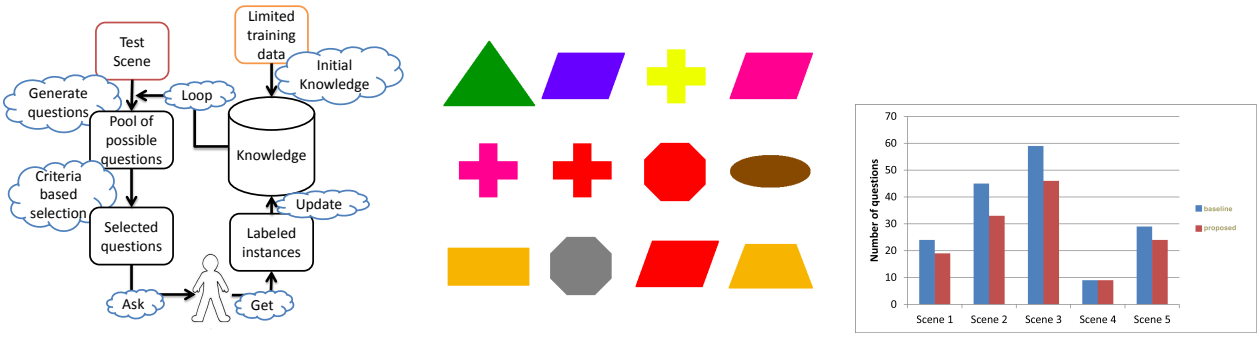

(a) Architecture (b) Illustrative scene (c) Experimental results

Figure 1: Figure 1: (a) Proposed architecture; (b) An example scene; (c) Proposed architecture requires less number of queries than baseline algorithm.

humans. Humans frequently use contextual cues to draw attention to an object of interest. Contextual cues can take different forms, and positional context with reference to a known object can be very useful in disambiguating the object of interest. For instance, instead of referring to a "1965 Ford Mustang" in a busy street intersection, we may refer to the "Red car behind the bus", using both feature labels (e.g., color and object labels) and positional reference to a known object.

2.3 Query types

We consider candidate queries based on different contextual cues, and rank them based on measures of information gain, ambiguity and human confusion. Some examples of possible queries are posed below:

- *Object Label*: "What is the label of the object in the bottom right of the scene?"
- *Feature Label*: "What is the label of the color of object in top left of the scene?"
- *Affirmation*: "Is there a blue object in the scene?"
- *Location*: "Where is a blue object?"

3. RESULTS AND DISCUSSION

We report preliminary results of evaluating our architecture in a simulated domain [2]. We abstract away the non-determinism in object recognition and speech understanding; objects are recognized once the individual features are learned, and speech gets translated into text and parsed to generate the labels. The trials summarized below consider simulated images with objects characterized by 10 different colors and 15 different shapes. As an illustrative example of query generation, consider the scene in Figure 1(b). Assume that the color and shape labels of four objects are known a priori:*pink star, green arrow, blue heart,* and *yellow cross.* The following are some of the queries generated; each line ends with the answer provided to the question:

- *Iteration 4*: "What is the label of the object in the bottom right of the scene?" **Orange Trapezoid**.
- *Iteration 6*: "What is the label of the object that is to the left of the orange trapezoid?" **Red Parallel**.
- *Iteration 13*: "What is the label of the object that is above the red parallel?" **Red Octagon**.

The information obtained by posing questions is used to formulate and pose subsequent questions, and questions may refer to more than one object. Figure 1(c) summarizes results for five scenes. These scenes differ in terms of the number and type of objects in the scene. For instance, "Scene 1" corresponds to Figure 1(b), while "Scene 2" and "Scene 3" have 20 and 30 objects respectively. For each scene, the robot started with the same initial knowledge about a subset of the objects. As a baseline for comparison, we used an algorithm that started with the same initial knowledge but selected queries randomly from the set of candidate queries. We observe that the number of queries required to acquire the desired labels increases as the scene becomes more complex. However, for each set of paired experimental trials, our architecture results in the desired labels of objects and features being acquired by posing a much smaller number of queries. Similar results were obtained over 100 different (randomly generated) scenes with different number and types of objects. Future work will implement and evaluate the architecture in more complex scenes, and on physical robots interacting with non-expert humans.

REFERENCES

[1] M. Cakmak and A. Thomaz. Designing robot learners that ask good questions. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 17–24, March 2012.

[2] B. Myagmarjav and M. Sridharan. Incremental knowledge acquisition with selective active learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, May 4-8, 2015.

[3] H. Raghavan, O. Madani, and R. Jones. Active learning with feedback on both features and instances. *Journal of Machine Learning Research*, 7:1655–1686, 2006.

[4] S. Rosenthal, A. K. Dey, and M. Veloso. How robots' questions affect the accuracy of the human responses. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 1137–1142, 2009.

[5] B. Settles. *Active Learning*. Morgan & Claypool publishers, 2012.

[6] B. Settles, M. Craven, and S. Ray. Multiple-instance active learning. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1289–1296. Curran Associates, Inc., 2008.