in Fig. 7. The video data is captured with a FireWire digital camera with the standard frame size of 640 x 480. Human-robot interaction and image analysis is implemented using c++. The robot reactions are realized using the V++ for the robotic arm. The robot is communicated the cooperative instructions using the TCP/IP connection for assigning different operation, e.g., pick, place and move to a certain location, etc. Skin detection and Fourier descriptors are used for the image analysis.

Figure 7. Human robot interaction workspace.

In order to evaluate the human-robot cooperation by recognizing the explicitly and implicitly communicated human intentions, different human-robot interaction scenarios are considered. The interaction activities corresponding to the 5 explicitly and 2 implicitly communicated intentions are described below. The explicitly communicated intentions are as under.

1. *Picking and placing intention of an object.* The human intends to move an object from one place to another place in the human-robot collaboration workspace. The human explicitly communicate his intention by performing the corresponding actions.
2. *Passing intention of the human.* The human has the intention of passing an object to the robot and performs the concerning action.
3. *Placing intention of the human.* The robot places the already picked up object at a specific place according to the human intention.
4. *Picking and holding intention of an object.* The human intends that the robot picks up a specific object in the human-robot collaboration workspace.
5. *Giving a pointed object intention.* The robot provides the human the intended object that exists in the human-robot collaboration workspace.

The above described intentions from 1 to 5 were tested with 3 persons. The results of the tested intentions and the recognized intention for the explicitly communicated tested intentions are given in the Table 2 given below.

TABLE II.
RESULTS OF EXPLICITLY COMMUNICATED INTENTIONS

Tested Intentions	Recognized intention					
	Int1	Int2	Int3	Int4	Int5	Experiments
Int1	19	0	0	0	0	20
Int2	0	20	0	0	0	20
Int3	0	0	18	0	0	20
Int4	0	0	0	20	0	20
Int5	0	0	0	0	20	20

The implicitly communicated intentions are described as under.

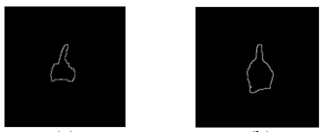
1. *Piling up of the objects.*
   Human comes into the scene and starts working without engaging the robot actively. The human starts pilling of the objects. The robot estimates the human intention by observing the human actions and the changes occurring in the scene information. After understanding the human intention of pilling up, the robot cooperates with the human by pilling the objects.
2. *Unpiling of the objects.*
   Human comes in the human-robot collaboration workspace and starts an operation of unpiling the objects without engaging actively the robot. The robot understands the human intention and unpiles the objects.

The results of the tested intentions and the recognized intention for the implicitly communicated tested intentions are given in the Table 3 given below.

TABLE III.
RESULTS OF IMPLICITLY COMMUNICATED INTENTIONS

Tested Intention	Recognized Intentions		
	Int1	Int2	Experiments
Int1	7	0	10
Int2	0	9	10

The false results shown in the Tables II and III are due to the unrecognized human hand posture, e.g., the pointing hand posture shown in Fig. 8(a) is recognized as pointing hand while the hand posture in Fig. 8(b) is not recognized as the pointing hand. In case if no expected action sequence is observed then no intention is recognized.



(a) (b)

Figure 8. Extracted outlines of pointing hand posture

The Figs. 9 and 10 represent how the weights of the intentions represented by different state machines change during the intention recognition process.

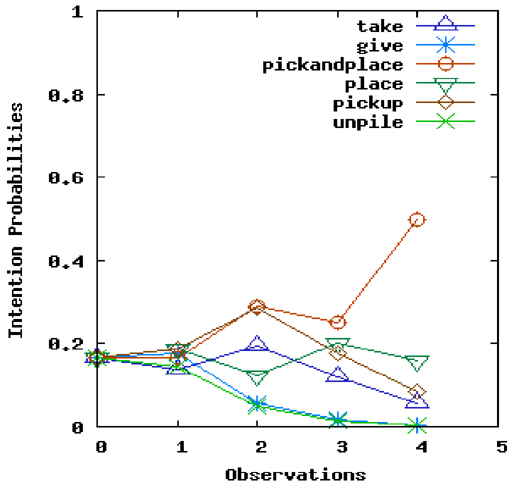


Figure 9. Picking and placing intention of an object.

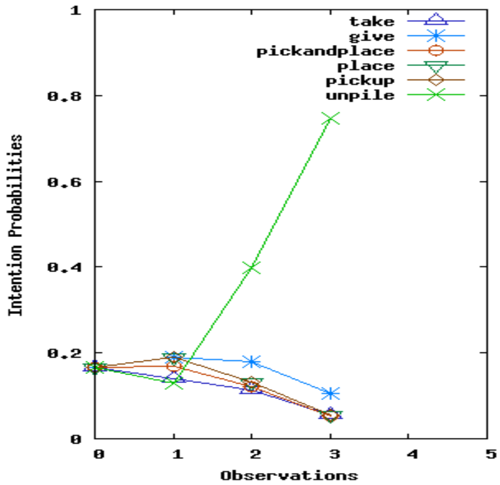


Figure 10. Intention recognition for unpiling the objects.

The graph in Fig. 9 describes the intention recognition of picking an object from one place and placing that object at another place. At the start all the intentions have equal probabilities as shown at observation 0 in Fig. 9. The first observation made is not directly related to any particular intention of the human. Therefore all the intentions get almost the same weight at observation 1. At the next step when the 2nd observation is made then the performed human action has high observation probability for *pickup*, *pickandplace* and *take*. Therefore the weights of these intentions go up and the weights of others go down at the observation 2 as shown in the Fig. 9. The state transitions occur in the state machines for which the observed human action is highly probable. Therefore state transitions occur in *pickup*, *pickandplace* and *take* state machines. At the observation 3 the perceived human action was unintentionally performed human action. The unintentional action stance may occur if a human changes his actions stance from one action to another. At observation 3 the perceived human action has high probability for *place* intention. Therefore the intention weight for *place* goes up and weights for others goes down. The state transition only occurs in the *place* state machine. At the observation 4 the performed human action has high probability for *pickandplace* and low probability for others. Thus the state transition only occurs in the *pickandplace* state machine and the weight of the intention also increases. The final state of the *pickandplace* state machine is reached and it has also has the high intention probability as compared to others. Thus the intention of *pickandplace* is recognized.

The intention graph shown in Fig. 10 describes the recognition of implicitly communicated intention of unpiling the objects. At observation 2 the recognized human action mainly corresponds to the *unpile* intention. Therefore the weight of the *unpile* intention increases and also the state transition occurs in *unpile* state machine. At this observation step the scene information is also taken into account as the human action concerned to an implicitly communicated intention is performed. The weights of other intentions decrease at observation 2. At observation 3 along with the human action, the scene information is inspected to check the increase or decrease in the unpiled objects. The human action and the scene information relates significantly to the unpile intention. Therefore the state transition occurs in the sate machine and the intention weight increase significantly. The state transition at observation 3 brings *unpile* state machine in the end state and the weight of the *unpile* intention is also high. Thus *unpile* intention is recognized.

IV. CONCLUSION

In this paper we have presented a probabilistic state machines based intention recognition algorithm. The suggested solution is applicable for both explicitly and implicitly communicated intention recognition. Explicit intention communication addresses to all the situations where human actively cooperate with the robot and implicit intention communication addresses to all the situations where human does not engage the robot but robot proactively starts the cooperation by recognizing the intention through scene information and human actions.

Addressing both explicitly and implicitly communicated intentions recognition make the human-robot collaboration more intuitive.

REFERENCES

[1] Andreas J. Schmid, and Oliver Weede, and Heinz Wörn. *Proactive Robot Task Selection Given a Human Intention Estimate*. Robot and Human interactive Communication, RO-MAN 2007. The 16th IEEE International Symposium 726-731.

[2] Hochul Jeon, Taehwan Kim, and Joongmin Choi. *Ontology-based user intention recognition for proactive planning of intelligent robot behaviour*. International Conference on Multimedia and Ubiquitous Engineering, 2008.

[3] Karim A. Tahb. *Intention recognition of a human commanding a mobile robot*. Proceedings of the IEEE, Conference on Cybernetics and Intelligent Systems, Singapore, pages 896:1-6, 1-3 December, 2004.

[4] Oliver C. Schrempf, and Uwe D. Hanebeck. *A Generic Model for Estimating User-Intentions in Human-Robot Cooperation*, in Proceedings of the 2nd International Conference on Informatics in Control, Automation and Robotics, ICINCO 05, 2005.

[5] Richard et al. *Understanding Human intentions via Hidden Markov Model in Autonomous Mobile Robots*, Proceedings of ACM/IEEE international conference on Human-robot interactions, pages 367-374, Amsterdam, Netherlands, March 2008.

[6] So-Jeong Youn, and Kyung-Whan Oh. *Intention Recognition using a Graph Representation*. World Academy of Science, Engineering and Technology, 25, 2007.

[7] Wenji Mao and Jonathan Gratch. A utility-based approach to intention recognition. *Workshop on Agent Tracking: Modeling Other Agents from Observation*, 2004.