Modeling structured activity to support human-robot collaboration in the presence of task and sensor uncertainty

Kelsey P. Hawkins Shray Bansal Nam Vo Aaron F. Bobick

***Abstract*— A representation for structured activities is developed that allows a robot to probabilistically infer which task actions a human is currently performing and to predict which future actions will be executed and when they will occur. The goal is to enable a robot to anticipate collaborative actions in the presence of uncertain sensing and task ambiguity. The system can represent multi-path tasks where the task variations may contain partially ordered actions or even optional actions that may be skipped altogether. The task is represented by an AND-OR tree structure from which a probabilistic graphical model is constructed. Inference methods for that model are derived that support a planning and execution system for the robot that attempts to minimize a cost function based upon expected human idle time. We demonstrate the theory in both simulation and actual human-robot performance of a two-way-branch assembly task. In particular we show that the inference model can robustly anticipate the actions of the human even in the presence of unreliable or noisy detections because of its integration of all its sensing information along with knowledge of task structure.**

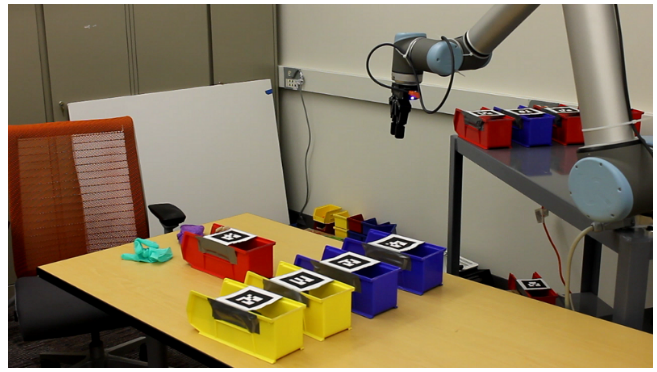

Fig. 1: Station where a Universal UR10 robot assists a human by fetching and removing bins as needed by anticipating actions of the human.

I. Introduction

Robots are potentially powerful tools for assisting people in a broad range of applications including industrial manufacturing and assembly [1], as well as personal services [2]. Most prior work in human-robot collaboration has focused on two research topics: acquiring skills by demonstration or teaching, and how to properly engage with users [3]. However, the uncertainty of real-world action detection — of belief as to what the human agent is doing at any given time — has been largely engineered out of the problem. For robots that share an environment with humans, the timing and identity of the human's actions change how and when the robot should react to make the interaction fluent, effective, and safe. Thus, maintaining a representative belief about the past, current, and future state of the human in the face of perceptual uncertainty and human variability is important for improving daily human-robot interaction.

In this paper we focus on the specific interaction scenario in which a human is performing one of a variety of assembly tasks during which the robot must anticipate which parts the human will need and when. Even though the partial ordering of possible human actions is available to the robot before collaboration, the branch choices the human selects are never explicitly communicated. Futhermore, the timing of human actions is naturally variable due to sources such as mistakes, differing expertise, and physiological state. Therefore, understanding when and what human actions occur must be implicitly inferred from perception, action duration priors, and task constraints. However, noisy and miscalibrated sensors, systematically occluded vision, and tracking failures render action detection uncertain. To recover from this uncertainty, the robot should utilize its information as effectively as possible.

We developed a probabilistic representation of human action which models the human's task, inferring both which action branches the human chooses and when each action is likely to have occurred, if ever. By modelling detector reliability, incorporating timing distributions, and reasoning over the full sensor history, the robot makes efficent use of its knowledge, integrating evidence over time to improve its belief about the human. Furthermore, by encoding human task constraints and action ordering structures, the system improves its robustness as the robot understands how its actions effect how the human acts. After inferring the human's state, we utilize a cost-based planner to optimize the robot's action plan with respect to the posterior human action distributions, reducing the expected cost for an arbitrarily defined system cost function.

The work present here greatly expands prior results in [4] by accommodating task structures that are not simple linear, sequential actions. In that work, the only inference to be performed was when future actions were likely to be performed and for the robot to anticipate accordingly. Here the system can respond to acyclic task structures with partially ordered actions we call multi-path tasks, where the human only performs actions along one path through

All authors are affiliated with the Center for Robotics and Intelligent Machines and The School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA. {kphawkins, namvo, sbansal34}@gatech.edu, afb@cc.gatech.edu This work was supported in part by BMW Project# RD441.

the graph. This representation is powerful since it allows us to encode different task goals, arbitrary action ordering possibilities, and optional actions.

We organize the remainder of our paper as follows. After discussing selected related work, we develop the representation and inference method for modeling multi-path tasks, assessing the likelihood that a given branch of the task is being performed, and predicting when a particular subtask will occur. Using these predictions we construct a planning and action system for the robot that attempts to minimize a cost function based upon expected human idle time. We demonstrate the theory in both simulation and actual human-robot performance of a simple two-way-branch assembly task. In particular, we show that the inference model can robustly anticipate the actions of the human even in the presence of unreliable or noisy detections because of its integration of all its sensing information along with knowledge of task structure.

II. Related Work

In robotics there has been significant recent study on the role of prediction on the fluency of human-robot interactions, along with the development of learning and planning algorithms that perform action selection in a collaborative context; such work usually presumes sensing is straightforward and that the challenge is making the right action decision. For example, [5] uses an adaptive Markov model to assign confidence about predictions of the human partner's actions. The uncertain predictions are used in a cost-based framework to select the best action. In both that work and subsequent efforts [6] the benefits of employing anticipatory actions in a human-robot task are well observed in human trials. In all these systems the actions of the human are presumed to be clearly and reliably observed.

In the robotics literature there is a variety of approaches to anticipating the actions of humans. These efforts vary in how much a priori knowledge the system has about the task or domain. Huber et.al. [7] provides the robot has complete knowledge of the sub-tasks performed by the human. Fish et. al. [8] and Tenorth [9] collect detailed statistics about the human performance of the specified task and predict duration variability over time. Koppula and Saxena [10] learn likely sequences of human action from observation training data. At run time, the robot instantiates a set of probabilistically weighted "anticipatory temporal conditional random fields" to predict which actions the human may take and when. The work presented here also explicitly models possible future sub-task sequencings and maintains a probability for each based upon prior info and current observations. But our possible futures are defined by an a priori task description.

Wilcox et. al. use strict temporal constraints to develop robotic schedules for human-robot collaborative assembly with the addition of preferences which optimize the plan over the constraints [1]. While they accommodate human variability by using different preferences for different behavior models, they do not address the issue of perceptual ambiguity. We note that the work presented here also frames action selection as minimum cost planning in the face of probabilistic beliefs about when the human will perform various sub-tasks.

Finally, computer vision research, specifically activity recognition, has also developed many approaches to modeling activities composed of sequences of actions. Perhaps the most relevant work is that of Shi et. al. [11] where a Dynamic Bayes Network variant was proposed to recognize partially ordered sequential action. Albanese et al. [12] uses probabilistic Petri nets to detect events while [13] learns an activity's decomposable structure of "actionlets" with a probabilistic suffix tree; given that data structure, early prediction of sub-action can be done. In [14], Tang et. al. demonstrated how to use variable-duration Hidden Markov Models to learn an action's latent temporal structure and showed it helps to improve detection results in the presence of noisy sensors. This is analogous to the work here where sensing information is integrated with a structural description of the task to improve action detection.

III. Representation and Inference

In [4], we developed a representation, inference procedure, and reactive planning system in the context of linear chain tasks. The system modeled the task as a known sequence of human actions, incorporating duration knowledge, task constraints, and detector observations simultaneously. Given a history of task constraints and detector observations up to the current time and an estimate of these values in the future, the system inferred the distribution over when human actions occurred or will occur.

The key development in that work was representing the chain as a sequential Bayes net where the state variables were the beginning and ending times of each of the actions. Duration models allowed for conditioning the end times upon start values, detectors were designed that provided diagnostic information as to when an action was occurring, and on-line inference procedures were developed that incorporated not only all detections viewed up to the current time but also task constraints such as whether the robot had performed a necessary action that would enable the human to progress in his task.

In this paper we significantly extend that work to allow for task variations where the human is not limited to a strict sequential chain but can be considered as "multi-path": the task may be a partially ordered one where certain actions can occur in a variety of orderings, or even a set of multiple tasks where some actions may be skipped altogether. We achieve this extension by allowing the tasks to be specified by an AND-OR tree structure and automatically constructing a probabilistic graphical model that reflects the branches of the specification. In addition we extend the state representation to include the possibility that an action never occurs — it never starts and thus never completes.

A. AND-OR representation of a task

We begin by defining a representation for the multi-path task. While a thorough exposition would involve a discussion of a formal grammar representation, here we

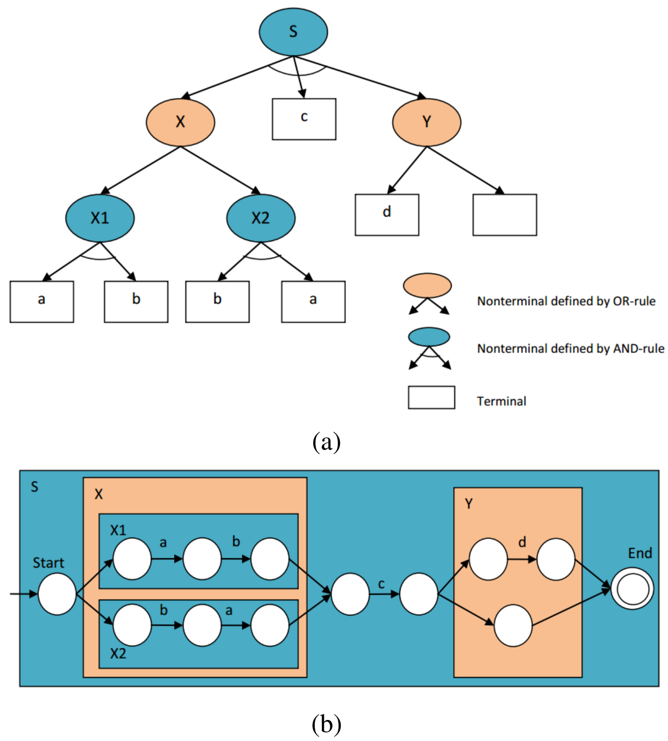

Fig. 2: An complex activity represented by (a) an AND-OR tree representation of a task and (b) the equivalent acyclic finite state machine

simply state that the tasks are described as AND-OR trees of sub-tasks where sub-tasks are either *primitives actions* or *compositions*. Primitives are analogous to the actions in our earlier work on linear chains [4] and, as described below, are assigned probabilistic detectors. The composition elements are represented by ANDs or ORs of smaller elements.

Figure 2 provides an example of a task whose components are primitives *a,b,c* and *d*. The task requires that first *a* and *b* are performed without an ordering constraint, followed by *c* and then by an optional action *d*. This task structure can be represented by the illustrated AND-OR tree or the analogous finite state machine.

As in our earlier work we will produce a Bayes network representing the temporal structure of the task with network variables being the start and end time of each element. And as before, inference over the network will determine the probability density of these variables given the sensor detections. However, these variables will no longer be restricted to representing only primitive actions.

The proposed representation is analogous to grammar-like or FSM descriptions in prior work from both robotics (e.g. [2]) and computer vision [11, 15]. One restriction for this work is that the length of the "string" of actions produced must be finite since the resulting Bayes net must explicitly represent every action that occurs.

B. Primitive actions and detectors

We refer to atomic actions of a task as *primitives*. These are actions that are not decomposed within our system and which are defined to have a start and end time; for some primitive action A we denote these variables A.s and A.e. These variables admit integer value between 1 and T (the assumed maximum length of the task), as we operate in discrete-time fashion.

For each primitive actions we have two important densities. The first is a duration model $P(A.e|A.s)$ representing the prior information about the duration of action A. In our implementation we use Gaussian model derived from training data: $P(A.e|A.s) \propto N(A.e - A.s; \mu_{Dur(A)}, \sigma_{Dur(A)})$ if $T \geq A.e \geq A.s \geq 1$, or 0 otherwise.

The second density is an observation likelihood $P(Z^A|A.s, A.e)$. For each primitive action we construct a visual detector that outputs a detection score $D_A[\alpha, \beta]$ representing how consistent the observed data is with the action starting at time α and ending at time β for every possible(α, β) of the entire input sequence. Then the likelihood can be computed based on that detection: $P(Z^A|A.s = \alpha, A.e = \beta) = h_A D_A[\alpha, \beta]$ for some constant h_A.

The choice of D_A reflects the sensitivity and reliability of the sensing system in being able detect the action A. If, for example, there was no available sensing, then D_A would be a constant, effectively eliminating any impact on the inference. In the example scenario that will be test in section V our sensing will involve tracking hand positions as the human reaches for a bin of parts. The detector for the human reaching into a specific bin is a mixture of two factors. The first is a Gaussian with mean $\mu_{Pos(A)}$ being the bin location and variance Σ_0 determined by the accuracy of the hand tracker while successfully tracking the hand; the variance is the same for all bins. The second factor is a constant w_m — representing a uniform distribution that arises when the tracker has failed ("missed") and is returning arbitrary values. The larger w_m is, the less confident the system is in its sensing. A w_m of infinity would correspond to the constant value mentioned earlier — no available sensing information.

C. Sequence of actions: AND

First we desribe the simpler composition: the task S is a fixed sequence of a number of primitive actions, for example A, B, C. To represent this composition we use the notation AND-rule: $S \rightarrow ABC$ As before we represent the start and end time as S.s and S.e. Also, as before Z^A denotes the visual observation of primitive action A, and define $Z = Z^S = Z^{A,B,C}$. The example network is shown in figure 3.

In our network, we define the start and the end of a task according to its subtasks (S.s = A.s, S.e = C.e). Also we assume the end of an action is the start of the next one (A.e = B.s, B.e = C.s). These are denoted as red edges in the network. The remaining conditional probabilities required are the duration values and the observation likelihoods of the primitives.

Given all the conditional probability tables computed, we use a message-passing algorihtm with forward and backward phase to perform inference and the output will be the posterior distributions of the start and end of every actions: $P(Z)$ and $P(S.s|Z), P(S.e|Z), P(A.s|Z), P(A.e|Z), \ldots$ [4].

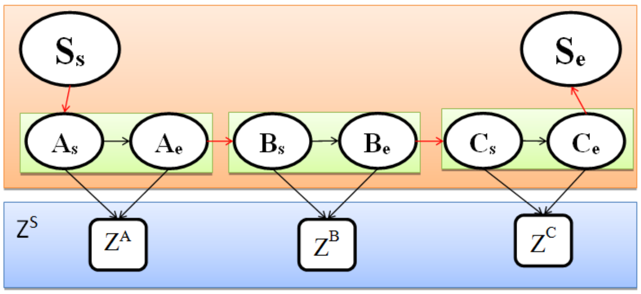

Fig. 3: A composition element represented by a sequential AND of primitive actions

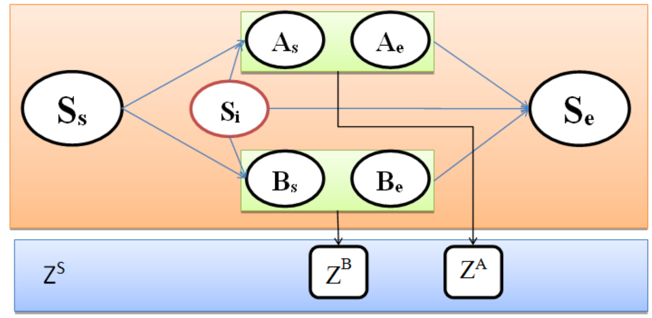

Fig. 4: A task element S represented by A OR B.

D. Branching: OR

Here we describe a more complicated composition: the task S is defined as either subtask A or B (with some prior probability, for example 30%, 70%), denote as the OR-rule: $S \rightarrow A(30\%)|B(70\%)$ where A, B can be primitive actions or further compositions. We call this situation "branching": the task either takes the branch A or branch B.

The network will include the nodes S.s, S.e and the all components in A, B (that can also be described in recursive manner). Two possible cases can happen: S is A, or S is B. In the first cases, we would have S.s = A.s, S.e = A.e, and we need to denote that B, which we write as B.s = B.e = -1. We use $\exists A$ and $!A$ to denote the event A happens (A.s, A.e $>$ 0) or not (A.s = A.e = -1).

A standard approach to realizing "OR" in Bayes network is using a "switching" variable [16]. In this example, it will be S.i, shown in Figure 4.

The timing of S can be presented in term of timing of A and B, such as:

$$P(S.e = \alpha, Z) = W_A P(A.e = \alpha, Z^A | \exists A) + \\ W_B P(B.e = \alpha, Z^B | \exists B)$$

$$\text{where} W_A = P(\exists A) P(Z^B | !B) \text{and} W_B = P(\exists B) P(Z^A | !A)$$

Practically the inference's forward and backward processes are perform on A, B, then the results are combined for S. The weight of A and B depends on 2 factors: the prior probabilities $P(\exists A)$, $P(\exists B)$ and the likelihood $P(Z^x | x.s, x.e)$ for every primitive action x in A and B. For example strong detection of actions in subtask A would make S more likely to be A than to be B.

E. Inference

Combining the AND-rule and OR-rule for composition, one can define more complicated sequence of actions. For example $S \rightarrow (AB|BA)C(D|\emptyset)$ is the task that consists of action A, B in any order, follows by action C, and ends with optional action D.

The message-passing algorithm is used to perform inference. Besides values between 1 and T, the distribution of the timings now also include special value -1, which means the action does not happen. We have a very rich output: the posterior probabilities of all branchings (probabilities of an action happens or not) and the distribtion of when an action starts or ends. From this output, we can also compute probability that the action x is being performed at time step t, for every x and t. Note that inference is performed at every time-step. The network is initialized with all likelihoods initialized to be a prior uniform value using the expected detection score ($D_v[-1,-1]$). As new observations are obtained, the likelihoods are recomputed.

IV. Example Application Description

We first present a human-robot collaborative application we use to motivate our investigation. A human sits at a table across from a robot collaborator who is safely out of reach of the human, but who can move a set of bins both into and out of the reach of the human (Fig. 1). Each bin contains a variable number of Baufix toys, a wooden construction set of screws, nuts, and bolts, which can be used to make small model vehicles and other designs. The bins are kitted so that a number of the bins could be used to construct a few different models.

For the task, the human is instructed to begin building a model from the pieces in the bins,. Their reaches are generally restricted to widthdrawl one part from a bin at a time. Since the human cannot withdraw from a bin not in reach, this imposes a task constraint which the robot must satisfy for the pair to complete the task. When the human needs to reach for a part from a bin not in the workspace, they are instructed to wait until the robot has delivered the bin they need. Based on observations of the human gathered from sensors in the environment, combined with a model of the task, the robot begins delivering bins the human might need. There are only M slots ($M = 3$ for our experiments) in the human's workspace into which the robot can place bins, so eventually the robot must decide to remove unneeded bins and deliver more demanded ones. When more than one construction is possible, the knowledge of which one the human was performing is not made explicit and must be inferred by the activity of the human.

We presented in our previous paper [4] a cost-based planner which optimizes bin delivery and removal timings given the posterior distributions generated by the inference. The planner attempts to minimize expected sum squared wait times, which we use to reduce both total wait time and the maximum wait period. The planning is mostly identical, except that costs are now weighted by the posterior branch probabilities which come from the OR-rules.

V. Evaluation

A. Task Descriptions

We developed a simple, illustrative task to demonstrate the types of behavior our system exhibits in a collaborative assembly scenario. The human attempts to assemble one of two possible toys whose parts are each separated into 4 bins and the robot has no prior knowledge as to which toy the human will be assembling. The two toy models have an identical assembly structure and the base structure, in bin A_1, is the same for both. However, each model is a different color, and all successive parts past the base are in different bins. Thus, the human needs bins B_2, B_3, and B_4 for model B and C_2, C_3, and C_4 for model C. We require that the human perform one reach for each part in the bin and there are total of 14 parts that need to be assembled, 6 in bin A_1, 1 in bin B_2/C_2, 1 in bin B_3/C_3, and 6 in bin B_4/C_4. The bin A_1 is already in the human's workspace when the task starts and takes enough time that both bin B_2 and bin C_2 can be delivered before the human finishes with it. Since there are so few reaches into the second and third bins, the human does not provide the robot with a large amount of information to disambiguate between the two models.

Since the robot cannot determine which model the human is building before reaching into one of the bins in a branch, the robot almost always begins by delivering both B_2 and C_2. Assuming the robot does not remove bins preemptuously, a problem we explored more heavily in our previous work, the best case scenario is when the robot only delivers bins $*_3$ and $*_4$ in the branch the human performing. In the worst case, the same bins are delivered in the branch the robot is not performing, followed by the two in the branch they are.

B. Simulation

We developed a simulator which allowed us to evaluate our planner in a controlled environment. The human agent was programmed to reach towards bins based on random times drawn from our duration model. If a necessary bin was not available in the workspace, the agent would remain stationary and wait.

In order to investigate the behavior of our system in the face of action detection ambiguity, we modifed the detector to demonstrate a false hand position. For each reach, the target position was randomly perterbed from the center of the bin along the row of bins by a value drawn from a Gaussian. Thus, some reaches will mistakenly appear as they are reaching towards the middle of two bins or even into another bin. We use the variance this Gaussian to control the "detector reliability", as higher variances make it more difficult for the robot to disambiguate what action the human is performing. For high reliability, we use $\sigma = 1$ cm, and for low reliability, we use $\sigma = 20$ cm.

Likewise, in the robot's inference model, we alter the parameters of the sensor model to control the "detector confidence". By adjusting the variance weighing factor in the likelihood function, which determines how much to smooth

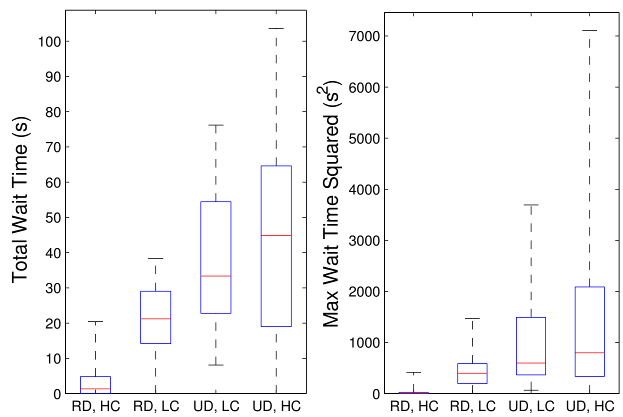

Fig. 5: Distributions of results from a set of $N = 50$ simulated trials for each detector performance/detector confidence combination: **Reliable Detector (RD)**, **Unreliable Detector (UD)**, **High detector Confidence (HC)**, and **Low detector Confidence (LC)**. For each condition, in half the trials the simulated human built model B and half built model C, each without prior knowledge by the robot. As the robot's estimation about its detector reliability better matches the the actual stocastic processes, the human's typical wait time improves, the variance of the wait times is reduced, and its worst-case wait time is limited.

the distance measure to the center of the bin reach location, we are able to expand and contract the area where bin reaches produce responses. Given the bins are about 15 cm apart, we use a high confidence value of $\sigma = 6$ cm and a low confidence value of $\sigma = 10$ cm.

The results of our simulation trials can be found in Fig. 5. Generally, when the robot has confidence in its good detectors (RD, HC), after perceiving a reach into the B_2 bin, it commits by immediately delivering B_3 and B_4. Thus, the human waits very little or not at all.

However, when the robot has confidence in a bad detector (UD, HC), a variety of cases can occur. Sometimes, detections are accurate and the robot acts similarly to the (RD, HC) case. Frequently, the detections are ambiguous, where there is not strong evidence supporting one bin being the target and not the other. In this case, the robot usually plans to deliver both of the next possible bins. Occasionally, a detection occurs which causes the robot to strongly believe the other branch is being built, thus causing a worst-case scenario where the human's next two anticipated bins are entirely wrong.

Having low confidence in a bad detector (UD, LC) generally causes the robot to cover its bases, delivering bins from both branches regardless of which detections are stronger. Having low confidence in a good detector (RD, LC) exhibits similar behavior, except that it is more likely to get a stronger response from the correct bin, so that it delivers the correct bins first. However, it is not nearly as fast as having a properly matched confidence (RD, HC).

C. Real World Example Cases

We also ran experiments with a real human-robot collaborative team. The robot was a 6-DOF Universal Robots UR-10

	Reliable Detect.	Unreliable Detect.
High Confid.	{5.0, 7.9, 20.4}	{12.2, 13.6, 21.0}
Low Confid.	{8.0, 18.1, 25.0}	{16.5, 19.6, 27.3}

Human-Robot Total Wait Times (s)

TABLE I: Sorted total wait times for $N = 3$ real-world trials for each condition. Consistent with Fig. 5, in the reliable detector case, high confidence performs better than low. Furthermore, the reliable cases perform better than the unreliable, for both confidence beliefs. However, more testing is required to confirm that occasionally, an unreliable detector with excessively high confidence should produce a very long waiting time.

mounted to a steel table with a Robotiq C-model parallel jaw gripper. Above the robot, a webcam was mounted to track the positions and orientations of the bins, affixed with Alternate Reality (AR) tags. To the side of the human and bins, a Kinect RGB-D sensor was mounted to sense the behavior of the human. The entire system is calibrated such that the locations of the bins are known with respect to both the robot and the human sensing.

The task the human performed is exactly the same linear task we tested in simulation. To track the human collaborator's hands, we used brightly colored surgical gloves and implemented a color blob tracker on the RGB-D sensor. In typical operation, we call the hand tracking system a reliable detector (RD). To produce an unreliable detector (UD) condition, we alter the RGB-D extrinsic calibration by 6 cm.

The results for each of the trials, sorted by total wait time, can be found in figure I. These results seem consistent with the results obtained in figure 5, but are difficult to verify due to the small sample size.

VI. Discussion and Conclusion

In this paper, we have proposed a significant extension of our previous work which allows us to model multi-path branching in a probabilistic manner. By accomodating the idea that some actions do not need to be performed by the human at all, we can greatly expand the application of our framework to non-deterministic tasks. Since real-world human-robot interactions rarely have a linear chain structure, this development will make the system far more useful. Futhermore, by encoding the task structure that still exists, the robot can continually integrate new information and propogate it forward and backwards in time, always improving its perception of the human's past, current, and future states.

By maintaining densities over multiple branch possibilities, the robot can act in a way that does not require it to commit to a particular branch belief. Thus, the robot is both generally robust to mistakes, since it maintains a non-zero belief that alternate branches are the actual branches and that the absence of detections in the future provides information which can be used to reassess the likelihood of missed detections in the past. Furthermore, it can use its knowledge more effectively to be more or less conservative when it comes to making projections about which actions the human will need the robot to do next. The robot can leverage this information optimize its execution to reduce the number of supporting actions it must perform.

We have performed experiments which seem to confirm that by simply adjusting the confidence in your detectors, you can tune your system to behave more appropriately in the face of uncertainty. As robots are more broadly deployed, having simple, intuitive methods which allow users to correct for percieved mistake-ridden risky behavior, or speed up slow conservative behavior.

In future work, we should perform a more rigiorous real-world evaluation of the system with more trials and novice users. Furthermore, we should show that our model works under a large range of task structures by generating random, more complex tasks.

References

[1] R. Wilcox, S. Nikolaidis, and J. Shah, "Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing," in *Robotics: Science and Systems*, 2012.

[2] H. Goto, J. Miura, and J. Sugiyama, "Human-Robot Collaborative Assembly by On-line Human Action Recognition Based on an FSM Task Model," in *Human-Robot Interaction 2013 Workshop on Collaborative Manipulation*, 2013.

[3] B. Hayes and B. Scassellati, "Challenges in Shared-Environment Human-Robot Collaboration," in *Collaborative Manipulation Workshop at HRI*, 2013.

[4] K. Hawkins, N. Vo, S. Bansal, and A. Bobick, "Probabilistic human action prediction and wait-sensitive planning for responsive human-robot collaboration," in *2013 13th IEEE-RAS International Conference on Humanoid Robots*. Ieee, Oct. 2013.

[5] G. Hoffman and C. Breazeal, "Cost-Based Anticipatory Action Selection for HumanRobot Fluency," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 952–961, Oct. 2007.

[6] ——, "Effects of anticipatory perceptual simulation on practiced human-robot tasks," *Autonomous Robots*, vol. 28, no. 4, pp. 403–423, Dec. 2009.

[7] M. Huber and A. Knoll, "When to assist?-Modelling human behaviour for hybrid assembly systems," in *Robotics (ISR), 2010 41st ...*, 2010, pp. 165–170.

[8] L. Fish, C. Drury, and M. Helander, "Operatorspecific model: An assembly time prediction model," *Human Factors and ...*, vol. 7, no. 3, pp. 211–235, 1997.

[9] M. Tenorth, F. D. Torre, and M. Beetz, "Learning Probability Distributions over Partially-Ordered Human Everyday Activities," in *International Conference on Robotics and Automation*, 2013, pp. 4524–4529.

[10] H. S. Koppula and A. Saxena, "Anticipating Human Activities using Object Affordances for Reactive Robotic Response," in *Robotics: Science and Systems*, Berlin, 2013.

[11] Y. Shi, Y. Huang, D. Minnen, A. Bobick, and I. Essa, "Propagation networks for recognition of partially ordered sequential action," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. Ieee, 2004, pp. 862–869.

[12] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V. Subrahmanian, P. Turaga, and O. Udrea, "A constrained probabilistic petri net framework for human activity detection in video," *Multimedia, IEEE Transactions on*, vol. 10, no. 6, pp. 982–996, 2008.

[13] K. Li, J. Hu, and Y. Fu, "Modeling complex temporal composition of actionlets for activity prediction," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 286–299.

[14] K. Tang, L. Fei-Fei, and D. Koller, "Learning latent temporal structure for complex event detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1250–1257.

[15] Y. A. Ivanov and A. F. Bobick, "Recognition of visual activities and interactions by stochastic parsing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 852–872, 2000.

[16] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)*, 2009.