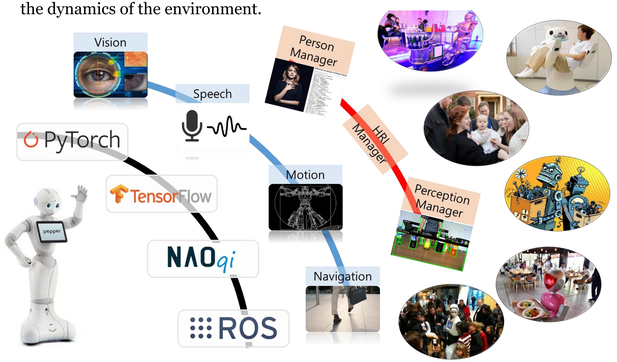
AUPAIR: A Personal Assistant Robot at Home

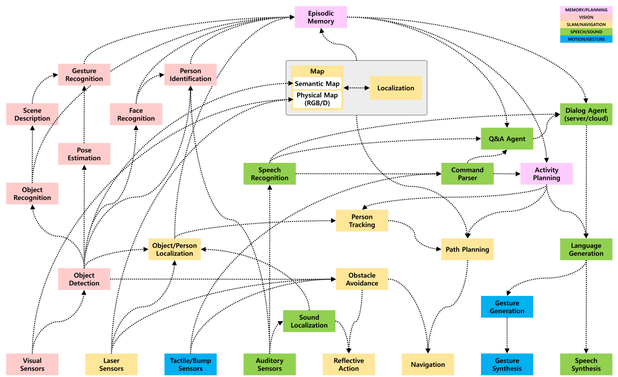
Chung-Yeon Lee, Kibeom Kim, Sungjae Cho, Seung-Jae Jung, Joonho Kim, Hyung-Suk Lim, Hyundo Lee,
Injune Hwang, JooHyun Cho, JunHa Chun, Min Whoo Lee and Byoung-Tak Zhang

http://aupair.ml

Overall Architecture

We aim to develop a new class of machine learning algorithms for mobile cognitive robots that actively perceive information by multimodal sensors and interacting with the real world environment. We work towards this goal by developing a multi-module integrated system for mobile robots to perceive information (objects, people, actions) from the environment, act (schedule, interact) according to the perceived information and building models that learn the dynamics of the environment.

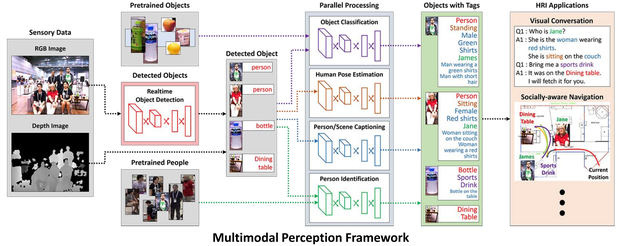




Overall System Architecture

Multimodal Perception

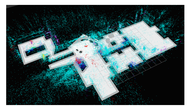
We developed IPSRO (Integrated Perception for Service RObots) framework, which is ROS-friendly integrated perception system. IPSRO can flexibly integrate several perception modules including deep learning models to extract rich and useful perceptual information from the environment based on a unified perception representation.

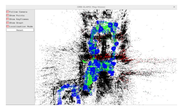


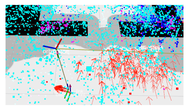
Multimodal Perception Framework

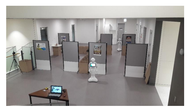
Visual SLAM

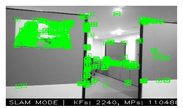
We applied a visual SLAM to make Pepper robot possible to overcome a fatal localization error and draw a more accurate map. The visual SLAM is adopted by customizing ORB-SLAM2 and the AMCL module in ROS navigation stack for our system and Pepper. Our visual SLAM system improved stability and performance of the Pepper robot's @home scenario execution in a complex environment.


Constructed Map w/ ORBs

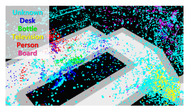

Cartography w/ Visual SLAM


AMCL Particle Resampling


Experimental Environment



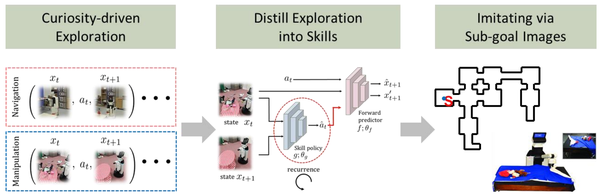
ORB Detection w/ RGB-D Camera



Semantic Space Labeling

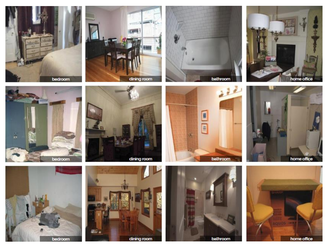
Visual Imitation Learning

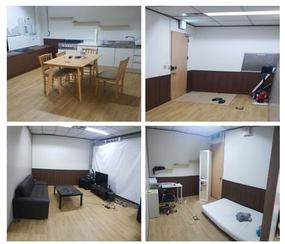
Recently, we begin to do our research about a visual imitation learning which a robot randomly explores the environment and then distills its own experience into a goal-conditioned skill policy. In this framework, the learned policy is employed to imitate the human expert after observing a visual demonstration. We plan to evaluate the visual imitation learning framework in two real-world settings: tidy a table using a HSR robot (Toyota) and navigation in previously unseen home environments using a Pepper robot (Softbank Robotics).



Embodied Spatial Cognition

One of our ultimate goal is an autonomous navigation based on visual cues in an environment where its map is not built yet. We try to model an embodied spatial representation by learning rooms, objects, and dialogues related to them.

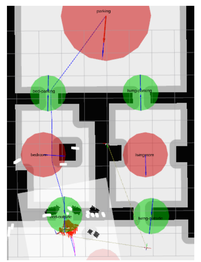

Matterport 3D Dataset (VR)

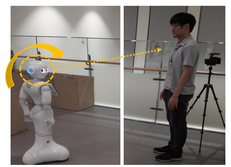

Our Experimental Model House (Real Space)

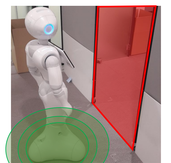
Experiments for Vision-and-Language Navigation (Embodied Q&A, Visually-Grounded Navigation)

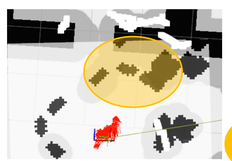
Robust Navigation

Developing particular action modules based on the perception results and Pepper's sensors, we could improve important abilities (e.g., tracking, navigation, approaching, guiding) for home robots to perform better services.


Path Planning based on Waypoints


Simultaneous Person Identification and Approaching


Robust Door Passing


Noisy Costmap Cleansing


Obstacle Detouring

References

- Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. arXivpreprint arXiv:1804.02767, 2018.
- Ejaz Ahmed, Michael Jones, and Tim Marks. An improved deep learning archi-tecture for person re-identification. CVPR, pp. 3908–3916, 2015.
- Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. CVPR, pp. 4724–4732, 2016.
- Justin Johnson, Andrej Karpathy, and Fei-Fei Li. Densecap: Fully convolutionallocalization networks for dense captioning. CVPR, pp. 4565–4574, 2016.
- Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. IEEE Transactions on Robotics, 33(5):1255–1262, 2017.
- Peter Anderson et al., Vision-and-Language Navigation: Interpreting visually-grounded navigation instructions in real environments, arXiv:1711.07280v3, 2018.
- Deepak Pathak et al., Zero-Shot Visual Imitation, ICLR-2018

- https://github.com/team-aupair
- https://github.com/tensorflow/tensorflow
- https://github.com/raulmur/ORB_SLAM2
- https://github.com/pjreddie/darknet
- https://github.com/Ning-Ding/Implementation-CVPR2015-CNN-for-ReID
- https://github.com/CMU-Perceptual-Computing-Lab/openpose
- https://cloud.google.com/speech-to-text/
- https://www.kairos.com/

Special Thanks to ...

과학기술정보통신부 Ministry of Science and ICT
IITP 정보통신기술진흥센터
VIC
SURROMIND ROBOTICS
국방과학연구소 Agency for Defense Development
SoftBank Robotics
AFRL
SAMSUNG