Enabling Robust Human-Robot Cooperation through Flexible Fully Bayesian Shared Sensing

Nisar Ahmed
University of Colorado Boulder
Department of Aerospace Engineering Sciences
Boulder, CO 80309
Nisar.Ahmed@colorado.edu

Rina Tse and Mark Campbell
Cornell University
Department of Mechanical and Aerospace Engineering
Ithaca, NY 14853
{rt297,mc288}@cornell.edu

Abstract

Cooperative human-robot sensing has the potential to overcome many hurdles for networked field robotic applications, especially those with significant sensing, computational or communication constraints that make teleoperation or direct supervisory control difficult. The main idea is to augment robotic sensing horizons with the diverse and complementary capabilities of ‘human sensors’ for solving difficult hybrid nonlinear state estimation and perception problems, although robots may still plan and control their own actions autonomously. In this work, we examine two related issues through the use of sketch-based and semantic codebook perceptual interfaces for probabilistic target search problems: (i) how should autonomous machines/robots assess the trustworthiness of human sensors?; (ii) what strategies can be used to generate human-machine ‘dialog’ about complex uncertainties in random variables, without cognitively overburdening humans or undermining human trust in automation?

1 Introduction

Robustness can be generally characterized by the ability of an intelligent autonomous agent (machine, robot, human) to appropriately reason about and respond to uncertainty, which typically refers to the physical process/sensor noise, unknown model parameters, ambiguity in task requirements, etc. that collectively affect an agent’s local planning and perception processes. Robustness can also be understood in the context of mixed human-machine/human-robot systems, in which intelligent agents interact with each other to solve complex problems more effectively than any one agent can solve on its own. Such interactions lead to other more subtle uncertainties that make robust intelligent cooperation particularly challenging; these can be broadly characterized as either systemic perception uncertainties (due to unknown pedigree and accuracy of shared information) or systemic planning uncertainties (due to ignorance of the capabilities, preferences and intentions of other agents) (Sheridan 2002). Both local and systemic uncertainties are important to understanding the relationship between trust and robustness in mixed systems, and are the main focus of efforts to improve human-robot team interactions through cooperative intelligence.

Cooperative intelligence is based on the idea that human intelligence is a valuable, though constrained and imperfect, resource that complements (constrained and imperfect) machine intelligence. In other words, intelligent robots must not only know how to best exploit human reasoning to help accomplish tasks, but must also help support and avoid over-taxing human reasoning as much as possible. One exemplary study of this idea for single human-single robot teams is given by (Fong, Thorpe, and Baur 2003), in which a small autonomous mobile robot queries novice/expert human supervisors to help it trouble-shoot various pre-programmed perception and planning ‘faults’ it encounters. A more sophisticated variant of this idea is presented in (Kaupp and Makarenko 2008), which proposes a formal decision theoretic strategy for requesting human observations or control inputs using a probabilistic model of uncertainty and utility measures for interacting with novice/expert humans. These studies underscore the key role uncertainty plays in shaping trust and robust intelligence in mixed human-robot systems. In particular, the probabilistic approach proposed by (Kaupp and Makarenko 2008) highlights the difficulty of ensuring that human inputs are reliable and useful enough to be consumed by automation, even in a cooperative setting. Likewise, the results reported by (Fong, Thorpe, and Baur 2003) emphasize the importance of developing informative ergonomic human-robot interfaces that prevent (expert and non-expert) humans either from taking automation for granted (e.g. ignoring false alarms, or assuming that ‘the machine is always right’) or from becoming overwhelmed with supervisory tasks (which can make automation more burdensome than beneficial).

However, these previous studies were restricted to single human-single robot scenarios, so it is not clear how these approaches exactly scale to larger mixed teams of multiple humans and robots. As noted in (Lewis et al. 2009) in the context of urban search and rescue missions, issues of trust and robustness become even more complex to assess in large teams due to increased task complexity and the larger possible space of human-robot, human-human, and robot-robot interactions. Furthermore, humans in the aforementioned studies could only provide basic low-level commands or high-level observations for aiding in relatively simple
navigation or object classification problems. As such, the potential range and flexibility of human reasoning has not yet been fully explored for cooperatively solving complex cooperative perception and planning problems for multi-robot systems in unstructured uncertain environments. Some inroads have been made relatively recently within the robotics community, although work using probabilistic AI has primarily focused on interpretation of multimodal command inputs for human-assisted robotic planning and navigation, e.g., via natural language speech (Tellex et al. 2011), hand-drawn sketches (Shah and Campbell), or physical gestures (Bauer et al. 2009). However, these works do not consider cooperative perception tasks or multi-human/multi-robot interaction scenarios where information gathering is the primary objective. Moreover, they do not address the issues of determining the trustworthiness of human-generated observations and of relating complex general beliefs from probabilistic models back into a ‘human understandable’ form.

To address these issues, a unified probabilistic framework for cooperative perception with general mixed human-robot information sources must be developed. We argue that, as human-robot systems become more sophisticated, formal statistical methods for integrating robotic and human-generated information become essential for establishing quantitatively verifiable measures of agent trustworthiness, and are thus key ingredients for designing robust intelligent human-robot systems. To this end, we are exploring how dynamic hierarchical Bayesian estimation and learning methods can be used to incorporate formal statistical notions of agent trustworthiness into ergonomic interfaces for cooperative human-robot perception in large teams. Figure 1 illustrates the cooperative human-robot information exchange process that we ultimately envision, which is discussed here in the context of static target search applications. In this work, we first show how fully Bayesian learning techniques can be used to extract information from and account for uncertainties in probabilistic ‘human sensor’ model parameters for a set of real multi-human search experiments designed around flexible free-form sketch interfaces. We then later discuss the potential use of ‘flipped’ semantic human sensor models to enable human-robot dialog about complex probabilistic beliefs, e.g., which may be represented by multimodal distributions such as Gaussian mixtures or hybrid probabilistic models in the case of target search applications.

2 Background

Refs. (Kaupp et al. 2005) and (Bourgault et al. 2008) showed that robots can make good use of human sensors for augmented physical perception if the Bayesian paradigm is adopted to formally characterize and fuse human observations with independent robotic sensor data. However, these works place strong limitations on the types of perceptual inputs that can be provided by humans. For instance, (Kaupp et al. 2005) assumes that humans provide numerical range and bearing measurement data for target localization (‘The object is at range 10 m and bearing 45 degrees’), while (Bourgault et al. 2008) assumes that humans provide binary ‘detection/no detection’ visual observations for target search.

Our recent work using hybrid Bayesian modeling and dynamic state estimation with vague semantic human sensor data in (Ahmed, Sample, and Campbell 2013) exploits the fact that humans are capable of providing a much broader range of information in such scenarios using natural language information (‘The target is nearby in front of me’ or ‘Nothing is next to the truck moving North’). Although more flexible and user-friendly, such ‘soft’ human sensor data present significant challenges for cooperative Bayesian perception, since semantic human sensors cannot be modeled solely from physical first principles and are quite sensitive to contextual as well as cognitive factors (Hall and Jordan 2010). In (Ahmed, Sample, and Campbell 2013), we showed that likelihood functions for semantic human observations of continuous physical states can generally be modeled via probabilistic discriminative classifiers, whose parameters can be rigorously learned from real data via maximum likelihood or Bayesian methods (Fig. 2).

While such models can gauge the overall precision of information provided by human sensors, they do not give a complete picture of their accuracy, i.e., is it possible that a human sensor is partially or entirely mistaken in her observations? If so, is it possible for the automation to detect and quantify this in a reliable and rigorous way? Can any such procedure be scaled to scenarios where multiple potentially unreliable human sensors are sharing information simultaneously, and can the automation even learn to extract some useful information from both reliable and unreliable human sensors for various perception tasks? In Sec. 3 show how these questions can be addressed through Bayesian modeling and inference techniques for information fusion that account for hierarchical uncertainties in human sensor models.

The human-robot interfaces developed for (Ahmed, Sample, and Campbell 2013) also allow human collaborators to directly ‘see what the robot thinks’ about its search environment by displaying a complex, continually updated probability distribution on a GUI. Since most human collaborators are largely not expected to be experts in estimation theory, human-machine dialog could instead be used to directly share such information and foster reliable human perceptual input without overburdening humans or undermining human.
as the ones shown in Fig. 3 could help address these issues.

cally ‘flipping around’ semantic human sensor models such as

To my left’. In Sec. 4, we will also discuss how probabilistic con-

an 10% chance it is behind that wall 3 feet away. I am 90% cer-

am 90% certain the target is 100 yards from my current lo-

trust in automation, e.g. by explaining important yet subtle
details of state uncertainties to human collaborators (e.g. ‘I am 90% certain the target is 100 yards from my current location, but there is a 10% chance it is behind that wall 3 feet away’). In Sec. 4, we will also discuss how probabilistically ‘flipping around’ semantic human sensor models such as the ones shown in Fig. 3 could help address these issues.

3 Fully Bayesian Sketch-based Target Search

Assume all human and robot agents share a fixed metric map \( \mathcal{M} \) of the search space and are aware of their own positions on \( \mathcal{M} \). At time \( k \), let \( X_k = X \in \mathbb{R}^n \) be the unknown static target location with prior pdf \( p_0(X) \) (the initial belief). Given a vector of observations \( \mathbf{c}_k \) with conditional likelihood \( p(\mathbf{c}_k | X) \), robot/human agent \( i \) recursively updates its local belief in \( X \) via Bayes’ rule,

\[
p(X|\mathbf{c}_{1:k}) \propto p(X|\mathbf{c}_{1:k-1})p(\mathbf{c}_k | X),
\]

where \( \mathbf{c}_{1:k} = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\} \) and \( p(X|\mathbf{c}_{1:k-1}) = p_0(X) \) for \( k = 0 \). The LHS of (1) summarizes all information gathered by agent \( i \) up to time \( k \) and thus permits efficient information exchange among multiple agents in a large scale search network without requiring storage/transmission of \( \mathbf{c}_{1:k} \) (Bourgault 2005). However, \( p_0(X) \) and \( p(\mathbf{c}_k | X) \) are typically highly non-Gaussian functions in \( X \), so that (1) generally cannot be found in closed-form and must be approximated. Conventional approximate Bayesian filtering techniques for robotic sensor data (e.g. using grid, particle, or Gaussian mixture representations) also naturally extend to ‘human sensor’ data. However, suitable models \( p(\mathbf{c}_k | X) \) must be first found to properly account for uncertainty in human-generated information and avoid inconsistencies when deployed with expert/non-expert humans (e.g. trained search and rescue workers vs. civilian volunteers, who may have little or no experience with robots).

While the semantic codebook likelihood approach illustrated in Fig. 2 is generally effective and allows users to supply useful information using a simple high-level language, it can lead to suboptimal fusion results and user frustration in actual human-robot search applications, since humans cannot ‘tweak’ inferred \( p(\mathbf{c}_k | X) \) models provide more accurate/precise spatial information with respect to \( \mathcal{M} \) that is otherwise cumbersome or impossible to express (Sample, Ahmed, and Campbell 2012). For example, the model for ‘nearby (location)’ in Fig. 3 is restricted to imply that a target is nearby a single recognizable 2D map coordinate, and thus cannot readily express information that the target is somewhere nearby a large irregularly shaped landmark (e.g. a lake or building) or a vaguely defined region (e.g. part of a town). While it is always possible in principle to augment codebooks online to account for such cases, in practice this increases the complexity of the user interface and can make learning of \( p(\mathbf{c}_k | X) \) harder. This can be especially problematic if only sparse/limited human sensor calibration data are available, since (1) must then be modified to account for the parametric uncertainty of \( p(\mathbf{c}_k | X) \).

3.1 Probabilistic Human Sketch Sensor Model

This section describes our newly proposed sketch-based data fusion interface for human search agents to address the limitations of fixed semantic codebook interfaces. The sketch interface allows users to quickly convey their search observations by indicating the possible presence/absence of the target in different regions of the search space \( X \) in \( \mathcal{M} \), which are denoted by free-form sketch encirclements drawn directly on \( \mathcal{M} \) by the human. This idea is illustrated in Fig. 4(a), which shows that a human search agent can specify the boundaries of ad hoc spatial regions in which the target may either be ‘inside’ or ‘outside’. This protocol allows the human to roughly ‘classify’ the search space based on his/her observations during the search mission. For instance, the blue sketch regions in Fig. 4(a) roughly imply that ‘the target may not be in these areas’ (i.e. summarizing negative information obtained by a visual sweep of the areas), while the red sketch region imply ‘the target may be around here’ (i.e. which might collectively summarize positive information from clues in the search environment).

Although such sketch observations might be intuitively simple for human search agents to understand and use, they are also potentially very inaccurate and imprecise, and can
Figure 4: (a) Example search map with ‘inside’ and ‘outside’ region free-form human sketches; (b) discretization of sketches on search space grid for $X$ to form $S_{in}$ and $S_{out}$, where filled red/blue cells are 1 and empty cells are 0.

therefore be misleading if only interpreted superficially. In particular, intended sketch region boundaries will not be drawn precisely (especially in time critical situations) and can be contradictory or inconsistent with each other, even for the same human agent. Similar issues were considered by (Wang et al. 2012) in the context of social detection of binary events (e.g. existence of graffiti in public spaces) using intermittent verbal discrete event reports from multiple human agents. As noted in (Wang et al. 2012), the ability to extract useful information from a human observation $\zeta^i_k$ relies only on the ability to accurately model the probabilistic dependence of $\zeta^i_k$ on the actual value of $X$ through $p(\zeta^i_k | X)$, i.e. the exact reason human $i$ reports $\zeta^i_k$ when $X$ is in a given state at time $k$ is irrelevant. Rather, for our purposes, the degree to which human $i$ is expected to give false, misleading, ambiguous, or conflicting information should be captured by $p(\zeta^i_k | X)$, so that (1) remains valid.

Hence, to be treated as useful data which updates the belief in $X$ via (1), each human-sketched ‘inside’/‘outside’ region must be parsed into an observation vector $\zeta^i_k$ conditioned on $X$. The main problems addressed here are: (i) how to construct a likelihood $p(\zeta^i_k | X)$ that formally accounts for the uncertainties embedded within each sketch? (ii) how to learn the parameters for $p(\zeta^i_k | X)$ and carry out (1), especially with large parameter uncertainties due to limited training data?

### 3.2 Probabilistic model for sketch data fusion

**Sketch parsing** Without loss of generality, we assume that human only draws closed regions over the search space $X$ in $\mathcal{M}$ that can be easily converted into distinct simple closed (non-convex) polygons. Furthermore, assume that the human can only choose to label all sketches drawn together at time $k$ as either ‘inside’ or ‘outside’. Next, consider a discretized grid model for the searchable space $X$ in $\mathcal{M}$, such that $X \in \{1, 2, ..., n\}$, where $n_X$ is the number of grid cells. As shown in Fig. 4 (b), sketches drawn at time $k$ are discretized and converted into a corresponding grid of labeled ‘in’ (or ‘out’) cells, $S_{in}(k)$ (or $S_{out}(k)$), such that cells in the corresponding regions are labeled $S_{in}(k)[s] = 1$ (or $S_{out}(k)[s] = 1$) and all other cells are set to $S_{in}(k)[s] = 0$ (or $S_{out}(k)[s] = 0$), for grid index $s \in \{1, ..., n\}$. Denote the set of all $S_{in}(k)$ and $S_{out}(k)$ observation grids as $S_{in}$ and $S_{out}$, respectively.

**Likelihood modeling** Consider a hypothetical sketch in which in exactly one grid cell $s \in \{1, ..., n\}$ is labeled by the human as either ‘inside’ (i.e. containing the target) or ‘outside’ (not containing the target). Then, given the true value of the target location $X \in \{1, ..., n\}$, the human’s binary label for $s$ must either be correct or incorrect with some probability. In particular, assuming for now that the events $S_{in}(k)[s] = 1$ or $S_{out}(k)[s] = 1$ are only conditionally dependent on $X$, we can define the following conditional probabilities for the lone marked cell $s$ in $S_{in}(k)$,

$$a_i = p(S_{in}(k)[s] = 1 | X = s) \text{ (true positive)}$$

$$b_i = p(S_{in}(k)[s] = 1 | X \neq s) \text{ (false positive)}.$$

Thus, we also get for a lone marked cell $s$ in $S_{out}(k)$,

$$1 - a_i = p(S_{out}(k)[s] = 1 | X = s) \text{ (false negative)}$$

$$1 - b_i = p(S_{out}(k)[s] = 1 | X \neq s) \text{ (true negative)}.$$

Hence, given $X$, the human either chooses to label any single cell $s$ as ‘inside’ or ‘outside’ with the likelihoods

$$l_{in}(s; a_i, b_i, X) = a_i \delta(X, s) \cdot b_i^{1-\delta(X, s)}$$

$$l_{out}(s; a_i, b_i, X) = (1 - a_i) \delta(X, s) \cdot (1 - b_i)^{1-\delta(X, s)},$$

where $\delta(X, s)$ is the Kronecker delta.

Consider the full ‘inside’ sketch of Fig. 4, which implies that the human has ‘not detected’ the target inside any of the blue cells of $S_{out}(k)$ in Fig. 4 (b) (i.e. ‘the target may not be here and here and here...’). If
we were to adopt a ‘naive Bayes’ approach to determine the joint observation likelihood of the set \( S^-(k) \) of all \( N_{out,k} \) colored grid cells for which \( S_{out}(k)[s] = 1 \), we obtain

\[
l_{NB}(S_{out}(k); a_i, b_i, X = x) = \prod_{s \in S^-(k)} l_{out}(s; a_i, b_i, X = x) = (1 - a_i)^{S_{out}(k)[x]} \cdot (1 - b_i)^{(N_{out,k} - S_{out}(k)[x])},
\]

(9)

This does not account for any unknown conditional dependencies that exist between the different grid values in \( S_{out}(k) \) given \( X = x \); such dependencies are expected to arise since the human assigns ‘outside’ labels en masse within each region boundary sketch, rather than individually labeling grid cells. The naive Bayes assumption behind (9) thus treats each grid cell in \( S_{out}(k) \) as if it were an independent observation. Although simple and convenient, we have found in practice that this tends to significantly overestimate the amount of new independent information contained within \( S_{out}(k) \) observations. To address this issue and bypass the need for exactly modeling the dependencies within \( S^- \) while still maintaining a tractable likelihood expression, we apply a simple conservative log opinion pool approximation to (9), which was also applied in Fox to address similar issues for modeling WiFi signal strength likelihoods with unknown correlations for robot localization. This results in the likelihood model

\[
l_{LOP}(S_{out}(k); a_i, b_i, X = x) = (1 - a_i)^{\omega_x S_{out}(k)[x]} \cdot (1 - b_i)^{(1 - \omega_x)(N_{out,k} - S_{out}(k)[x])}
\]

(10)

Eq. (10) downweights the contribution from all \( S_{out}(k) \) observations by shifting the contribution between the ‘false negative’ and ‘true negative’ terms via parameter \( \omega_x, i \in [0, 1] \). In practice, \( \omega_x,i \) should be relatively high to account for the fact that there will be many more true negatives than false negatives in a given ‘outside’ sketch (i.e. the target is only in one cell and there are many unoccupied cells). For our applications, we have found that setting \( \omega_x,i = 0.5 \) often produces reasonable results. However, a more formal fully Bayesian learning approach can also be used to simultaneously capture uncertainty in \( X \) as well as the model parameters \( \omega_x,i, a_i \) and \( b_i \), which govern the ‘trustworthiness’ of human sketch sensor \( i \).

### 3.3 Bayesian Learning and Sensor Fusion

We propose a hierarchical fully Bayesian approach for simultaneously learning \( \omega_x,i, a_i, b_i \) and estimating \( X \), in which these variables are all assigned suitable prior probability distributions and hyperparameters for these distributions are also treated as uncertain random variables. The hierarchical Bayesian modeling approach allows us to be ‘agnostically’ about the exact form of the human sketch sensor likelihood in the presence of probabilistic model ‘meta-uncertainty’ (Mankell 2008), while also allowing us to determine the (un)reliability of particular human sketch sensors

1Hence, the human is ambiguous here about which cells he/she ‘meant to exclude’ from \( S^- \), in logical complementarity to the assignment ambiguity for ‘inside’ sketches and form a consistent global belief about \( X \) from human and robot data. Our proposed approach is similar to the one proposed by (Kaupp et al. 2005) for fully Bayesian model and state estimation with human range-bearing sensors, except that we consider a centralized information fusion scenario (instead of a distributed/decentralized one), and we require more sophisticated Bayesian inference approximations to handle more flexible and ambiguous inputs from multiple humans at once.

There are two possible learning modes. For supervised learning, assume that \( T \) recorded search mission training data sets \( D_{SV}(i, t) = [S^i_{in}, S^i_{out}], x_t \) are available for human \( i \) in training instance \( t \in {1, \ldots, T} \), where \( x^*_t \) is the true known value of \( X \) in search mission \( t \). Hence, mock training/calibration exercises and/or results from actual (successful) search missions can be used to generate sketch training data sets. For unsupervised learning, the data for human \( i \) becomes \( D_{US}(i, t) = [S^i_{in}, S^i_{out}] \), which corresponds to online ‘in situ’ learning, e.g. with human agents we have never seen before and where the true target location is unknown (and possibly also being estimated online by a search robot). Data sets for unsupervised learning can be collected over the course of an actual mission and/or combined afterwards for ‘group-wise learning’, since the parameters for different humans participating in at least one common mission \( t \) will become dependent if \( X_t \) is unknown.

In addition to learning \( \omega_x,i, a_i \) and \( b_i \), we also want to estimate \( X \) from the sketch data (or, more generally, find its posterior distribution via eq. (1)). This is straightforward if we can obtain precise point estimates of \( \omega_x,i, a_i \) and \( b_i \) via supervised learning, but this requires a sufficiently large number of training samples \( T \) to achieve reliable learning results. For supervised/unsupervised learning where \( T \) is very small (e.g. less than ‘inside’ and ‘outside’ 10 sketches total over the course of an entire mission in a large search area), the uncertainty associated with the learned parameters becomes very significant and the fully Bayesian approach naturally takes this into account when deriving an estimate for \( X \).

### 3.4 Probabilistic Graphical Model and Gibbs Sampling Inference

Fig. 5 shows the probabilistic graphical model (PGM) used to perform learning and state estimation with human sketches for both the supervised and unsupervised learning cases. Eqs. (8) and (10) give the likelihoods for the \( S^i, k \) nodes, while separate Beta distribution priors are assigned to \( a_i, b_i \), and \( \omega_x,i \) with shape/scale hyperparameter pairs \((k_1, k_2), (k_3, k_4)\), respectively, which in turn are all assigned separate Gamma distribution hyperpriors with shape and scale parameters of 1. Note that the hyperparameters for all human sensor distributions are assumed to be tied together. This implies that the human sensor model specifications are all generated by the same statistical distribution and are thus conditionally dependent on each other whenever \( X_t \) and all hyperparameters \( k \) are unknown, i.e. when information about the (un)reliability of one human sensor is obtained, this also influences our belief in the (un)reliability of another human sensor, since our belief in the distribution
Figure 5: PGM for supervised/unsupervised human sketch sensor learning and state fusion; shaded nodes denote hidden/unknown variables to be learned/estimated (unsupervised case shown here only with $X_t$ unknown in all training instances).

of (un)reliability of human sensors in general has been altered by the consistency and accuracy of evidence we receive from sketches (or robot data) regarding $X_t$. $X_t$ is assumed here to be a uniform distribution, but can generally be assigned whatever prior distribution is deemed appropriate for the initial phase of a particular search mission.

In the case of general unsupervised learning, it is desired to obtain the posterior distributions for $X_t$, $\omega_{x,t}$, $a_t$, $b_t$, and $k_{1:i}$ for all humans $i = 1 : H$ and all training instances $t = 1 : T$. Since the exact posterior distribution for the PGM in Fig. 5 is analytically intractable for the distributions specified here, approximate Bayesian inference techniques such as Markov Chain Monte Carlo sampling must be used. Although not derived here due to limited space, it is straightforward to show that an efficient Gibbs sampler can be constructed for all the variables of interest and implemented with fast adaptive rejection sampling techniques to obtain sample points from the exact joint posterior distribution over all variables (Gilks and Wild 1992).

3.5 Experimental Results

To assess the utility of our proposed sketch fusion interface in the context of actual multi-agent target search missions, this section presents offline parameter learning and state estimation results for data sets collected from outdoor search experiments. The experiments were conducted with a real team of 6 human participants (1 female; all 20-26 years old) who had no prior experience with the experimental software or hardware.

3.6 Experimental search mission setup

The human search team was tasked to find a small object (a blue-green keychain) hidden somewhere on Cornell University’s Pew Engineering Quadrangle (a ~6800 m$^2$ search space, shown in Fig. 4) in 15 minutes or less using a networked soft data fusion interface programmed on mobile tablet PCs. Seven similar search scenarios were used to study the team’s performance with the sketch interface. The humans were shown the target beforehand and were free to use any search strategy they wanted, as long as they did not directly communicate with each other (verbally or visually) unless they were explicitly told to do so. The large open quad search space was divided into 8 small ‘villages’ marked by labeled red traffic cones on the quad to provide a semi-structured map $M$. The humans were told beforehand that target could be anywhere within approximately 10m of any village center. The target was always placed on the ground and partially buried under dirt/grass so that it could only be seen about 0.5 m away from any viewing direction. This made the difficulty of finding target about the same for each scenario, as confirmed by preliminary search trials with other human subjects. To engage both the human reasoning and visual detection skills during the searches, 48 printed ‘clues’ taped onto plastic balls were randomly placed throughout the quad to simulate external evidence that the humans could freely choose to interpret and incorporate/ignore (e.g. ‘maybe look near Village X’ or ‘it’s not near a bush’). No clues are deceptive, but some were intentionally very vague. One ‘special’ clue in each village was based on the actual target location for each scenario; the remaining clues were always true and recycled between scenarios.

3.7 Bayesian Learning and Estimation Results

The sketch data provided by each human search agent were logged and analyzed offline with the proposed Bayesian learning and estimation methods, using grid-based approximations to carry out the desired inference calculations. A total of 310 individual sketches were provided by the humans (143 ‘inside’, 32 of which encompassed the actual target location).

Fig. 6 shows raw sketch data obtained from all 6 agents for Mission 7 alongside the estimated Bayesian posterior probability distribution for $X$. This underscores the huge variability in the way humans search agents can interpret and report simple binary information about the same search environment. Although not visible from this figure, agents appeared to prefer using ‘inside’ or ‘outside’ sketch observations to noticeably different extents, an factor which is not yet accounted for in our sketch likelihood model. In particular, agent 5 made the most prolific use of negative information via ‘outside’ sketches, whereas most other agents tended to ‘speculate’ on the location of the target and drew positive information via ‘inside’ sketches. Despite this discrepancy, it is interesting to note that the posterior distribution for $X$ remains consistent, i.e. the true location of the target still has a relatively high probability, and other parts of the search map are correctly downweighted with negative information.

Fig. 7 shows sample histograms for the Bayesian posterior distributions for the unknown sensor model parameters for two human agents under different learning conditions. These results show that the Bayesian learning algorithm can learn to assess which humans are actually providing reliable information, and which ones are likely speculating. As noted earlier, human agent 5 made the most prolific
Figure 6: Raw sketch data obtained from all 6 agents for Mission 7 alongside the estimated Bayesian posterior probability distribution for X (true target location denoted by green star, white/black indicates highest/lowest probability; magenta/blue sketches indicate ‘inside’/‘outside’ observations)

use of negative information, which should tend to decrease \(b_2\) since this is most often the correct observation for any given cell in \(M\). The results for unsupervised learning using only single mission data shows that the distribution for \(A_1\) and \(b_5\) are slightly skewed towards smaller and larger values, respectively, while the values for the remaining sensor model parameters remain concentrated near 0.5. This shows that the parameter estimates for the sensor models are ‘diluted’ by uncertainty in \(X\). This dilution effect vanishes for \(b_5\) when \(X_i\) is observed in the supervised learning results: much sharper posteriors are obtained that indicate agent 5 is consistently reliable in terms of not providing false positives, while agent 1’s false positive parameter still ‘hovers’ near 0.5. Note, however, that the parameter estimates for \(A_1\), \(A_5\), \(w_{x,1}\) and \(w_{x,5}\) still remain fairly diluted even in the supervised learning case, which indicates that these parameters are only very weakly observable from the gathered data. In the case of \(A_i\) parameters, this can be explained by inherent data sparsity, i.e. the fact that the parameter is only defined for data where a sketched cell happens to contain the true target location, which is a relatively uncommon event. Future work will examine more alternative experimental data collection strategies for overcoming these biases in the training data.

4 Expressing Robotic Uncertainty via Semantic Sensor Models

The robot’s beliefs stored as a probability distribution is a valuable piece of knowledge which could be useful in the tasks such as target search and rescue, target tracking and localization. To make the best use of such knowledge, the key question is how can we communicate such beliefs to the other agents like the human users who may not be good at interpreting dense quantitative representation. In other words, how can a robot translate a probability distribution of its beliefs into structured sentences that can be easily understood by humans? One approach considered here is a prediction on the applicability of each word in a semantic likelihood dictionary (cf. Fig. 3) to the current robot’s belief.

4.1 Word Applicability Prediction

The belief distribution of a state \(x\) can be defined as \(b(x) = p(x|I, \Theta_{i=1:M}, x_l)\), where \(I\) is all the information acquired by the robot through its own perception or through interactions with the other agents. \(\Theta_{i=1:M}\) are the parameters describing each \(i \in \{1, \ldots, M\}\) word in an a priori learned dictionary with respect to a reference (e.g. landmark) \(l\) with a state \(x_l\). The applicability \(D_{x_l}\) of each word \(i\) in the dictionary with respect to a reference \(l\) can be predicted from the meaning of each word which is defined by

\[
p(D_{x_l}|I, \Theta_{i=1:M}, x_l) = \int p(D_{x_l}|x, \Theta_{i=1:M}, x_l)p(x|I, \Theta_{i=1:M}, x_l)dx
\]

(11)

The relationships between all the random variables are illustrated in the probabilistic graphical model in Figure 8. As shown in (Ahmed, Sample, and Campbell 2013), in the situation when the word meaning \(p(D_{x_l}|x, \Theta_{i=1:M}, x_l)\) is defined as a mixture of softmax model (MMS) and the belief \(p(x|I, \Theta_{i=1:M}, x_l)\) is represented in a Gaussian mixture, the integral above can be evaluated by performing variational Bayesian importance sampling (VBIS).
Figure 9: The robot’s belief $b(x) = p(x|\mathcal{I}, \Theta_{l=1:M}, x_l)$ on the target’s state $x \in \mathbb{R}^2$ given a human input statement “The target is <preposition> <landmark>,” where the landmark location is $x_l = [15 - 5]^T$, and the preposition is (a) “Nearby” (b) “Far” (c) “Next To”.

To demonstrate the word applicability prediction method, simple simulations are performed as follows: First, a dictionary of English spatial prepositions, and hence the MMS parameters describing each word $\Theta_{l=1:M}$, is constructed from the training data collected from human as described in reference (Ahmed, Sample, and Campbell 2013). A structured English statement is then given as an input to the robot’s information fusion system in the following format: “The target is <preposition> <landmark>,” where preposition $\in \{“Nearby,” “Far,” “Next To”\}$. The landmark location $x_l \in \mathbb{R}^2$ is also provided to both the human and the robot beforehand in the form of a shared map of the environment. A uniform prior on the target location is used for the human information fusion process. A belief distribution $p(x|\mathcal{I}, \Theta_{l=1:M}, x_l)$ of the target state $x \in \mathbb{R}^2$ is then computed by the same principle as in (Ahmed, Sample, and Campbell 2013) except that a grid-based calculation is used for the fusion process. A starting belief pdf is provided in Figure 9 for each case of human input tested.

Eq. (11) is then used calculate the applicability probability of each output statement “The target is probably <preposition> <landmark>,” preposition $\in \{“Nearby,” “Far,” “Next To”\}$, landmark $x_l = [15 - 5]^T$ in describing each case of robot’s belief pdf in Figure 9. Table 4.1 shows the predicted applicability result using all three cases of information $\mathcal{I}$ received by the robot (‘Output’ denotes robot’s interpretation of resulting posterior according to human’s semantic codebook).

Table 1: $p(D_{l_i} = 1|\mathcal{I}, \Theta_{l=1:M}, x_l)$ results for $i \in \{1, 2, 3\}$

<table>
<thead>
<tr>
<th>Human Input</th>
<th>Nearby</th>
<th>Far</th>
<th>Next To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearby</td>
<td>0.5145</td>
<td>0.0030</td>
<td>0.4602</td>
</tr>
<tr>
<td>Far</td>
<td>0.2553</td>
<td>0.9967</td>
<td>0.0530</td>
</tr>
<tr>
<td>Next To</td>
<td>0.2302</td>
<td>0.0003</td>
<td>0.4867</td>
</tr>
</tbody>
</table>

References


