Bayesian-inferred Flexible Path Generation in Human-Robot Collaborative Networks

William Bentz and Dimitra Panagou

Abstract— **This paper presents a novel method for generating the trajectory of a robot assisting a human in servicing a set of tasks embedded in a convex 2-D domain. This method makes use of Bayesian inference to predict human intent in task selection. Rather than following optimal trajectory towards a single task, the robot computes a set of potentially optimal tasks each weighted by the human's posterior probability and superimposes them into a cost function that is designed to minimize the weighted Euclidean distance relative to set. The effect is a flexible path human-robot collaborative network that is shown in simulation to complete all tasks in a given domain in less time than existing methods for a certain class of highly impulsive humans, i.e., humans that tend to randomly switch tasks at times generated by a Poisson counting process. The algorithm is also illustrated through an experimental demonstration.**

I. Introduction

Over the past decade, the reductions in cost and improved performance of embedded computers has transformed small unmanned aircraft systems (sUAS) from hobbyists' toys into aerial robotic research platforms. Members of the sUAS research community now envision a future in which humans and robots will collaborate in tasks such as: surveillance [1], search and rescue [2], and environmental monitoring [3]. The latter task, as well as inventory keeping, are primary applications of NASA's Astrobee: a free-flying robot designed to assist astronauts in intravehicular activities [4]. These flying assistants serve as a motivating example for this work.

Much of the existing literature on human-robot interaction has focused on human supervisory control, i.e., an operator is required to process information provided by the robot, usually through a computer interface, and then direct the robot's behavior [5]. In these systems, the human is often required to possess detailed knowledge of the robot's various operating modes and transition characteristics in order to properly predict future actions. Automation surprises, due to human fatigue or lack of mode awareness, may have potentially fatal consequences [6], [7].

Human-robotic networks of the future need to operate in a manner that is both intuitive and natural for the human, i.e., free of the information transfer overhead present in supervisory protocols. This will enable a truly collaborative "peer-to-peer" network where humans and robots can operate independently of one another to service various sets of tasks within a given domain [8]. Central to this idea is the requirement that the robot must effectively infer the intent of the human collaborator. The converse problem has been studied at length in [9]–[12], where a typical solution is to design legible motion planning, i.e. visually communicative robot trajectories that are arced or exaggerated to aid in human perception. Inferring human intent, the focus of our work, has typically been approached using tools from Bayesian statistics [13]–[17] which we shall likewise employ.

Many authors have studied human intent inference in the context of single-task collaboration. In [13], the authors consider a human and robot collaborating within a factory setting on the construction of a cart. The robot is required to fetch, return, and use various tools to assemble the cart at a workbench while the human stocks cart components. The authors model the workbench state as a first-order Markov process and demonstrate experimentally that Bayesian inference of human stocking intent offers a more continuous and natural assembly process than their strictly reactive control trials. In [14], the authors use the related technique, Intention-Driven Dynamics, in games such as table tennis.

Our focus lies in the context of multiple-task collaboration. Specifically, we consider a 2-D domain which contains a set of spatially distributed tasks which must each be visited by either the human or the robotic agent. This problem has previously been studied in [16], [17] where the authors have modelled task-allocation under the assumption that humans will behave in a roughly optimal manner. In these works, robots take optimal trajectories towards their most likely target and must recompute the target sequence only when a change in most likely human intent is detected: e.g., the human impulsively changes their target mid-transit. In contrast, we take a flexible path approach, i.e, the possible robot target sequences exist in superposition and inform a cost function that weighs the likelihood of all targets at any given time. In this way, our trajectories tend to curve towards targets as they are continually updated with new posterior probability data. This approach is demonstrated in simulation to have shorter total task completion times than prior approaches for cases involving highly impulsive humans.

This paper is organized as follows: Section II presents our human and robot kinematic models as well as our task domain. Section III describes the Bayesian inference of human goals and presents our method for robot trajectory design. In Sections IV and V, we discuss simulation results and provide an illustrative experimental case study. Section VI presents our conclusions and intended future work.

William Bentz is with the Department of Aerospace Engineering, University of Michigan, Ann Arbor; wbentz@umich.edu.

Dimitra Panagou is with the Department of Aerospace Engineering, University of Michigan, Ann Arbor; dpanagou@umich.edu.

The authors would like to acknowledge the support of an Early Career Faculty grant from NASAs Space Technology Research Grants Program

II. Problem Formulation

Consider a convex domain $\mathcal{D} \in \mathbb{R}^2$ which contains a set $\mathcal{T}$, referred to as the *task queue*, of N spatially distributed tasks denoted $T_i, \forall i \in \{1, ..., N\}$. Each task is located at position p_{T_i} and all positions are measured relative to a global coordinate frame $\mathcal{G}$ with origin $\mathcal{O}$. A heterogeneous network consisting of a single human agent $\mathcal{H}$ and a single robotic agent $\mathcal{R}$ are charged with servicing each of the N tasks. Human and robot positions are denoted $p_{\mathcal{H}}(t_k)$ and $p_{\mathcal{R}}(t_k)$ respectively where $t_k = k\Delta T$, $k \in \mathbb{Z}$, and ΔT is the discrete time step.

Definition 1: T_i is considered complete and is removed from $\mathcal{T}$ the instant that either the human or the robot has passed within a task capture radius of r_c, i.e., $\|p_{\mathcal{H}}(t_k) - p_{T_i}\| \leq r_c \vee \|p_{\mathcal{R}}(t_k) - p_{T_i}\| \leq r_c$.

The discrete time kinematics of the human and robot evolve as:

$$p_{\mathcal{H}}(t_{k+1}) = p_{\mathcal{H}}(t_k) + v_{\mathcal{H}}(t_k)\Delta T, \tag{1a}$$

$$p_{\mathcal{R}}(t_{k+1}) = p_{\mathcal{R}}(t_k) + v_{\mathcal{R}}(t_k)\Delta T, \tag{1b}$$

respectively where $v_{\mathcal{H}}(t_k)$ and $v_{\mathcal{R}}(t_k)$ denote agent velocities which are constrained by a maximum value V, i.e., $v_{\mathcal{H}}(t_k) \leq V$ and $v_{\mathcal{R}}(t_k) \leq V, \forall t_k$. Thus, at time t_k we may constrain the set of reachable points at t_{k+1} such that $p_{\mathcal{H}}(t_{k+1}) \in \bar{B}_{V\Delta T} p_{\mathcal{H}}(t_k)$ and $p_{\mathcal{R}}(t_{k+1}) \in \bar{B}_{V\Delta T} p_{\mathcal{R}}(t_k)$, where $\bar{B}_{V\Delta T} p_{\mathcal{H}}(t_k)$ is the closed ball of radius $V\Delta T$ centered at $p_{\mathcal{H}}(t_k)$. This task domain and agent kinematic model is similar to that of [17] which serves as our baseline technique for comparison.

Modeling human behavior in a collaborative environment is incredibly complex and thus we shall restrict ourselves to three potential models for human task selection. The first is referred to as the *human-optimal sequence*. It is the sequence of tasks that would result in the shortest overall path length for the human assuming straight paths. It does not account for the actions of the robot. The second model is referred to as the *human-greedy sequence* and assumes that the human will always proceed towards their closest target. The third model, *human-collaborative sequence*, assumes that the human and robot will collaborate perfectly and travel at maximum speeds in order to service all targets in the theoretical minimum time. Note that these sequences are recomputed whenever a task is removed from the queue.

Humans are often unpredictable and thus it is necessary to perturb these baseline models in a manner consistent with the seemingly sub-optimal aspects of human behavior. The authors of [17] took an approach, referred to as "human action randomness", in which a human shall nominally walk straight towards their intended target with probability $\alpha \in (0, 1)$ of taking a random step in transit. The intent is to capture the indecisiveness of human nature; however, we postulate that a more realistic model is that of a homogeneous Poisson process. This generates events, at random arrival times, in which the human impulsively "changes their mind" as to their intended target. They select a new one at random while always proceeding along straight paths. We shall refer to this as our *impulsivity perturbation model*. Thus, upon every reduction event in the task queue, the human shall select a target, consistent with their task selection model, and proceed towards it with some probability of an impulsivity event occurring in transit.

In a homogeneous Poisson process, the interarrival times are independent identically distributed (i.i.d) exponential random variables with mean λ^{-1} and a probability density function (pdf) $f(x) = \lambda e^{-\lambda x}$ where $x \in [0, \infty)$. We shall refer to λ as our *human impulsivity factor*. Using a cumulative distribution inverse method, we define impulsivity event times with a discrete approximation of the well known method presented in [18].

Definition 2: Impulsivity event ϵ_{n+1}, occurs at a time:

$$t_{\epsilon_{n+1}} = t_{\epsilon_n} - \Delta T \left\lceil \frac{\lambda^{-1} \ln(U_{n+1})}{\Delta T} \right\rceil, \tag{2}$$

where U_{n+1} is a uniformly distributed random variable in $(0, 1)$ and the upper-bracketed delimiters represent the ceiling function. Note that the initial event ϵ_1 occurs at $t_{\epsilon_1} = -\Delta T \left\lceil \frac{\lambda^{-1} \ln(U_1)}{\Delta T} \right\rceil$.

Remark 1: The purpose of this work is to design a flexible robot trajectory generation method that facilitates shorter total task completion times than existing methods for highly impulsive humans, i.e., $\lambda >> 1$.

III. Approach

A. Human Intent Inference

To predict human intent, we use the standard Bayesian inference model presented in [15] and later adopted in [16], [17]. Essentially, the robot observes the translation of the human at each time step and computes the probability that any given task in $\mathcal{T}$ is the intended target of the human. The posterior probability for T_i evolves as:

$$P\big(T_i | p_{\mathcal{H}}(t_{1:k+1})\big) \propto P\big(p_{\mathcal{H}}(t_{k+1}) | T_i, p_{\mathcal{H}}(t_{1:k})\big) P\big(T_i | p_{\mathcal{H}}(t_{1:k})\big). \tag{3}$$

Assuming a memoryless system, we make the simplification that our likelihood function obeys $P\big(p_{\mathcal{H}}(t_{k+1}) | T_i, p_{\mathcal{H}}(t_{1:k})\big) = P\big(p_{\mathcal{H}}(t_{k+1}) | T_i, p_{\mathcal{H}}(t_k)\big)$.

We model the relationship between human action and intent following a standard Boltzmann, i.e., softmax, policy:

$$P\big(p_{\mathcal{H}}(t_{k+1}) | T_i, p_{\mathcal{H}}(t_k)\big) \propto e^{\beta Q_i(p_{\mathcal{H}}(t_{k+1}))}, \tag{4}$$

where β is the adjustable rationality index characterizing the tendency of the human to take an optimal path towards their intended task. This should be distinguished from λ which characterizes the tendency of a human to switch tasks. For $\beta \to \infty$, the robot assumes that the human will be perfectly rational and take straight paths while $\beta \to 0$ implies the robot will assume the human follows a random walk.

$Q_i(p_{\mathcal{H}}(t_k))$ is the Q-value for the action taken by the human over $p_{\mathcal{H}}(t_{k:k+1})$, i.e., the expected discounted reward of stepping from $p_{\mathcal{H}}(t_k)$ to $p_{\mathcal{H}}(t_{k+1})$ and then proceeding optimally towards T_i. Our robots assume that a rational human would take steps that carry them closer to their target

along a minimum path length. Thus, $Q_i\left(p_{\mathcal{H}}\left(t_k\right)\right)$ is the value function of a γ-discounted optimal control problem:

$$Q_i\left(p_{\mathcal{H}}\left(t_k\right)\right) = R\gamma^{\|p_{T_i} - p_{\mathcal{H}}(t_k)\|} - C\frac{\gamma - \gamma^{\|p_{T_i} - p_{\mathcal{H}}(t_k)\|}}{1-\gamma}, \tag{5}$$

where R is a terminal reward and C is a running cost. For additional descriptions of reinforcement and Q-learning, see [19].

B. Robot Task Superposition

Assume for a moment that at time t_k, T_i is the current intended task of $\mathcal{H}$ at $p_{\mathcal{H}}\left(t_k\right)$ and $\mathcal{R}$ is positioned at $p_{\mathcal{R}}\left(t_k\right)$. Furthermore, let us assume that both the robot and human shall proceed along straight paths at maximum velocity towards each intended task. Then, there exists an optimal sequence of tasks for both $\mathcal{H}$ and $\mathcal{R}$ from these initial conditions. Define T_j^ℓ, $\forall j \in \{1, ..., N\}$ and $\ell \in \{1, ..., (N-1)!\}$ | $T_1^\ell = T_i$. T_j^ℓ is the j^{th} task in the ℓ^{th} permutation of possible network task sequences beginning with T_i. Also define m as the last task in the human's sequence and $m+1$ as the first task in the robot's sequence. Note that the human and robot will complete tasks simultaneously.

The robot must find the indices $\bar{\ell}$ and $\bar{m}$ that correspond to the task sequence and agent task allocation that minimizes the completion time for the network, i.e, that minimize the maximum path length of either agent:

$$\begin{aligned}\{\bar{\ell},\bar{m}\} = \underset{\substack{\ell\in\{1,\ldots,(N-1)!\},\\ m\in\{1,\ldots,N-1\}}}{\operatorname{argmin}} \max\Big(&\sum_{j=1}^{m}\|p_{T_j^\ell} - p_{T_{j-1}^\ell}\|\left(1-\delta_{j,1}\right) \\ &+ \|p_{\mathcal{H}}\left(t_k\right) - p_{T_1^\ell}\|\delta_{j,1}, \sum_{j=m+1}^{N}\|p_{T_j^\ell} - p_{T_{j-1}^\ell}\|\left(1-\delta_{j,m+1}\right) \\ &+ \|p_{\mathcal{R}}\left(t_k\right) - p_{T_{m+1}^\ell}\|\delta_{j,m+1}\Big),\end{aligned} \tag{6}$$

where $\delta_{j,1}$ is the Kronecker delta function. Note that (6) may be solved in real-time with a brute-force method for approximately $N \leq 10$. For networks with $N > 10$, one may employ mixed integer linear programming such as in [20].

Thus for each T_i, our robot should choose $T_{\bar{m}+1}^{\bar{\ell}}$ with probability $P\big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\big)$. We superimpose the potential choices of $T_{\bar{m}+1}^{\bar{\ell}}$, weighted by both proximity and $P\big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\big)$, into a cost function that selects the robot's next translation step:

$$\begin{aligned}p_{\mathcal{R}}\left(t_{k+1}\right) = \underset{p\in\bar{B}_{v_{\Delta}r}p_{\mathcal{R}}(t_k)}{\operatorname{argmin}} \sum_{i=1}^{N}\|p - p_{T_{\bar{m}+1}^{\bar{\ell}}}\|\Bigg(&P\Big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\Big) \\ &- \min\Bigg(0,\ln\Big(\frac{\|p_{\mathcal{R}}\left(t_k\right) - p_{T_{\bar{m}+1}^{\bar{\ell}}}\|}{r_a}\Big)\Bigg)\Bigg),\end{aligned} \tag{7}$$

where r_a is a tunable parameter referred to as the task attractive radius. Note that $T_{\bar{m}+1}^{\bar{\ell}}$ is different for each index i in the summation. This is implicit as to not further complicate the notation. The generation of this trajectory is summarized in Algorithm 1.

The effect of (7) is that a robot, with distance greater than r_a from any task, will travel along a trajectory purely informed by the probabilities that each $T_{\bar{m}+1}^{\bar{\ell}}$ is its optimal task. If $\mathcal{R}$ passes within r_a of an adjacent task during its travel, the logarithmic term in (7) will output a negative value and thus the task's weight $P\Big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\Big)$ shall be supplemented with an additional positive value. This results in a logarithmic attraction to adjacent tasks that $\mathcal{R}$ passes while approaching tasks with high values for $P\Big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\Big)$.

Trajectories resulting from (7) are less direct and tend to curve towards tasks as $P\Big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\Big) \to 1$ rather than taking straight paths. This offers more flexibility in task selection than existing methods, such as [17], that direct the robot along straights paths towards the task which is optimal with probability greater than 0.5. We postulate that our method shall perform better for highly impulsive humans, $\lambda >> 1$, than the existing strategy as we are able to avoid sharp and sudden changes in the robot's trajectory that arise during the discrete event in which one task replaces another as having probability greater than 0.5. Furthermore, the existing strategy suffers in that it is often hard to identify any task as having a probability greater than 0.5 when there are a substantial number of tasks in the network. The authors of [17] address this problem by only recalculating the robots optimal sequence when a new task has risen above the probability threshold. Essentially, the human intent inference data is not utilized between these recalculation events.

IV. Simulation Results

Our hypothesis, that this method will offer shorter total task completion times than [17] for humans with high values of λ, has been verified through a series of simulations. In total, 1000 unique task domains were generated each with uniformly distributed random values for $p_{\mathcal{H}}\left(t_1\right)$, $p_{\mathcal{R}}\left(t_1\right)$ and p_{T_i}, $\forall i \in \{1, ..., N\}$. Simulations take place in a 100×100 unit domain with $N = 10$ tasks. $R = 10$, $C = 1$, and $\gamma = 0.9$ were chosen as our Q-value parameters as they typically resulted in a smooth and gradual growth in $P\Big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\Big)$ that would exceed 0.9 immediately before the human entered the capture radius r_c of a task.

$V = 2\sqrt{2}$ was chosen as the maximum velocity because $Q_i\left(p_{\mathcal{H}}\left(t_{k+1}\right)\right)$ is computed only at discrete grid points in our simulation. These points intersect the reachable set of each agent and thus $V = 2\sqrt{2}$ implies that $Q_i\left(p_{\mathcal{H}}\left(t_{k+1}\right)\right)$ is computed over a geometrically symmetric 6×6 square of points. It was observed that choosing values for V that are not multiples of $\sqrt{2}$ tended to prevent diagonal movement. $r_c = \frac{1}{2}V + 1$ was chosen as it was the smallest value to reliably guarantee task capture without any possibility of infinite oscillations around the task location. Finally, $r_a = 40$ was chosen empirically as it was observed to guarantee capture of any adjacent task within approximately 10 units of

Algorithm 1 Compute Robot Trajectory

Inputs: $N, \Delta T, \beta, r_c, r_a, \mathcal{T}$=$\{T_i, p_{T_i}\ \forall i \{1, ..., N\}\}$, $p_{\mathcal{H}}(t_k), p_{\mathcal{R}}(t_0), V, R, C, \gamma$
Initialize: $k \leftarrow 0$
$P(T_i|p_{\mathcal{H}}(t_k)) \leftarrow \frac{\|p_{\mathcal{H}}(t_k)-p_{T_i}\|^{-1}}{\sum_{i=1}^{N}\|p_{\mathcal{H}}(t_k)-p_{T_i}\|^{-1}}$, $\forall i \in \{1, ..., N\}$
while $N > 0$ **do**
  while 1 **do**
    while $i < N + 1$ **do**
      Compute $Q_i(p_{\mathcal{H}}(t_{k+1}))$ via (5).
      $P(p_{\mathcal{H}}(t_{k+1})|T_i, p_{\mathcal{H}}(t_k)) \leftarrow P(T_i|p_{\mathcal{H}}(t_k))e^{\beta Q_i(p_{\mathcal{H}}(t_{k+1}))}$
      $T_j^{\ell} \leftarrow permutations(\{1, ..., N\}) \mid T_1^{\ell} = T_i$
      Compute $\{\bar{\ell}, \bar{m}\}$ via (6).
      $i \leftarrow i + 1$
    end while
    Normalize $P(p_{\mathcal{H}}(t_{k+1})|T_i, p_{\mathcal{H}}(t_k))$, $\forall i$.
    Compute $p_{\mathcal{R}}(t_{k+1})$ via (7).
    $k \leftarrow k + 1$
    if $\exists j \mid \|p_{\mathcal{R}}(t_k) - p_{T_j}\| < r_c \vee \|p_{\mathcal{H}}(t_k) - p_{T_j}\| < r_c$ **then**
      $\mathcal{T} \leftarrow \mathcal{T} \backslash T_j$, p_{T_j}
      $N \leftarrow N - 1$
      Break
    end if
  end while
end while

$\mathcal{R}$. We consider this to be a reasonable detour in a trajectory proceeding towards tasks with high $P\left(T_i|p_{\mathcal{H}}(t_{1:k+1})\right)$.

Each of our 1000 unique task domains was tested in simulations with values for λ set at integer steps between 0 and 24 and under all three human behavior models described in Section II. This resulted in a total of 150,000 individual simulations. The average total task completion time over the 1000 task domains is presented in Fig. 1. As one would assume, the time required to complete all tasks was highly dependent upon the spatial arrangement of tasks in the domain. Therefore, the standard deviations associated with the average values in Fig. 1 tended to be very large, e.g., approximately 10 seconds for trials with small λ and upwards of 30 seconds for trials with large λ. These standard deviations are presented in Fig. 2.

In order to prevent domains containing highly distributed tasks from holding an unfair weight in these averages, a set of corrected trends were produced for Fig. 3. For these, each simulation task completion time is normalized against the theoretical minimum time (control robot with $\lambda = 0$ in human-collaborative task sequence) for that given map before the averaging occurs. This was done to properly scale the relative performance for each unique map. However, this only affected the percent change in performance by at most a few fractions of a percent.

For all three human models, the control robot performs better for small values of λ. As λ increases a trade-off occurs at which point the flexible path robot shows better performance. This trade-off occurs for λ in an approxi-

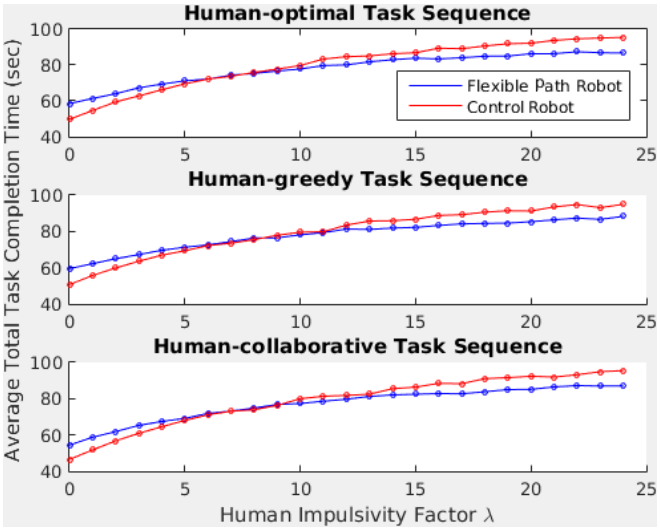

Fig. 1. 1000 task domains were generated with simulations performed on successive integer values of $\lambda \in [0, 24]$ for three human behavior models. Each point is an average total task completion time over all task domains.

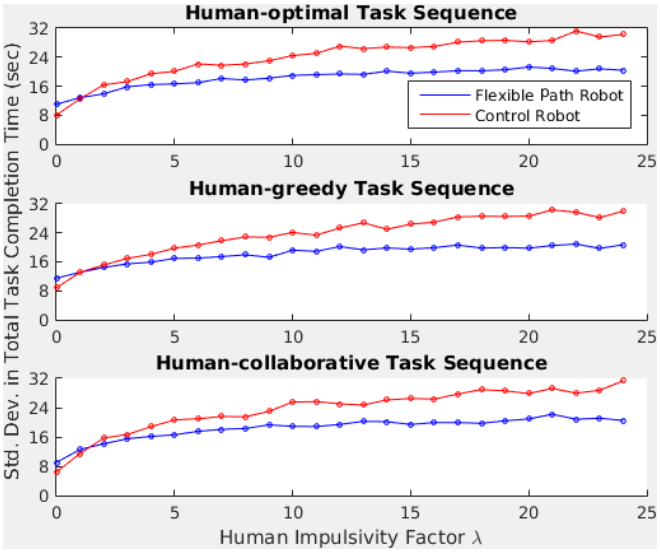

Fig. 2. These standard deviations are computed in reference to the average total task completion times in Fig. 1.

mate range of $6 - 8$ for the human-optimal task sequence, $6 - 11$ for the human-greedy task sequence, and $7 - 9$ for the human-collaborative task sequence. For $\lambda = 0$, the improvements in total task completion time for our three human models are -17.7% -17.3% and -17.1% respectively while $\lambda = 24$ they are 9.2%, 7.0% and 8.9%. These results are presented in Fig. 4. The trajectories from a sample trial of the human-optimal task sequence with $\lambda = 13$ are presented in Fig. 5 which illustrates a direct comparison of our method with the existing one. For the full animation of this figure, see the supplemental video available at `https://www.dropbox.com/s/q59l0p77ls0h5yq/HRI_FEB_2018.mp4?dl=0`.

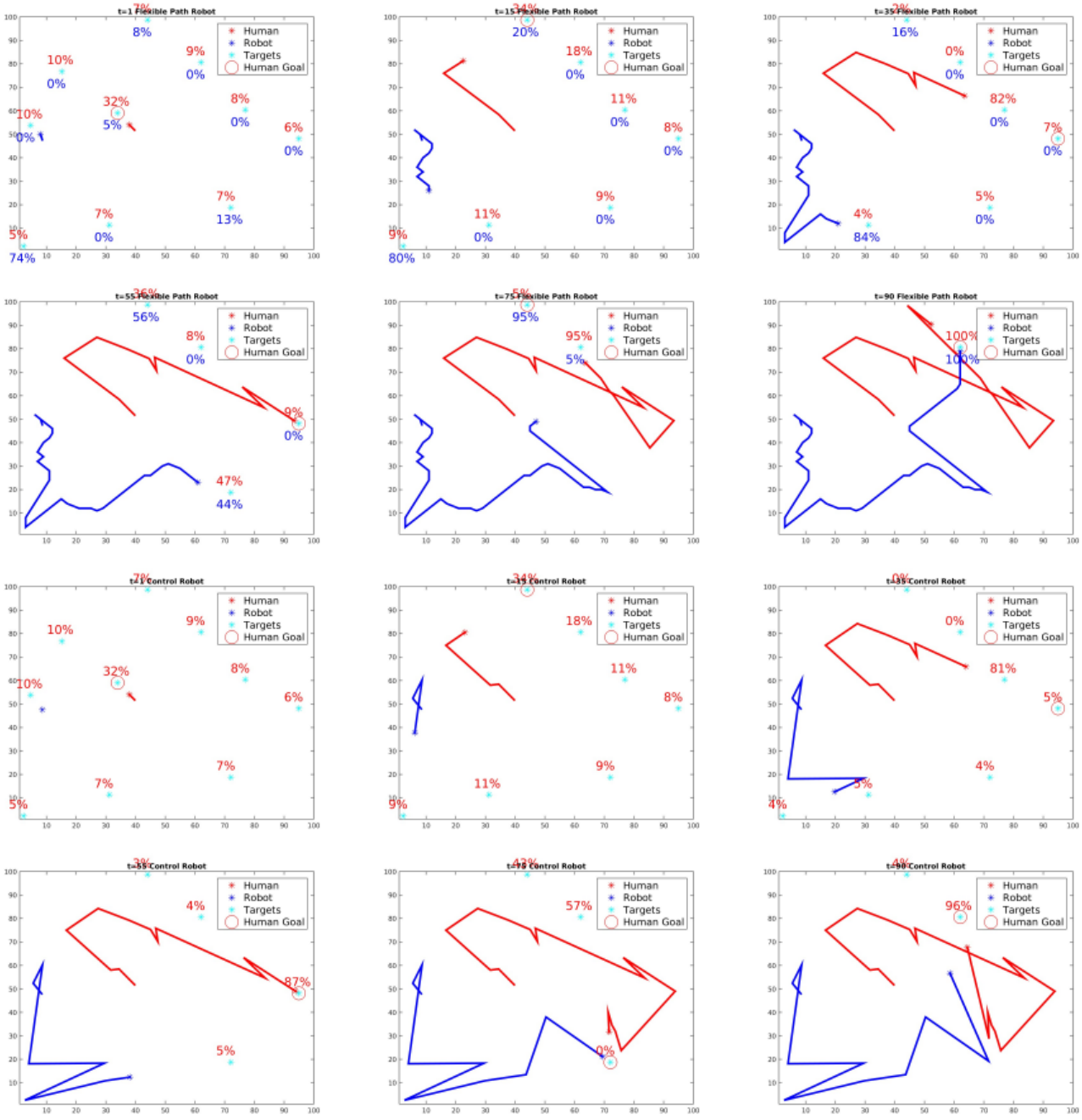

Fig. 5. A sample trial illustrates a direct comparison of the flexible path robot (FPR) with the control robot (CR). The upper six and lower six images are time lapses of the FPR and CR simulations respectively for $\lambda = 13$ and for the same set of initial conditions and impulsivity event times t_{ϵ_n}. At $t = 1$, the FPR has already started towards its closest task while the CR remains stationary as it waits for event $P\Big(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\Big) > 0.5$ to trigger the initial robot sequence computation. While $1 < t < 15$, the robot completes this closest task in both simulations. The CR then proceeds north east towards the same task as the human as it is still operating under its initial sequence computation. It is not until a probability event triggers a recomputation that the CR takes a south west turn. Meanwhile, the FPR had already begun proceeding in a generally southward direction. While $15 < t < 35$, the human impulsively switches targets from the northernmost target to the easternmost target. This eventually triggers a sharp deviation in the CR's path away from the southernmost target while the FPR is unaffected as it is already well within r_a of the southernmost task. By $t = 55$, the CR is well behind the FPR.

V. Experimental Case Study

Our algorithm has also been tested experimentally within an indoor test environment with five tasks placed within a $5\,m \times 2.5\,m$ domain. Our robotic agent is an Ascending Technologies Hummingbird quadrotor which is controlled via an XBee wireless serial module from our ground desktop computer. Position data of both the human the quadrotor are made available to our algorithm at 250 Hz via a Vicon motion

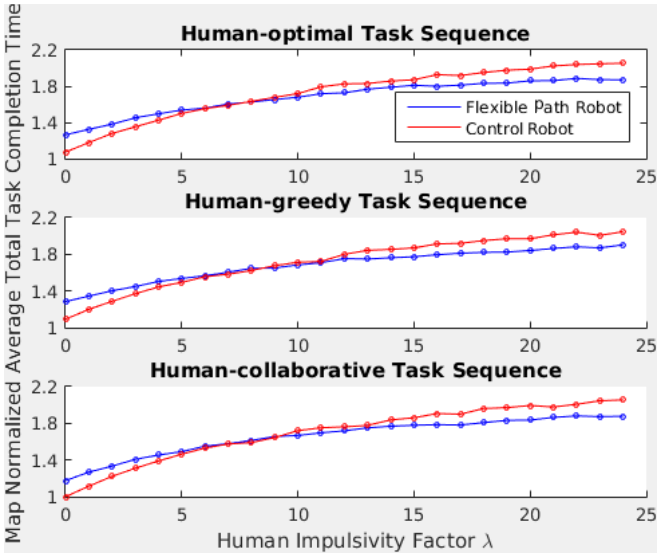

Fig. 3. For these corrected trends, each simulation task completion time is normalized against the theoretical minimum time for that task domain before averaging.

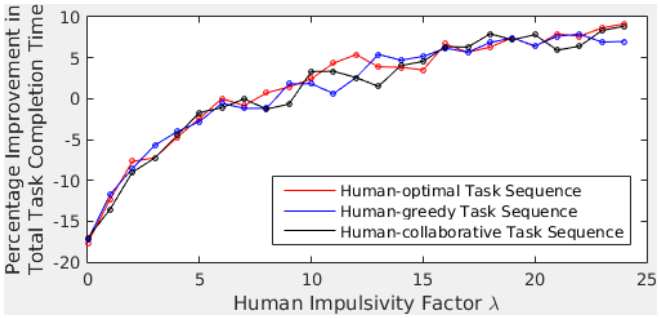

Fig. 4. The flexible path robot shows a definite improvement in total task completion time across all three human behavior models for $\lambda > 11$.

capture system. The algorithm is implemented in MATLAB on the ground computer which interfaces with the vehicle via the Robot Operating System (ROS) and the MATLAB Robotics Systems Toolbox.

In the experiment, the human initially proceeded towards task 1 while the robot was drawn towards tasks 3-5 due to a combination of their attractivity terms and the high probability that task 5 was the optimal robot task. Upon detecting a shift in the human's intent towards tasks 3-5, the robot's trajectory curved back towards tasks 1 and 2. This is illustrated in Fig. 6 and the supplemental video.

One noticeable difference between the experiment and simulation is that $P\left(T_i|p_{\mathcal{H}}\left(t_{1:k+1}\right)\right)$ showed much greater oscillations as the human approached intended targets. This may be a result of the fact that the simulated human proceeds at a constant velocity while the actual human's trajectory is noisy. This may result in a noisy evolution of $Q_i\left(p_{\mathcal{H}}\left(t_k\right)\right)$. Our immediate future experimental goals are to implement a finer mesh task grid and to filter the human's trajectory.

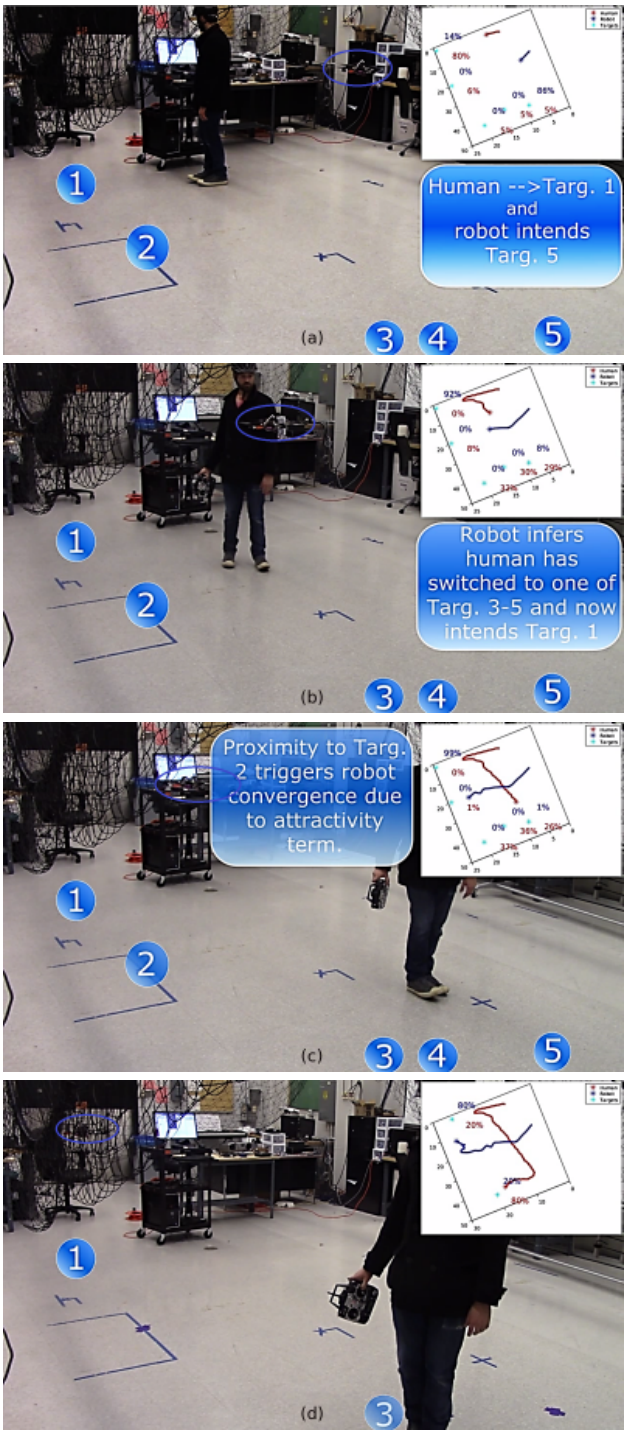

Fig. 6. In our experimental trial, the human initially proceeds towards task 1 and then impulsively switches to task 5. This prompts the robot to switch to tasks 1 and 2.

VI. Conclusions

In this paper, we presented a novel method for generating the trajectory of a robot assisting a human in servicing a set of tasks in a convex 2-D domain. Like existing methods, we used Bayesian inference to estimate the human's intent which informed the computation of potentially optimal task sequences for the robot. However, our approach is novel in that all of these potential task sequences are superimposed into a cost function weighted by the posterior probability of the human's intent with an additional consideration for task proximity. This results in flexible robot path that increases the performance of the network for a class of highly impulsive humans, i.e., typically humans with $\lambda > 11$. We verified these results through 150,000 simulations of various domain maps and agent initial conditions as well as values for λ.

Furthermore, we demonstrated the efficacy of this algorithm experimentally. Qualitatively, the experimental agent performed well; however, the posterior probability tended to oscillate much more rapidly than in the simulated environment. We are considering potential solutions to this problem, e.g, filtering position measurements and using a higher resolution spatial discretization. Future work will extend this network to multiple humans and robots and formalize a hybrid automaton with modes dedicated to additional assistive tasks other than point-to-point servicing.

References

[1] V. Srivastava, "Stochastic search and surveillance strategies for mixed human-robot teams," Ph.D. dissertation, University of California Santa Barbara, 2012. [Online]. Available: http://www.egr.msu.edu/~vaibhav/papers/2012e-s.pdf

[2] L. Marconi, C. Melchiorri, M. Beetz, D. Pangercic, R. Siegwart, S. Leutenegger, R. Carloni, S. Stramigioli, H. Bruyninckx, P. Doherty, A. Kleiner, V. Lippiello, A. Finzi, B. Siciliano, A. Sala, and N. Tomatis, "The sherpa project: Smart collaboration between humans and ground-aerial robots for improving rescuing activities in alpine environments," in *2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Nov 2012, pp. 1–4.

[3] E. S. Berman, M. Fladeland, J. Liem, R. Kolyer, and M. Gupta, "Greenhouse gas analyzer for measurements of carbon dioxide, methane, and water vapor aboard an unmanned aerial vehicle," *Sensors and Actuators B: Chemical*, vol. 169, pp. 128 – 135, 2012.

[4] M. Bualat, J. Barlow, T. Fong, C. Provencher, T. Smith, and A. Zuniga, "Astrobee: Developing a free-flying robot for the international space station," in *AIAA SPACE 2015 Conference and Exposition*, Pasadena, CA, 2015.

[5] J. R. Peters, V. Srivastava, G. S. Taylor, A. Surana, M. P. Eckstein, and F. Bullo, "Human supervisory control of robotic teams: Integrating cognitive modeling with engineering design," *IEEE Control Systems*, vol. 35, no. 6, pp. 57–80, Dec 2015.

[6] N. B. Sarter and D. D. Woods, "Team play with a powerful and independent agent: Operational experiences and automation surprises on the airbus a-320," *Human Factors*, vol. 39, no. 4, pp. 553–569, 1997.

[7] T. Shanker and M. Richtel, "In new military, data overload can be deadly," *The New York Times*, 2011. [Online]. Available: http://www.nytimes.com/2011/01/17/technology/17brain.html?pagewanted=all.

[8] C. K. L. F. J. S. R. A. R. B. R. S. L. H. A. S. J. G. T. M. B. Terrence Fong, Illah Nourbakhsh and J. Scholtz, "The peer-to-peer human-robot interaction project," in *SPACE 2005*, Long Beach, CA, 2005.

[9] D. Szafir, B. Mutlu, and T. Fong, "Communication of intent in assistive free flyers," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 358–365. [Online]. Available: http://doi.acm.org/10.1145/2559636.2559672

[10] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, 2015, pp. 51–58. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696473

[11] J. F. Fisac, C. Liu, J. B. Hamrick, S. Sastry, J. K. Hedrick, T. L. Griffiths, and A. D. Dragan, "Generating plans that predict themselves," in *Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2016.

[12] Y. Zhang, S. Sreedharan, A. Kulkarni, T. Chakraborti, H. H. Zhuo, and S. Kambhampati, "Plan explicability and predictability for robot task planning," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 1313–1320.

[13] G. Hoffman and C. Breazeal, "Effects of anticipatory action on human-robot teamwork: Efficiency, fluency, and perception of team," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2007, pp. 1–8.

[14] Z. Wang, M. Deisenroth, H. Amor, D. Vogt, B. Schlkopf, and J. Peters, *Probabilistic modeling of human movements for intention inference*. MIT Press Journals, 2013, vol. 8, pp. 433–440.

[15] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329 – 349, 2009.

[16] C. Liu, J. B. Hamrick, J. F. Fisac, A. D. Dragan, J. K. Hedrick, S. S. Sastry, and T. L. Griffiths, "Goal inference improves objective and perceived performance in human-robot collaboration," in *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems*, Richland, SC, 2016, pp. 940–948.

[17] C. Liu, S.-Y. Liu, E. L. Carano, and J. K. Hedrick, "A framework for autonomous vehicles with goal inference and task allocation capabilities to support peer collaboration with human agents," in *Proceedings of the ASME 2014 Dynamic Systems and Control Conference*, San Antonio, TX, 2014.

[18] R. Pasupathy, *Generating Nonhomogeneous Poisson Processes*. John Wiley Sons, Inc., 2010.

[19] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge, 1989.

[20] C. Schumacher, P. Chandler, M. Pachter, and L. Pachter, "Constrained optimization for UAV task assignment," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2004.