The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Perception-Action-Learning System for Mobile Social-Service Robots using Deep Learning

Beom-Jin Lee,[1] Jinyoung Choi,[2] Chung-Yeon Lee,[1] Kyung-Wha Park,[3] Sungjun Choi,[1] Cheolho Han,[1] Dong-Sig Han,[1] Christina Baek,[3] Patrick Emaase,[1] Byoung-Tak Zhang[1 2 3]
[1]School of Computer Science and Engineering, [2]Cognitive Science Program, [3]Interdisciplinary Program in Neuroscience,
Seoul National University, Seoul, Korea, Republic of

Abstract

We introduce a robust integrated perception-action-learning system for mobile social-service robots. The state-of-the-art deep learning techniques were incorporated into each module which significantly improves the performance in solving social service tasks. The system not only demonstrated fast and robust performance in a homelike environment but also achieved the highest score in the RoboCup2017@Home Social Standard Platform League (SSPL) held in Nagoya, Japan.

Introduction

Recent release of various standardized social-service robots with their idealistic promotional videos are drawing the public's interest in the possibility of co-existing with robots. This is also increasing the public's expectation of social-service robots to become more human-like, capable of providing natural social services to the customers in dynamic environments such as houses, restaurants, hotels and even airports. However, this has been a challenging goal for researchers in the field of social-service robotics.

One promising approach is developing an integrated system of methodologies from many different research areas. This multi-module integrated intelligent robotic system has been widely accepted and its performance has been well known from previous studies (Brooks 1986; Siepmann et al. 2014). However, with the individual roles of each module in the integrated system, perception modules mostly suffered from desynchronization between each other and difficulty in adapting to dynamic environments (Rodríguez et al. 2016). This occurred because of the different process time and scale of coverage of the adopted vision techniques (Collet et al. 2011). To overcome such difficulties, developers usually upgraded or added expensive sensors (hardware) to the robot to improve performances. Though this may have provided some solutions to the limitations, it made companies difficult to release a standardized robot because of increased cost of the robot.

We propose a novel and robust integrated system for mobile social-service robots that at least includes an RGB-D camera and any kind of obstacle detecting sensors (laser, bumper, sonar). We incorporated state-of-the-art deep learning methods to overcome the conventional perception issues for robots to perform various social service tasks in real-time with robustness. Moreover, by designing the system inspired by the cognitive perception-action-learning cycle (Badre 2008), we achieved adaptability to dynamic environments.

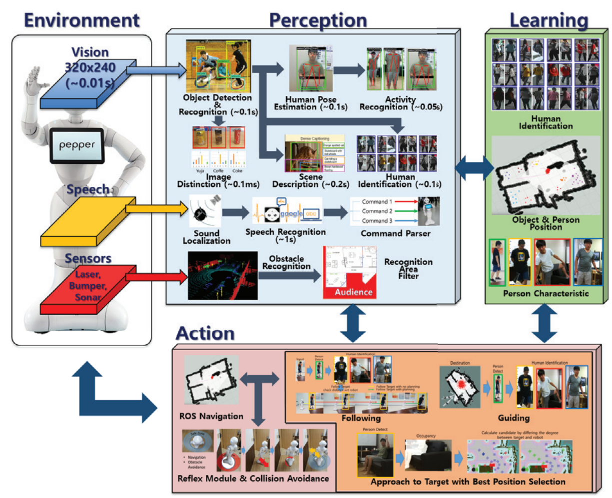

Figure 1: Perception-Action-Learning System for Mobile Social-Service Robots using Deep Learning

System Architecture and Methodology

As illustrated in Figure 1, our system's perception-action-learning cycle works in real-time ($\sim$0.2 s/cycle) where the arrows indicates the flow of each modules. The system was implemented on a server of I7 CPU, 32Gb RAM and GTX Titan 12Gb GPU. Using ROS topics, the communication between the server and the robot were achieved and the ROS topics were passed through 5GHz Wi-Fi connection. The overall information about the implemented modules are organized in Table 1, 2 and 3.

We used our system on SoftBank Pepper, a standardized mobile social-service robot, and achieved the highest score in every scenario performed at the RoboCup2017@Home SSPL, winning first place overall.

Table 1: Perception Modules

No.) Module	Input; Description; Time
1) Object Detection & Recognition	RGB Image; YOLOv2 (2016) performance of 78.6 mAP on VOC2007, real-time speed; ∼0.1 s
2) Human Pose Estimation	1); Realtime multi-person 2d pose estimation using part affinity fields, 79.7 mAP (2016); ∼0.01 s
3) Activity Recognition	2); Rule based recognition; ∼1 ms
4) Scene Description	1); Densecap (2016): Fully convolutional localization networks for dense captioning; ∼0.2 s
5) Human Identification	1), 16); Siamese network based person re-identification (2015); ∼0.1 s
6) Image Distinction	1); Histogram based method ; ∼0.1 ms
7) Sound Localization	Mic; Localize the sound source by the difference between mic input; instant
8) Speech Recognition	Mic; Google Speech API; ∼1 s
9) Command Parser	8), 16), 17), 18); Python Natural Language Toolkit (NLTK) based rule based parsing; instant
10) Obstacle Recognition	Obstacle Sensors; Rule based recognition from sensors (0 or 1); instant

Table 2: Action Modules

No.) Module	Input; Description
11) Navigation	9); Robot Operating System (ROS) Navigation Stack, Adaptive Monte Carlo Localization (AMCL) based SLAM. Default global planner and the Dynamic Window Approach (DWA) local planner
12) Reflex for Collision Avoidance	10); Backing up in the opposite direction of the obstacle and apply curvy path planner to avoid obstacle
13) Following	5), 11); Control for following the target person perceived by 5)
14) Guiding	5), 11); While navigating to designated location, periodically observe for target person
15) Approach	Depth Image, 1), 10), 11); Navigate near the target object and person

Table 3: Learning Modules

No.) Module	Input; Description
16) Human Identification	1); Save person image for 5) re-identification
17) Object & Person Position	1), 5), 11); Save objects and person related information with respect to the map built by ROS Navigation Gmapping
18) Person Characteristic	5), 6); Save person's characteristics described by the perception modules

Conclusion

Our system allows robots to perform social service tasks in real-life social situations with high performance working in real-time. However, our system is yet to fulfill every individual's expectations on performance and processing speed, our demonstration highlights the importance of research on not only the individual elements, but the integration of each modules for developing a more human-like, idealistic robot to assist humans in the future. Related demonstration videos can be found at https://goo.gl/Pxnf1n and our open-sourced codes at https://github.com/soseazi/pal_pepper.

Acknowledgments

This work was supported by the Air Force Office of Scientific Research under award number FA2386-16-1-4089.

References

Ahmed, E.; Jones, M.; and Marks, T. K. 2015. An improved deep learning architecture for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3908–3916.

Badre, D. 2008. Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in cognitive sciences* 12(5):193–200.

Brooks, R. 1986. A robust layered control system for a mobile robot. *IEEE journal on robotics and automation* 2(1):14–23.

Cao, Z.; Simon, T.; Wei, S.-E.; and Sheikh, Y. 2016. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*.

Collet et al., A. 2011. The moped framework: Object recognition and pose estimation for manipulation. *The International Journal of Robotics Research* 30(10):1284–1306.

Johnson, J.; Karpathy, A.; and Fei-Fei, L. 2016. Densecap: Fully convolutional localization networks for dense captioning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4565–4574.

Redmon, J., and Farhadi, A. 2016. Yolo9000: better, faster, stronger. *arXiv preprint arXiv:1612.08242*.

Rodríguez, F. J.; Matellán, V.; Conde, M. Á.; and Rico, F. M. 2016. A motivational architecture to create more human-acceptable assistive robots for robotics competitions. In *Autonomous Robot Systems and Competitions (ICARSC), 2016 International Conference on*, 144–149. IEEE.

Siepmann, F.; Ziegler, L.; Kortkamp, M.; and Wachsmuth, S. 2014. Deploying a modeling framework for reusable robot behavior to enable informed strategies for domestic service robots. *Robotics and Autonomous Systems* 62(5):619–631.