2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Anticipating Human Activities for Reactive Robotic Response

Hema Swetha Koppula[1] and Ashutosh Saxena[1]

I. INTRODUCTION

An important aspect of human perception is anticipation, which we use extensively in our day-to-day activities when interacting with other humans as well as with our surroundings. Anticipating which activities will a human do next (and how to do them) can enable an assistive robot to plan ahead for reactive responses in the human environments.

In this work, our goal is to enable robots to predict the future activities as well as the details of how a human is going to perform them in short-term (e.g., 1-10 seconds). For example, if a robot has seen a person move his hand to a coffee mug, it is possible he would move the coffee mug to a few potential places such as his mouth, to a kitchen sink or just move it to a different location on the table. If a robot can anticipate this, then it would rather not start pouring milk into the coffee when the person is moving his hand towards the mug, thus avoiding a spill. We represent each possible future using an anticipatory temporal conditional random field (ATCRF) that models the rich spatial-temporal relations through object affordances. We then consider each ATCRF as a particle and represent the distribution over the potential futures using a set of particles.

We evaluate our anticipation approach extensively on CAD-120 human activity dataset [1], which contains 120 RGB-D videos of daily human activities, such as *microwaving food*, *taking medicine*, etc. For robotic evaluation, we measure how many times the robot anticipates and performs the correct reactive response. The accompanying video shows a PR2 robot performing assistive tasks based on the anticipations generated by our proposed method.

II. METHOD

We model three main aspects of the activities. First, we model the activities through a hierarchical structure in time where an activity is composed of a sequence of sub-activities [1]. Second, we model their inter-dependencies with objects and their affordances. The object affordances are represented in terms of the object's relative position with respect to the human and the environment.[1] Third, we model the motion trajectory of the objects and humans, which tells us how the activity can be performed. Modeling trajectories not only helps in discriminating the activities, but is also useful for the robot to reactively plan motions in the workspace.

For anticipation, we present an anticipatory temporal conditional random field (ATCRF), where we start with modeling the past with a standard CRF (based on [1]) but augmented with nodes/edges representing the object affordances, sub-activities, and trajectories in the future. Since there are many possible futures, each ATCRF represents only one of them. In order to find the most likely ones, we consider each ATCRF as a particle and propagate them over time, using the set of particles to represent the distribution over the future possible activities.

III. EXPERIMENTS

We performed an extensive evaluation on CAD-120 human activity RGB-D dataset. For a new subject (not seen in the training set), we obtain an activity anticipation accuracy (defined as whether one of top three predictions actually happened) of 75.4%, 69.2% and 58.1% for an anticipation time of 1, 3 and 10 seconds respectively. Fig. 1 shows how the performance changes with the future anticipation time.

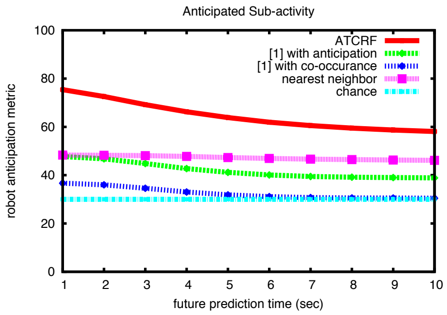

Fig. 1. Plot showing how *robot anticipation metric* changes with the future anticipation time.

We also consider the following scenario for evaluating our algorithm on the robot for an assistance task: Robot is instructed to refill water glasses for people seated at a table, but when it anticipates an interaction with the cup, it waits for the interaction to complete before refilling. The robot considers the three top scored anticipations for taking the decision. We considered 40 pour instructions given during 10 interaction tasks, and obtained a success rate of 85%, which is the fraction of times the robot correctly identifies its response ('to pour' or 'not pour').

The accompanying video shows PR2 robot performing two assistive tasks based on the generated anticipations. In the first task, the robot assists in the activity by opening the fridge door when it sees the person approaching the fridge with an object. In the second task, the robot serves a drink without spilling by anticipating the person's interactions with the cup.

REFERENCES

[1] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," *IJRR*, 2013.

[1] Department of Computer Science, Cornell University. {hema,asaxena}@cs.cornell.edu

[1] For example, a *drinkable* object is found near the mouth of the person performing the *drinking* activity and a *placeable* object is near a stable surface in the environment where it is being placed.

978-1-4673-6357-0/13/$31.00 ©2013 IEEE