

Enhancing Robot Perception Using Human Teammates*

(Extended Abstract)

Jean Oh, Arne Suppe, Anthony Stentz, and Martial Hebert
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA
{jeanoh,suppe,axs,hebert}@cs.cmu.edu

ABSTRACT

In robotics research, perception is one of the most challenging tasks. In contrast to existing approaches that rely only on computer vision, we propose an alternative method for improving perception by learning from human teammates. To evaluate, we apply this idea to a door detection problem. A set of preliminary experiments has been completed using software agents with real vision data. Our results demonstrate that information inferred from teammate observations significantly improves the perception precision.

Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Intelligent agents

General Terms

Human Factors

Keywords

Robot perception, robot-human hybrid teams

1. BACKGROUND

Robot perception is generally formulated as a problem of analyzing and interpreting various sensory inputs, *e.g.*, camera feeds. In this paper, we approach robot perception from a completely different direction. Our approach utilizes a team setting where a robot collaborates with human teammates. Motivated by the fact that humans possess superior perception skills relative to their robotic counterparts, we investigate how a robot can take advantage of its teammate's perfect vision.

In general, an agent acquires new *information* through *perception*, and in turn, the agent chooses *actions* based on the information acquired. Let us suppose that a robot has a mental model of its human teammate such that a causal relationship is specified between information and actions. Then, by understanding the human mental model of such decision making (or *planning*), the robot can infer what the human teammate has seen based on the human's behavior. In other words, an observation of a human teammate can be

*This work was conducted (in part) through collaborative participation in the Robotics Consortium sponsored by the U.S Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement W911NF-10-2-0016.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

Copyright © 2013, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

used as evidence to infer the information perceived by the human. This, in turn, can be used to reduce uncertainty in robot perception.

In this paper, we specifically focus on a motivating problem of door detection in the following scenario. Consider a team consisting of a robot and a human performing a military operation in a hostile environment. According to intelligence, armed insurgents are hiding in an urban street. The team is deployed to cover the buildings in the surrounding area, focusing on doors from which the insurgents may try to egress. This is a stealth operation.

We make two specific assumptions that are reasonable in a team context. First, observing a teammate is generally more manageable than perceiving an unfamiliar environment. Second, team members share common objectives in reaching the team's goals.

2. PERCEPTION USING VISION

This section describes a purely camera-based approach. First, we find a likely semantic image segmentation using a computer vision technique called stacked hierarchical labeling [3] which uses machine learning to classify regions within an image. This labeling approach generally demonstrates high accuracy in identifying man-made objects, *e.g.*, in our scenario, the precision and recall rates of detecting buildings are 0.937 and 0.934, respectively.

It is not constrained by shape grammars and can model a more general class of objects, but its method of constructing a hierarchical segmentation does not convey semantic meaning at a finer detail, as would be necessary to detect doors on a building. It is, however, reliable in detecting buildings as a whole, significantly reducing the search space for detecting doors in the next step.

Once buildings are identified, we can apply a broad feature detector to detect likely openings on the façade of the building. As in [2] we observe that symmetry is a common hallmark of architectural elements. However, we use Kovese's phase symmetry detector [1] to find regions of high vertical symmetry instead of mutual information. This technique is particularly effective when coupled with range measurements so that the symmetry detector can be targeted to the proper scale. Unfortunately, for our experiment, that information was not available leading to much degraded performance. As shown in Figure 1b, the vision-based algorithm predicted that the entire regions labeled as "building" as possible doors but with very low confidence.

3. PERCEPTION USING TEAMMATES

This section describes enhancing perception accuracy using the eyes of a human teammate. Here, we define the causal relationship between information and actions in terms of inter-visibility. A position (*destination*) is said to be "exposed" to another position (*source*) if a line of sight can be established from the source to the destination. We assume that humans prefer stealthy (less exposed)

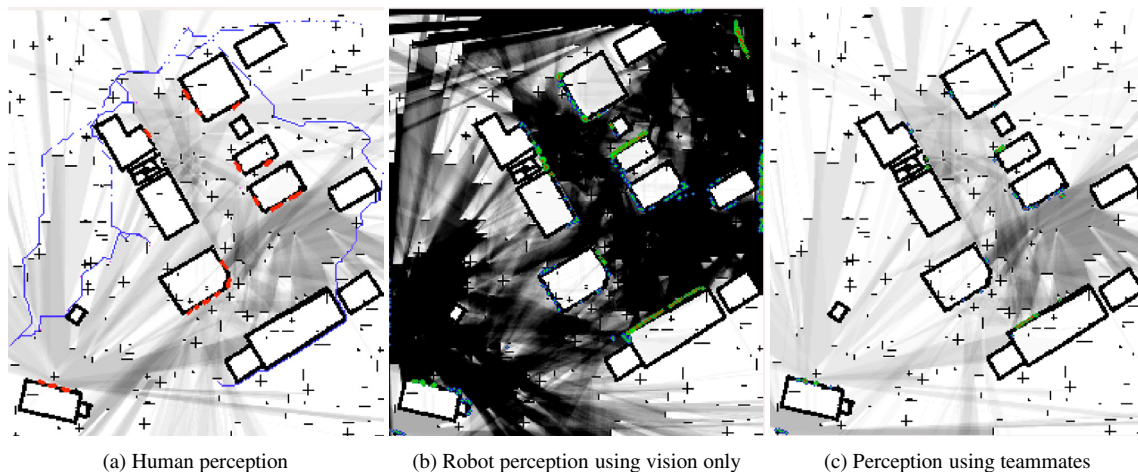


Figure 1: Door detection enhanced using SMM.

paths; and that humans stay low to the ground to reduce the risk of being exposed. More specifically, we assume that the *riskiness* of a height z is inversely proportional to the number of sources v to which the position is exposed, *s.t.* $z \propto v^{-1}$.

The robot maintains a set of candidate doors associated with confidence values. Using the current estimation of the candidates, it computes a map specifying the stealthiness of all destinations (or cells)—referred to here as the cost map. The cost maps shown in Figure 1 can be interpreted as follows: the darker the shade is, the riskier. The robot’s initial cost map (*e.g.*, Figure 1b) is derived from the camera-based detection results described in Section 2. Our goal here is to make the robot learn a cost map that is close to that of its human teammate.

We utilize teammate observations to enhance this cost map in two ways. First, we interpret each observation independently. Second, we also analyze a combined sequence of observations. Both methods follow the same basic principle of trying to match what is observed with what is expected.

Given a teammate observation $o = (x, y, z)$, the height (or z -value) of the observation provides key evidence indicating how stealthy the teammate is, *e.g.*, the maximum height corresponds to a riskiness value of 0 indicating that the position is not exposed to any doors, or vice versa. Using the current cost map, the robot computes the predicted number of exposures v' at position o and compares that with the number of exposures v inferred from observation. If the expectation matches the observation, then the robot’s cost map is consistent with that of the human, so the update process stops. Otherwise, we collect a set S of candidate sources to which position o is exposed and compute the new confidence value $c(s)$ for each source s , *s.t.* $c(s) = \frac{v}{|S|}, \forall s \in S$. We then update each source’s confidence value as a weighted sum of new and old values.

When the robot has access to the teammate’s subgoals—a set of milestones towards accomplishing the final goals—a sequence of observations can also be used to draw higher-level inferences. Given a sequence of observations, we formulate a planning problem by taking the first and the last observations as start and goal states. We then use a planner to find an optimal plan for that planning problem. If the observed path is different from the optimal path (*i.e.*, expectation) then it implies that the expected path must have a higher cost than the observed one. We lower the heights of positions in the expected path by using the average height of the observed path. Using this modified expectation, we update the cost map as if a stealthier path were observed.

Figure 1c demonstrates the robot’s cost map after the human teammate had achieved the 10th subgoal that closely resembles that of the human teammate in Figure 1a.

Let h^* and \tilde{h} denote the true and the estimated cost maps of teammate h , respectively. Our objective is to minimize the discrepancy between the two maps, such that $\lim_{t \rightarrow \infty} |\tilde{h} - h^*| = 0$ where t denotes time step. The performance is, therefore, measured in terms of mean squared error between the true and estimated cost maps.

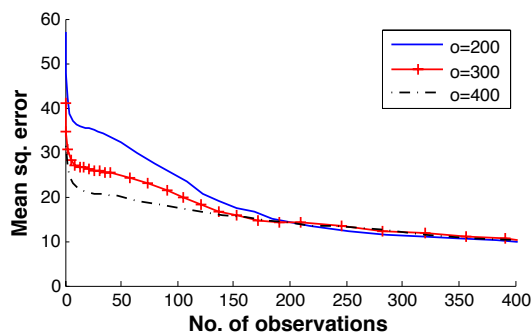


Figure 2: Error measuring difference between a robot’s cost estimation and that of a human; parameter o specifies the number of obstacles.

Figure 2 shows that robot perception improves over time as evidenced by measurements of the difference in costs between the robot’s estimation and that of the human agent. Here, the results are shown with varying degrees of visibility, parametrized by the number of obstacles o . In this set of results, it appears advantageous to have a higher number of objects (*i.e.*, a higher o value) especially in the beginning of the learning process. This is because additional obscuring objects reduce the average number of candidate sources. Overall, this result demonstrates that our algorithm effectively learns a teammate’s cost map after observing about 100 steps. In terms of mean squared error (MSE) in cost estimation, error is reduced by 80%.

4. REFERENCES

- [1] P. Kovesi. Symmetry and asymmetry from local phase. In *Australian Joint Conference on AI*, pages 2–4, 1997.
- [2] P. Müller, G. Zeng, P. Wonka, and L. Van Gool. Image-based procedural modeling of facades. In *SIGGRAPH, SIGGRAPH '07*, New York, NY, USA, 2007. ACM.
- [3] D. Munoz, J. A. Bagnell, and M. Hebert. Stacked hierarchical labeling. In *Proc. ECCV*, 2010.