Towards Task Understanding through Multi-State Visuo-Spatial Perspective Taking for Human-Robot Interaction

Amit Kumar Pandey, Rachid Alami

Towards Task Understanding through Multi-State Visuo-Spatial Perspective Taking for Human-Robot Interaction

Amit Kumar Pandey and Rachid Alami
akpandey@laas.fr rachid.alami@laas.fr
CNRS ; LAAS ; 7 avenue du colonel Roche, F-31077 Toulouse, France
Université de Toulouse; UPS, INSA, INP, ISAE ; LAAS ; F-31077 Toulouse, France

Abstract

For a lifelong learning robot, in the context of task understanding, it is important to distinguish the 'meaning' of a task from the 'means' to achieve it.

In this paper we will select a set of tasks in a typical Human-Robot interaction scenario such as show, hide, make accessible, etc., and illustrate that visuo-spatial perspective taking can be effectively used to understand such tasks' semantics in terms of 'effect'. The idea is, for understanding the 'effects' the robot analyzes the reachability and visibility of an agent not only from the current state of the agent but also from a set of virtual states, which the agent might attain with different level of efforts from his/its current state.

We show that such symbolic understandings of tasks could be generalized to new situations or spatial arrangements, as well as facilitate 'transfer of understanding' among heterogeneous robots. Robot begins to understand the semantics of the task from the first demonstration and continuously refines its understanding with further examples.

1 Introduction

The robots in the human centered environment will soon be expected to be able to acquire and enhance their knowledge life long, as humans do. In this context, from the task point of view, we identify 4 essential and complementary components: (i) Symbolic level understanding of the task's semantics, (ii) Situation dependent symbolic level planning to perform the task, (iii) Symbolic to execution level mapping of plan, (iv) Execution of the task. Hence it is important that the understanding of the task could be generalizable to a variety of situations, without any need of providing the learning data for each and every situation. In the context of Human-Robot Interaction the learning approaches could be broadly divided into two categories from "what is being learnt" point of view: (i) trajectory based (ii) symbolic primitive based. In [12], robot learns the trajectory for pick-and-place tasks with constraints on orientations. In [11], robot adapts the trajectory for '*pour*' task to avoid collision. In such approaches the robot is not aware about the 'meaning' of the task and in some sense it is bound to follow the learnt trajectory, which makes the generalization difficult in different scenarios and on different robots. On the other side, in symbolic primitives based approaches, which is the focus of this paper, the task is (a) either learnt based on the sequence of the sub-tasks or (b) based on the effect in terms of changes in the environment.

In [1], a set of symbolic predicates, such as *on*, *under*, etc., has been used for the incremental learning of the task precedence graph, for the tasks of *pouring the bottle* and *laying the table*. In other approaches the task performed by the human is inferred as symbolic descriptions of sub-tasks. For example "*place an object next to another object*" would be inferred as something like '*reach'*, '*grasp'* and '*transfer_relative',* [2], and "*Take a bottle out of the fridge*" would be sub-symbolized as '*Open the fridge*', '*Grasp the bottle*', '*Get the bottle out of the fridge*', '*Close the fridge*' and '*Put the bottle on the table in a stable position*', [10]. In [3], the robot grounds the task of assembling a table in terms of '*reach'*, '*pick'*, '*place'* and '*withdraw'*, and tries to learn the dependencies in order to reorder and adapt for different initial setups. In [9], a hybrid approach tries to represent the task in a symbolic sequence but also incorporates trajectory information to perform the task. But these approaches try to represent a task from the point of view of execution of sub tasks. The reasoning on the task semantics independent of the execution is missing.

On the other hand from the aspect of analyzing effects in terms of the task driven changes in the environment, [4] analyzes it in terms of '*holding object'*, '*hand empty'*, '*object at location'*, etc., for the pick-and-put task domain. In [6] robot performs different actions such as grasp, touch and tap on different objects to analyze the effects; once learnt could be used to select the appropriate action for achieving a particular effect, [7]. A survey on learning from demonstration can be found in [5]. But in these approaches also, the effects are analyzed from the point of view of different

This work has been partially conducted within the EU Project CHRIS (Cooperative Human Robot Interaction Systems) funded by the E.C. Division FP7-IST under Contract 215805.

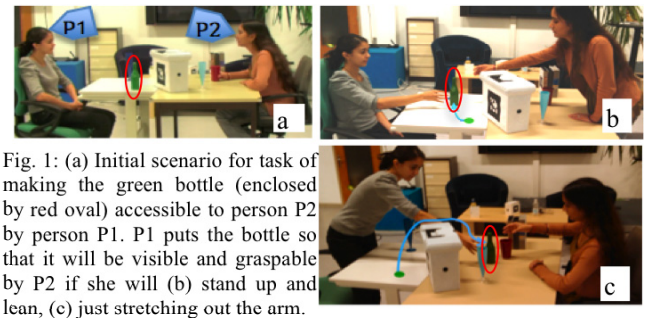

Fig. 1: (a) Initial scenario for task of making the green bottle (enclosed by red oval) accessible to person P2 by person P1. P1 puts the bottle so that it will be visible and graspable by P2 if she will (b) stand up and lean, (c) just stretching out the arm.

states of the object in the world frame. The effects on the object from the agents' perspective have not been explored.

In this paper we will (i) exploit the complementary issue of reasoning on the object from visuo-spatial perspectives of the agents, (ii) enable the robot to understand the task semantics independent of the task execution, (iii) separate trajectory information from task understanding. Although such information could be used during planning and execution as reference for 'how to' perform. All these will serve for better generalization of the task for unknown scenarios as well as transfer of understanding to heterogeneous robot.

We will use the term 'explanation-based understanding', because similar to explanation-based learning (see [8], for the learning aspects of explanation-based reasoning) the robot will be capable of deriving the 'effects' of a task through a single 'good' demonstration. If the demonstration is 'good' it will not result into unresolved or ambiguous effects. But if there is any misunderstanding or ambiguity robot could resolve it with further demonstrations. We consider a set of typical human-human interaction tasks such as make an object accessible, show or hide an object, etc. Robot benefits form its ability to perform visuo-spatial reasoning for agents not only from their current state/position but also from a set of possible achievable states by the agent.

1.1 Motivation

One of the common tasks in Human-Human Interaction is to make an object accessible to a person, which is currently invisible and/or unreachable for that person. As shown in fig. 1 depending upon the current state, relation, etc., person *P1* could take the bottle and put it at a place to make it visible and graspable by *P2* but the associated cost for doing so can vary, as in the cases (b) and (c) of fig. 1. The interesting point is: *P1* perceives various abilities of *P2* not only from her current state but also from the virtual state that if P2 will stand up and lean forward, she could get the bottle.

Now assuming a robot is observing the task as performed in fig. 1(c), and able to learn in the terms of symbolic subtasks such as '*grasp bottle*', '*carry bottle'* and *'put bottle'* at 'x' distance from the person P2 or put the bottle reachable for P2, then it will not be able to identify that the task performed in fig. 1(b) is same task. This is because what the robot has learnt is actually how to perform the task, not the semantics: "the object should become 'easier' to see, reach and grasp for the target person than it was before".

Moreover, if the robot will not be able to reason about multi-state abilities of the agents, it will fail to understand such semantics or could 'misunderstand' the task with poor generalization.

TABLE I
STATES FOR MULTI-STATE VISUO-SPATIAL PERSPECTIVE TAKING

Reachability States	Visibility States
Current	Current
Sitting Straight	Sitting Straight Head
Sitting Turn Around	Sitting Turn Head
Sitting Lean Forward	Sitting Lean Torso and Turn Head
Sitting Turn and Lean	Sitting Turn Torso and Turn Head
Standing Straight	Sitting Turn-Lean Torso and Turn Head
Standing Turn Around	Standing Straight Head
Standing Lean Forward	Standing Turn Head
Standing Turn and Lean	Standing Lean Torso and Turn Head
	Standing Turn Torso and Turn Head
	Standing Turn-Lean Torso and Turn Head

In this paper first we will enhance various states of an agent presented in [13] to perform multi-state visuo-spatial reasoning at object level. Then we will categorize the effort levels for state-transition, followed by approach of extracting various symbolic visuo-spatial facts and reason on them for task understanding. Successively we will analyze experimental results followed by discussion on potential applications of such symbolic understanding of tasks.

2 Methodology

2.1 Multi-state visuo-spatial reasoning

To reason on various abilities such as reachability, visibility of an agent for a particular object, robot virtually puts the agent in various states as shown in table I. An object is said to be reachable if at least one cell (dimension *5 cm* x *5 cm* x *5 cm* in current implementation) belonging to the object is within the length of the fingertip from the shoulder, that is how we mostly estimate reachability in a particular posture [16]. As an object might be reachable to touch, push, point, grasp, etc., robot is further equipped to distinguish whether the reachable object is graspable or not. Also it estimates how much the object is visible in a particular state of agent:

$$visibility\_score^{object}_{agent,state} = \frac{NP^{object}_{visible\_in\_FOV}}{NP_{FOV}} \quad ...(i)$$

Where NP denotes number of pixels in the image of visual perspective, i.e. in field of view (FOV) of agent.
Fig. 2(a) shows 3D representation of initial real world setup.

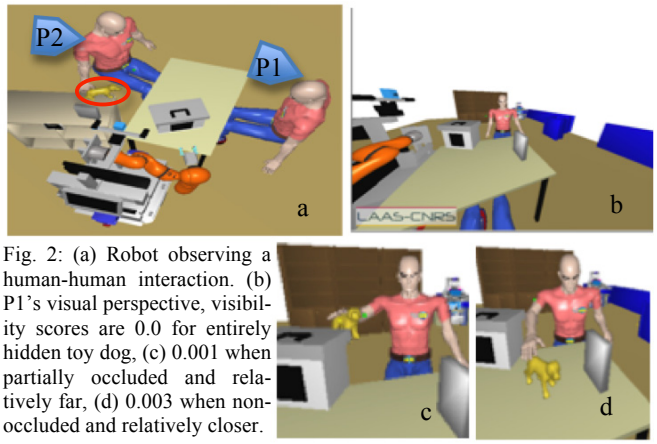

Fig. 2: (a) Robot observing a human-human interaction. (b) P1's visual perspective, visibility scores are 0.0 for entirely hidden toy dog, (c) 0.001 when partially occluded and relatively far, (d) 0.003 when non-occluded and relatively closer.

TABLE II
EFFORT CLASSES FOR VISUO-SPATIAL ABILITIES

Effort to Reach	Effort Level	Effort to See
No_Effort_Required	*Minimum*	*No_Effort_Required*
Arm_Effort		*Head_Effort*
Arm_Torso_Effort		*Head_Torso_Effort*
Whole_Body_Effort		*Whole_Body_Effort*
Displacement_Effort		*Displacement_Effort*
No_Possible_Known_Effort	*Maximum*	*No_Possible_Known_Effort*

Fig. 2(b) shows human *P1*'s visual perspective estimated by robot. The increasing visibility scores for the object toy dog, encircled in red in fig. 2(a), from *P1*'s perspective have been shown for three different cases fig. 2(b)-(d).

2.2 Classifying efforts for state transition

Fig. 3. Taxonomy of reach actions:(a) arm-shoulder reach, (b) arm-torso reach, (c) standing reach.

Robot dynamically categorizes the relative effort to attain a state from another state by the agent in terms of associated joints. This classification shown in table II is motivated from the studies of human movement and behavioral psychology, [14], [15], where different types of reach actions of the human have been identified and analyzed, fig. 3. The effort level categorization could be enhanced based on the studies of musculoskeletal kinematics and dynamics models such as [17].

2.3 Understanding task semantics

The robot will try to understand the task in terms of the changes on the agent's abilities to see, reach, grasp and visibility score of the object. We use the term '*performing-agent*' for the agent who will perform the task for a '*target-agent*', for whom the task is being performed, on a '*target-object*'. We will explain the approach through an example. Fig. 2(a) shows the initial world state before performing the task of making the yellow toy dog accessible to *P1* by *P2*.

Finding least effort state transition before task

For the *target-agent*, robot first finds whether the object is reachable, visible and graspable or not from the before-task state. If not then robot tries to find the least effort needed by the agent to reach, see or grasp the object. For this, depending upon the actual state of the agent robot tries to virtually put the agent in a series of states in the order of the efforts. For example if the agent is sitting then robot will first find the required yaw and pitch of the head to turn it towards the object, respecting the joint limits. If object is still not visible (because of joint limit or occlusion), robot will try to put the human in higher effort state such as turn torso and head, standup and turn around, etc., until the object becomes visible or the maximum allowed effort level has been reached. In this way for each ability type, robot finds the least effort required by the agent. For our example of fig. 2(a) robot finds that if *P1* will stand up and lean forward he will be able to see the toy dog thus categorizing the visibility effort as *Whole_Body_Effort,* from table II. Robot also found that *P1* could not reach the object from any of the states from his current position, so it categorizes the reachability effort as

TABLE III
VARIOUS AFTER TASK OBSERVATIONS

Reachability and Visibility	Ability to Grasp	Visibility Score
Easiest_Effort_Maintained (S)	*Graspability_Maintained (S)*	*Almost_Same (S)*
Effort_Becomes_Easier (S)	*Becomes_Graspable (S)*	*Increased (S)*
Effort_Becomes_Difficult (NS)	*Graspability_Lost (NS)*	*Increased_Significantly (S)*
	Still_Not_Graspable (NS)	*Decreased (NS)*
*(**S**: Supportive, **NS**: Non Supportive for an agent)*		*Decreased_Significantly (NS)*

'Displacement_Effort' (currently robot does not place the agent at new location to estimate abilities and assumes if human will move he will be able to see or reach the object).

Finding least effort state transition after-task

Similarly the robot finds least effort required by the *target-agent* for the new position of the *target-object*. But this time it will be calculated from the state of the *target-agent* at the end of the task, as he might adapt his state to favor the *performing-agent* or the task. For our example, fig. 2(d) shows the final state of the world from *target-agent*'s perspective. Robot estimates that *target-object* is visible by *target-agent* with an increased visibility score of 0.003 and will be reachable and graspable by him if he will lean forward. Hence it categorizes after-task reachability effort as *Arm_Torso_Effort* and for visibility as *No_Effort_Required*.

Extracting the effect of a task

Next step is to find the effects in terms of the changes from the visuo-spatial perspective of *target-agent*. For this robot compares the least effort state transitions before and after the task and categorizes the difference as one of the observations shown in table III. Robot found for our current example, for the toy horse, that for *target-agent P1,* to reach: *Effort_Becomes_Easier,* to see: *Effort_Becomes_Easier,* grasp: *Becomes_Graspable,* visibility score: *Increased*.

Finding the generalized meaning

After observing a demonstration, robot understands the task in terms of the *desirable changes* in the *target-agent*'s visuo-spatial abilities on the *target-object*. In the current example at this level of abstraction the robot understands: 'make object accessible means *target-object* should be easy to reach, grasp and see by the *target-agent*'. It further reasons at another level of abstraction to avoid over-constrained understanding as well as to facilitate continuous refining of the understanding as explained below.

Continuous Refining of the Understanding

It is possible that the robot has false belief about relevance of a predicate for a particular task. For example if the task is to hide an object, depending upon the places available for hiding, the *performing-agent* could put the *target-object* closer to the *target-agent* but behind some object which makes it invisible to *target-agent*. So, the robot will misunderstand that *target-object* should be difficult to be seen but easy to be reached by the *target-agent*. Hence reachability has been falsely associated as a relevant predicate for the task of hiding. So, there should be provision for continuous refinement with further demonstrations. For this, with every new observation of a task robot compares its past under standing for the '*consistency*' or '*contradiction*' about the belief of the relevance of a particular ability.

We define *Observation Occurrence Belief (OOB)* for a particular '*task_type*' for a particular '*ability_type*' as:

$$OOB^{task\_type,\,ability\_type}_{observation\_type} = \frac{N^{task\_type,ability\_type}_{observation\_occurred}}{N^{task\_type}_{demonstrations}} \quad ...(ii)$$

Numerator denotes the number of times, for the *target-object*, the particular observation, from table III (such as *Effort_Becomes_Easier, etc.*), has been observed about a particular ability (such as reachability, etc.), for a particular task (such as *make accessible*, etc.). The denominator is number of times the task has been demonstrated. We classify the observations as supportive or non-supportive for an ability of an agent, marked *(S)* and *(NS)* in table III. For example if after a task the agent's ability to reach the *target-object* has been maintained or has become easier then it is supportive to that agent, and so on. Then we define two beliefs: *Supportive Observation Occurrence Belief (SOOB)* and *Non-Supportive Observation Occurrence Belief (NSOOB)* for a particular task as follows:

$$SOOB^{task\_type,\,ability\_type} = \sum_{i=1}^{n_s} OOB_i^{task\_type,\,ability\_type} \quad ...(iii)$$

$$NSOOB^{task\_type,\,ability\_type} = \sum_{i=1}^{n_{ns}} OOB_i^{task\_type,\,ability\_type} \quad ...(iv)$$

Where n_s and n_{ns} are number of supportive and non-supportive observations for a particular ability of the agent. Now robot can detect a *contradiction* in the observations from two or more demonstrations of a task. If for a particular ability type, *SOOB* and *NSOOB* both are non zero then that particular ability might not be relevant for that particular task and the observations for that ability is just a side effect. Let us assume that the task of hiding an object has been observed by robot twice. Depending upon the availability of places to hide, performing agent puts the object at a place, which made it difficult for the *target-agent* to see. But in one demonstration the *target-object* was easier to reach and in another it was difficult to reach. Hence for the visibility, *SOOB* is non-zero and *NSOOB* is zero but for reachability, *SOOB* and *NSOOB* both becomes non-zero after these two demonstrations. So, robot detects a '*consistency*' in visibility but '*contradiction*' in reachability from the *target-agent's* perspective. At this state instead of directly concluding that the reachability is irrelevant for the task of hiding, we further define '*non-relevance factor*' for a particular task type, *'t'* and a particular ability type *'a'* as:

$$non\_relevance^{t}_{a} = 1 - \frac{abs\left(SOOB^{t}_{a} - NSOOB^{t}_{a}\right)}{\left(SOOB^{t}_{a} + NSOOB^{t}_{a}\right)} \quad ...(v)$$

Eq. (v) will result into non-relevance factor as 1, for a particular ability, if contradiction and consistency have been observed for equal number of demonstrations. On the other hand if there has been no contradictions it will return 0 meaning the ability is relevant for the task and the observed effects should be maintained.

Because of the inheriting conflict driven calculation of non-relevance factor and assuming that the demonstrations can contain ambiguous effects but are not to intentionally produce a incorrect effect (i.e. we are not trying to teach a child with wrong demonstration), an ability will be treated as non-relevant even if there is less evidence, green band in fig. 4. But if the non-relevance factor for a particular ability is very low but non-zero, as shown a red '*confusion zone*' in fig. 4, then it will put robot in a '*confusing*' state because something seems to be non-relevant but the supporting evidence is not sufficient. This confusing situation could be treated in a variety of ways: (i) Treat the conflicting ability as relevant but give least preference to satisfy the related supportive or non-supportive (whichever is having higher belief) observation while performing that task. (ii) Communicate the confusion to the human for help to resolve. (iii) Simply discard the current demonstration causing confusion assuming that the task understood and refined with time in past is more stable understanding than the current single demonstration, so non relevance factor will always be either 0 or 1. But this will limit the flexibility of refinement and understanding will become 'rigid' after few observations. For the current implementation we adapt (i) but the work is in progress towards a hybrid approach combining (i) and (ii).

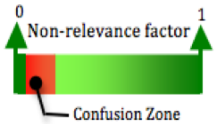

Fig. 4: Interpretation of non-relevance factor.

Separating execution preferences from task semantics

In the same framework the robot also extracts the 'effect' from the *performing-agent's* perspectives, which could serve as execution preferences. Also if the robot would be equipped with additional capabilities, it could further infer that *performing-agent* preferred to take the object from support plane-1 and put it on support plane-2, for the task of fig. 2. But there could be a case where the performing-agent would displace some other object, which is occluding and hindering the otherwise visible and graspable target-object from target-agent. So, if such task execution sequences would be used for task understanding, the robot will not be able to understand the 'desirable' meaning.

Robot could also have the trajectory of the object and the hand, but as mentioned earlier we prefer to put such information also in the execution preferences, which could facilitate robot for human preferable motion and behavior.

Planning and performing the understood task

Since the robot is equipped with the geometric interpretation of the symbolic terms, it can calculate a set of candidate space for performing a particular task. But this level of symbolic task understanding could provide more flexibility to the task planner about various ways to perform the task to achieve the desirable effect from the *target-agent's* perspective. We have adapted the framework presented in [13] to find candidate search space and perform a particular task.

3 Experimental Results and Analysis

Our robot is equipped with an integrated 3D representation and planning platform in which the models of all the agents and objects are updated online with the data through various sensors. Human gaze is simplified to the human head orientation. We have two ways of providing the robot with the data about human-human task performance: online and offline. The offline data is collected through markers based

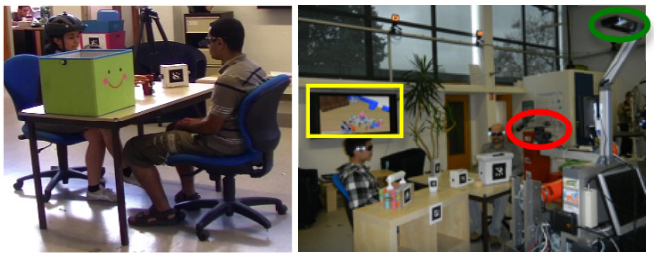

Fig. 5: Human-Human task performance data collection (a) by motion-capture system, (b) by observing online by the robot.

motion capture system, fig. 5(a). In the online process the robot directly observes the tasks, fig 5(b). It uses its stereo-vision system (enclosed by red oval) to identify and localize objects based on the tag and uses *Kinect* motion sensor system (enclosed by green oval) mounted on it to track the whole body of the humans. The yellow rectangle encloses the screen displaying online 3D representation of the environment by the robot. In both cases of getting the data, the task name, time stamps for starting and finishing of the task, information about the performing agent, target agent, target object are provided to the robot.

Fig. 5(a) is actually the final scenario of making the glass accessible to the person on the right by the person on the left. Following is the observation for this task by the robot, which is similar for the example scenario of fig. 2:

For *target-object*, for *target-agent*:

For reachability the *Effort_Becomes_Easier,* and for visibility the *Effort_Becomes_Easier,* and from the new easiest state to reach the *target-object Becomes_Graspable,* and from the new easiest state to see, the visibility score of the *target-object* has *Increased*.

Hence the robot is able to understand from a single demonstration that making an object accessible means the *target agent*'s ability to reach and see the object should be supported (which the robot interprets as the efforts should become easier or in worst case it should be maintained, while planning for a task). We have further demonstrated the same task twice with different environmental setups. Below is snap of the different beliefs after 3 demonstrations:

The *Observation Occurrence Belief (OOB)*:

for *reachability*:

Easiest_Effort_Maintained=0.33,
Effort_Becomes_Easier=0.66,
Effort_Becomes_Difficult=0

for *visibility*:

Easiest_Effort_Maintained=0.66,
Effort_Becomes_Easier=0.33,
Effort_Becomes_Difficult=0

…

Based on *OOB* robot further concludes that:

Supportive Observation Occurrence Belief (SOOB):
for reachability = 1, for visibility =1, …

Non-Supportive Observation Occurrence Belief (NSOOB):
for reachability =0, for visibility =0, …

Non relevance factor for: reachability =0, visibility =0, …

Hence till 3 demonstrations of the task '*make object accessible*' robot did not find any contradiction in belief and will maintain all the effects while performing the task.

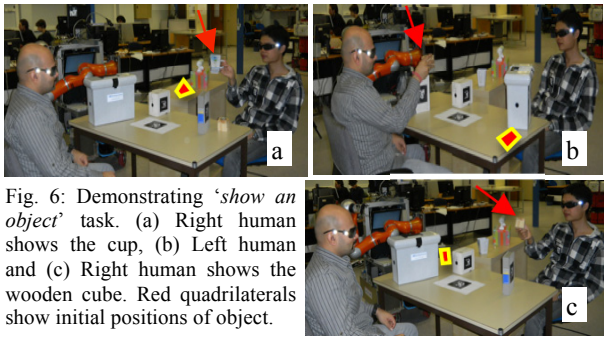

Fig. 6: Demonstrating '*show an object*' task. (a) Right human shows the cup, (b) Left human and (c) Right human shows the wooden cube. Red quadrilaterals show initial positions of object.

Now we will demonstrate how the robot refines its misunderstanding about relevance of 'reach' for the task of showing an object, similar to the task of hiding an object. Fig. 6(a)-(c) show final scenarios in three different situations for the task of showing an object. Red quadrilaterals show initial positions of the target object (which is cup in (a) and wooden cube in (b) and (c)), the red arrows mark the final position of object in hand. For (a) and (c) the person on the left was the *target-agent* whereas for (b) he was the *performing-agent*. Table IV shows the refined understanding of robot after each demonstration, for two abilities: visibility and reachability. Note that because of significant non-relevance factor after incorporating observation from 3rd demonstration, robot understood that reachability from the *target-agent's* perspective is not relevant for this task.

We have demonstrated various other tasks to the robot, such as *give an object*, *hide an object*, *hide an object away*, *put an object away*, etc. Table V shows the understanding of the robot in terms of effects on visibility and reachability from the *target-agent's* perspective. Table V also shows number of demonstrations per task, *N,* and the average processing time per demonstration, *T,* once the initial and final world states are known to the robot. It is interesting to observe that *T* is more for the tasks, which require the robot to put the target-agent in more number of states before getting first state of least effort satisfying reachability or visibility.

Another observation is that for the task of '*Make Accessible*' and '*Give*' robot understanding is similar. Perhaps they are same or there might be some differences, which are hid-

TABLE IV
ROBOT'S UNDERSTANDING FOR TASK SHOW AN OBJECT

Observation Occurrence Belief (OOB)						
After Demonstration ->	1		2		3	
Ability of target agent ->	See	Reach	See	Reach	See	Reach
Easiest_Effort_Maintained	0	0	0	0.5	0	0.33
Effort_Becomes_Easier	1	1	1	0.5	1	0.33
Effort_Becomes_Difficult	0	0	0	0	0	0.33
SOOB	1	1	1	1	1	0.66
NSOOB	0	0	0	0	0	0.33
Non Relevance Factor	**0**	**0**	**0**	**0**	**0**	**0.67**

TABLE V
ROBOT'S UNDERSTANDING OF DIFFERENT TASKS

Task	*Visibility*	*Reach*	*Vis. Score*	*Grasp*	*N*	*T (s)*
Show	Supp	Not Relv	Supp	Not Relv	4	0.48
Hide	Non-Supp	Not Relv	Non-Supp	Not Relv	3	0.67
Make Accessible	Supp	Supp	Supp	Supp	3	0.4
Give	Supp	Supp	Supp	Supp	2	0.42
Put Away	Supp	Non-Supp	Supp	Non-Supp	3	0.51
Hide Away	Non-Supp	Non-Supp	Non-Supp	Non-Supp	2	0.83

Supp: Supportive, **Non-Supp**: Non-Supportive, **Not-Relv**: Not-Relevant

den in the layers below the current level of abstraction. Such situations need exploration at lower levels to find relevant information to disambiguates the tasks understanding. For example, for the 'give' task the least effort for reachability is always *No_Effort_Required* whereas for '*make accessible*' task it is varying. And the inherited meaning of *No_Effort_Required* for reach is *target-object* should be in hand of *target-agent*. This is a pointer, which needs further investigation to disambiguate by involving other predicates.

4 Discussion on Potential Applications

Symbolic understanding of a task along with its geometric counterpart makes the robot more 'aware' about its behavior. Below we discuss few of the potential applications.

Generalization to novel scenario

As the understanding of task is independent of the relative arrangements of agents and objects, it facilitates the robot to perform the task in an entirely different way as well as scenario. As shown in fig. 7 robot is making the wooden cube (initially at the place of red rectangle) accessible to the human by putting it at the top of a box because robot was not able to plan a collision free trajectory for any other commonly reachable and visible place of less effort from human's perspective on the table.

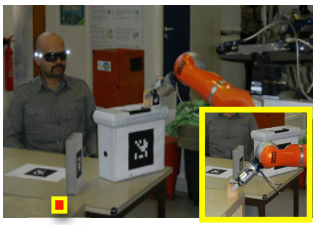

Fig. 7: Robot making an object accessible in a novel scenario.

Transfer of understanding among heterogeneous agents

Since the robot has understood the task independent of the trajectory planning and control level execution, it can easily transfer the task semantics to another robot of entirely different kinematics structure. And the other robot equipped with similar capabilities of visuo-spatial perspective taking of the agents could then interpret the understanding and perform it by respecting its own constraints.

Generalization for multiple target-agents

This level of symbolic understanding would also help in generalizing for multiple target humans. Such as hide from two humans, show to a group of people, etc.

Greater flexibility to the symbolic planner, interaction

If the planner at symbolic level knows the semantics of a task independent of execution, it could plan to achieve the task in a variety of ways. Such as it could decide to cover an object by another object to hide or to involve a third agent.

Such symbolic awareness about the task's semantics could also enrich the verbalize interaction with the human as well as could help in generating shared co-operative plan for achieving complex tasks. Such understanding could also be used to predict action and show proactive behavior.

5 Conclusion and Future Works

This paper is towards making the boundary between task primitives and execution primitives evident and enables the robot to understand task's semantics independent of the means to achieve it. As a primary step we have incorporated a complementary but important aspect, multi-state visuo-spatial perspective taking, to understand basic human-robot interaction tasks by the robot.

We have shown that such understandings would be easy to generalize to novel scenario as well as for heterogeneous robots. The presented approach could further be benefited by incorporating estimation of additional abilities and primitives to understand more complex tasks. Another interesting future work is to enable robot with the capabilities of autonomously finding inter-task relations, such as '*give*' could be '*show*' with some additional constraints.

References

[1] M. Pardowitz and R. Dillmann, "Towards life-long learning in household robots: The piagetian approach," in Proc. 6th IEEE International Conference on Development and Learning, 2007.

[2] A. Chella, H. Dindo, and I. Infantino, "A cognitive framework for imitation learning," Robotics and Autonomous Systems, Vol. 54, no. 5, May 2006, pp. 403-408.

[3] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching: extracting reusable task knowledge from visual observation of human performance," IEEE Transactions on Robotics and Automation, vol. 10, no. 6, Dec 1994, pp.799-822.

[4] S. Ekvall and D. Kragic, "Robot Learning from Demonstration: A Task-level Planning Approach," International Journal of Advanced Robotic Systems, vol. 5, 2008, pp. 223.

[5] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," Robotics and Autonomous Systems, Vol. 57, no. 5, May 2009, pp. 469-483.

[6] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-victor, "Modeling affordances using bayesian networks," IROS, 2006.

[7] M. Lopes, F. S. Melo, and L. Montesano, "Affordance-based imitation learning in robots," IROS, 2007.

[8] N. S. Flann and T. G. Dietterich, "A Study of Explanation-Based Methods for Inductive Learning," Machine Learning, vol. 4, no. 2, Nov. 1989, pp. 187-226.

[9] K. Ogawara, J. Takamatsu, H. Kimura, and K. Ikeuchi, "Extraction of essential interactions through multiple observations of human demonstrations," IEEE Transactions on Industrial Electronics, vol. 50, no. 4, Aug 2003, pp. 667- 675.

[10] R. Dillmann, "Teaching and learning of robot tasks via observation of human performance," Robotics and Autonomous Systems, Vol. 47, June 2004, pp. 109-116.

[11] M. M¸hlig, M. Gienger, S. Hellbach, J. J. Steil, and C. Goerick, "Task-level Imitation Learning using Variance-based Movement Optimization," ICRA, 2009.

[12] E. Gribovskaya, S. M. Khansari-Zadeh, and A. Billard, "Learning Non-linear Multivariate Dynamics of Motion in Robotic Manipulators," International Journal of Robotics Research, vol. 30, no. 1, Jan. 2011, pp. 80-117.

[13] A. K. Pandey and R. Alami, "Mightability Maps: A Perceptual Level Decisional Framework for Co-operative and Competitive Human-Robot Interaction," IROS, 2010.

[14] D. L. Gardner, L. S. Mark, J. A. Ward, and H. Edkins, "How do task characteristics affect the transitions between seated and standing reaches?," Ecological Psychology, vol. 13, 2001, pp. 245-274.

[15] H. J. Choi and L. S. Mark, "Scaling affordances for human reach actions," Human Movement Science, vol. 23, 2004, pp. 785-806.

[16] C. Carello, A. Grosofsky, F. D. Reichel, H. Y. Solomon, and M. T. Turvey, "Visually Perceiving What is Reachable," Ecological Psychology, Vol. 1, no. 1, March 1989, pp. 27-54.

[17] O. Khatib, E. Demircan, V. De Sapio, L. Sentis, T. Besier, and S. Delp. "Robotics-based Synthesis of Human Motion," Journal of Physiology Paris. Vol. 103, no. 3-5, August 2009, pp. 211-219.