Schedulebot: A Home Robot Learning and Acting Schedule Adaptively via Dynamic Environments

Chung-Yeon Lee, Sang-Woo Lee
Dept. of Computer Sci. and Eng.
Seoul National University
Seoul, Korea
{cylee, slee}@bi.snu.ac.kr

Chaeeun Lee
Dept. of Computer Sci. and Eng.
University of Minnesota
Minneapolis, MN, USA
leex8119@umn.edu

Byoung-Tak Zhang
Dept. of Computer Sci. and Eng.
Seoul National University
Seoul, Korea
btzhang@bi.snu.ac.kr

***Abstract*— It is important for a personal or service robot to be learned autonomously from the real world. This paper proposes an approach to developing an autonomous schedule learning and acting robot platform in dynamic home situations. Experimental setup of a pseudo-real home environment with a scenario and robot platform is described. Ongoing results including visual scene understanding and event prediction are presented.**

Keywords—home robot; schedule learning; dynamic environment; event prediction; multisensory data

I. Introduction

One of the primary challenges in service robotics is to allow robots to adapt to open-ended dynamic environments where they need to interact with naïve users with a little understanding of the strengths and limitations of the robotic systems [1]. To operate safely and efficiently in these complex real situations such as private homes, the robot needs to understand the environments and ongoing events.

To illustrate some of these problems, we employ the following example of a schedule that could happen in the morning at home, there existing a child going to an elementary school. A robot has to perform a role of a mother who takes care of the schoolchild. It can wake up the child, guides him/her to do some exercise and to take a shower, gives a breakfast, helps to wear appropriate clothes, and informs to pack study supplies or textbooks.

One solution to this problem is to hand-code the robot behaviors following a predefined user's schedule and home environment dataset. But there is no guarantee that the child would finish each task for exact time or every object would be located at the right position. Therefore, it is not feasible to hand-code all possible behaviors and to predict all possible exceptions under these complex real situations that the robot is faced with.

We aim to solve the problem by developing an autonomous schedule learning platform dubbed as Schedulebot. The bot freely moves around home and perceive the spatial information and objects, or follows family members and observe their activities and conversation. Based on the spatiotemporal data, the robot understands current events and perform right actions.

II. Experimental Setup

To show how mother and child normally spend a morning time to a robot, we built a pseudo-real home environment that family members and a robot can interact with each other based on the morning at home scenario as described in Table 1. The home space is divided into a bedroom, living room, kitchen, bathroom, and entrance, as shown in figure 1.

The robot can move all of the home spaces by wheels and observe the environments and people using a wide range of sensors, including RGB and depth camera, infrared and laser sensors, odometer, and microphones.

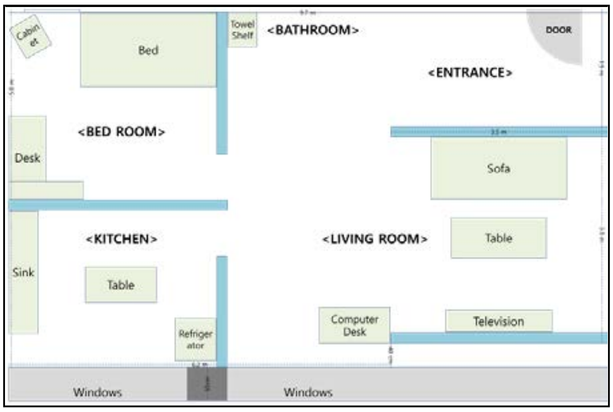

Fig. 1. Home environments setting for experimental scenario

TABLE I. Morning at Home Scenario

Time	Event	Location	Action
7:00	Wake up	Bedroom	Move, Speech, …
7:05	Exercise	Living room	Recognize, Speech, …
7:13	Washing	Bathroom	Move, Speech, …
7:20	Breakfast	Kitchen	Recognize, Speech, …
…	…	…	…

III. Ongoing Results

A. Understanding Visual Scene

The robot needs to understand what it observes from multimodal sensors. To integrate various types of sensory information, word captioning for each data will be used.

First of all, we generated captions that describe visual scenes recorded by an RGB camera. To understand visual scenes from real situations, the localization (e.g. object detection) and description task needs to be jointly performed. We used a fully convolutional localization network architecture [2] that processes an image with a single, efficient forward pass, requires no external regions proposals, and can be trained end-to-end with a single round of optimization. The architecture is composed of a convolutional network, a novel dense localization layer, and recurrent neural network language model that generates the label sequences.

Figure 2 shows the generated captions for example visual scene. There are several bounding boxes that are detected objects and people, and captions describe each box are generated. A robot then can integrate various information from the captions and better understand the visual scene.

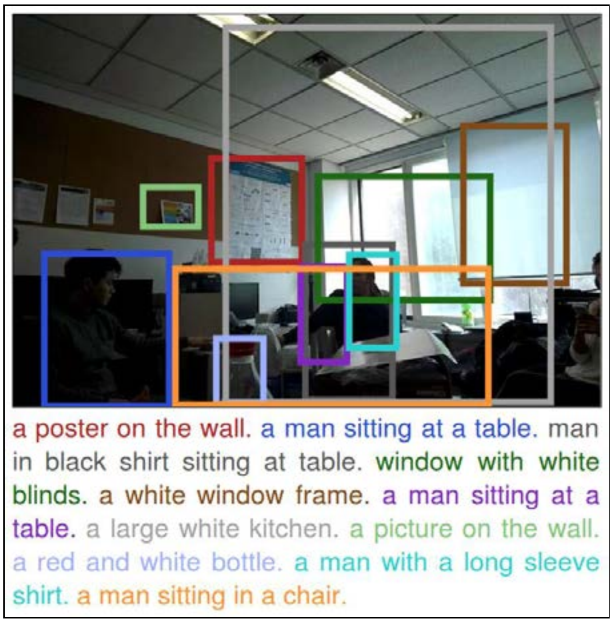

Fig. 2. Generated captions that describe the environment

B. Event Prediction

The robot also needs to predict the current and next events to do appropriate tasks. This event prediction generally will be processed based on the environmental understanding. However, in advance, we tried to perform this task by using a pseudo schedule data randomly generated by following the morning at home scenario. The schedule data is consists of time, location, action, and event labels.

To predict an event from the schedule data vectors, we used a Hypernetworks model [3] which is a molecular evolutionary architecture for cognitive learning and memory.

Predicted events, activities, and robot actions following the results are shown in figure 3. Every event in the scenario is scheduled reasonably except one event: studying event is omitted because the number of data is very small since it happens rarely.

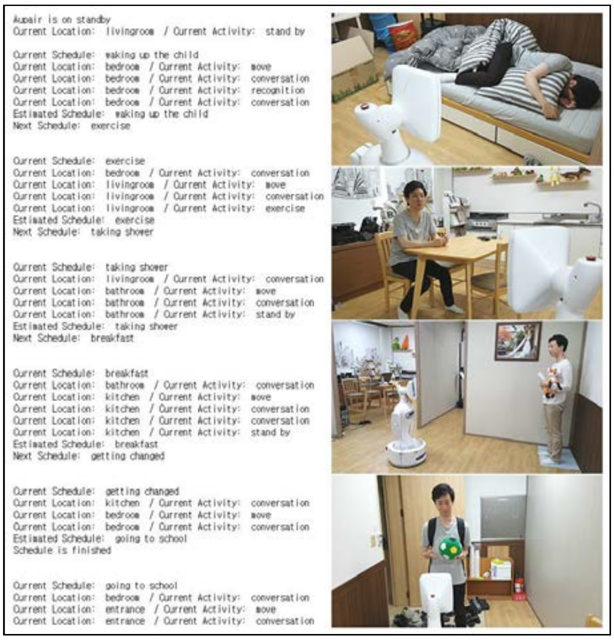

Fig. 3. Event prediction results and robot actions

Acknowledgment

This work was partly supported by the Institute for Information & Communications Technology Promotion (R0126-16-1072-SW.StarLab), Korea Evaluation Institute of Industrial Technology (10044009-HRI.MESSI, 10060086-RISF), and Agency for Defense Development (UD130070ID-BMRR) grant funded by the Korea government (MSIP, DAPA).

References

[1] M. Oliveira, G. H. Lim, L. S. Lopes, S. H. Kasaei, A. M. Tome, A. Chauhan, "A perceptual memory system for grounding semantic representations in intelligent service robots," In Proceedings of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2216-2223, 2014.

[2] J. Johnson, A. Karpathy, and L. Fei-Fei, "Densecap: Fully convolutional localization networks for dense captioning," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2016), pp. 4565-4574, 2016.

[3] B.-T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory," *IEEE Computational Intelligence Magazine*, Vol. 3, No. 3, pp. 49-63, August 2008.