Project AIR: Developing Artificial Social Intelligence for Human-Care Robots

Minsu Jang
Electronics and Telecommunications Research Institute
Daejeon-si, South Korea
minsu@etri.re.kr

Jaehong Kim
Electronics and Telecommunications Research Institute
Daejeon-si, South Korea
jhkim504@etri.re.kr

Jaeyeon Lee
Electronics and Telecommunications Research Institute
Daejeon-si, South Korea
leejy@etri.re.kr

ABSTRACT

This paper introduces a research project code-named AIR aiming to develop artificial intelligence technologies for human-care service robots that can provide personalized socially assistive services for elderly people. The core issues to solve with the project include deep understanding of human attributes and their changes over longer period of time, automated learning and generation of socially intelligent multi-modal robot behaviors based on machine learning methodologies, building large-scale multi-modal datasets in the domain of elderly care, and integration of core technologies as a scalable and extensible framework. The project spans 5 years from 2017, and its main results, including datasets, shall be open-sourced.

CCS CONCEPTS

• **Computer systems organization** → **Embedded and cyber-physical systems**; • **Computing Methodologies** → *Artificial Intelligence*; *Machine Learning*;

KEYWORDS

Human-care Service Robot, Social Intelligence, Artificial Intelligence

ACM Reference Format:
Minsu Jang, Jaehong Kim, and Jaeyeon Lee. 2018. Project AIR: Developing Artificial Social Intelligence for Human-Care Robots. In *Proceedings of HRI 2018 WORKSHOP ON SOCIAL HUMAN-ROBOT INTERACTION OF HUMAN-CARE SERVICE ROBOTS (HRI 2018 WORKSHOP).* ACM, New York, NY, USA, 3 pages.

1 INTRODUCTION

Problems of aging society are attracting ever increasing attention in many countries as the elderly population is growing fast with accompanied social cost getting quickly higher. In South Korea, the number of older people in the ages greater than 65 reached 6.8 million in 2016 which is 13.6% of the country's total population [9]. It is estimated that more than 50% of them will live alone by 2030, with fragile mind and body. The purpose of AIR project is to mitigate the problems around aging society by providing robot AI technologies to help older people avoid getting isolated and lonely, developing cognitive impairment and losing health.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HRI 2018 WORKSHOP, March 2018, Chicago, IL, USA
© 2018 Copyright held by the owner/author(s).

The technologies should be viable to continuously monitor long-term changes in personal preferences, health conditions, daily activity patterns and social relationships to better understand the elderly people in various aspects and make qualified reports to families or dedicated care givers. Also, it should be possible to build social robots with the technologies that can build emotional relationships and share experiences with older people.

The flux of technical approaches in AIR to achieve the goals can be summarized as follows:

- **Detailed and Continual User Profiling**: modeling personal profiles of older people based on video streams from robots by detecting, tracking, recognizing biometrics, human attributes, daily activities and owning objects, and understanding their long-term changes.
- **Learning Social Intelligence through Observation**: modeling interaction dynamics by observing interpersonal communications, imitating multi-modal human social skills, and incrementally improving them through feedbacks from users.
- **Life-Modeling and Health Monitoring**: modeling patterns of daily living and detecting abnormal situations.
- **Building Large-Scale Elder-care Datasets**: Capturing multi-modal data from real-world living labs where older people actually live, and augmenting them by generating synthetic human and action data with a lot of parametric variations in the virtual environment.
- **Designing a Software Framework for Social HRI**: Integration of perception, deliberation and multi-modal action generation modules into a coherent system based on plausible cognitive theories.

In this project, we hypothesize that if robots succeed in establishing emotional and trusted *relationship* with users, it will improve sustainability and efficacies of the caring services provided by human-care roots to elderly people at home environments. According to Knapp's relational development model, interpersonal relationship reciprocally develops with the growth of knowledge about each other [4], which we suppose is also true for human-robot relationship development. User profiling is the process of gradually accumulating knowledge about a user's preferences and activities. With recent advances of computer vision technologies, we aim to recognize face, gender, facial expressions as well as more than 14 kinds of human attributes including clothing properties like colors, styles, patterns, accessories and facial make-up, 55 kinds of daily activities and more than 20 object instances. As such, user profiling in AIR contributes to understanding an elder user to the level of very small details, and provides multi-faceted cues for service personalization [3].

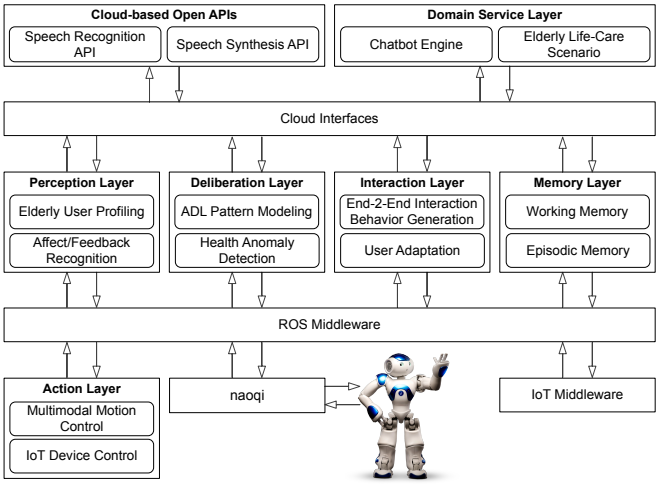

Figure 1: Subsystems of the AIR framework

Social intelligence modeling makes a robot automatically perform multi-modal communicative behaviors in a social and affective manner, with which elder users can communicate more effortlessly with robots. In our framework, social behaviors are not planned but generated according to end-to-end models that are trained with human-human social interaction data to map multi-modal input data to robot joint control sequences, which is pioneered in recent research efforts based on deep learning [5, 7]. This approach can improve responsiveness and adaptiveness of social behaviors and enables nuanced expressions upon variety of context changes.

Developing these wide variety of perceptual and generative technologies using recent AI techniques requires large-scale training data. Collecting datasets from elderly people is not feasible due to problems regarding ethics, safety and efforts. We plan to implement several living labs and collect real-life multi-modal data continuously throughout the whole project years, which will allow us to build large-scale training datasets in the domain of elder-care. Also we plan to employ human body modeling, motion animation and virtual environment to generate large variety of synthetic training data. We expect the datasets will constitute several hundreds of hours of annotated video clips.

2 SYSTEM ARCHITECTURE DESIGN

The AIR system architecture is designed as a hierarchical structure with 6 layers, interacting via ROS, Internet-of-Things and cloud middle-wares, as shown in Fig. 1.

The perception layer performs user profiling; the deliberation layer processes long-term user life modeling and anomaly detection; the interaction layer learns and generates multi-modal social behaviors; the action layer controls robots and IoT devices; the memory layer intermediates percepts and knowledge among modules and remembers episodic facts; the domain service layer delivers context-based dialogs and scenarios.

The current system is capable of maintaining short scenario-based dialogs with multi-modal perception, social context understanding with verbal/non-verbal motion replay.

The core modules in each subsystem and interactions among them are illustrated in Fig. 2.

2.1 Core Modules

We succinctly describe important core modules and show current performance test results.

2.1.1 Facial/Human Attributes Recognition. This module recognizes identity, gender and age by facial features. A light-weight noise-robust CNN model is employed [10] and three datasets comprised of 1.65 million images were used for training and testing. Overall performance is 98.2% for face, 92% for gender and 92.4% for age recognition.

This module also recognizes 9 clothing styles and 6 accessories. We collected 10,168 images of elder people in various clothings for training and testing a YOLO v2-based object detector [8], and the system performs with 60% mAP.

2.1.2 Non-Verbal Interaction Gesture Generation. This module generates communicative gestures by end-to-end mapping from visual stimuli to motor control [6]. Currently, we trained NAO robot to ring a bell that is positioned in an arbitrarily position in front of it. YOLO v2 was used to detect the position of the bell and LSTM-based sequence-to-sequence model was trained to generate a sequence of motor control commands to actuate NAO's body joints.

2.1.3 Co-Verbal Gesture Generation. This module generates co-verbal gestures by mapping input sentences to robot animations [1, 2]. The input sentence could be the output from a chatbot engine a robot should pronounce via text-to-speech. We collected and extracted 129 clips of TED videos and annotated them with synchronized subtitles and 44 co-verbal gestures. Later, each gesture class was mapped to a NAO animation for testing. LSTM-based encode-decoder model was trained with the data for generating a gesture label per input word. In a preliminary qualitative evaluation, more than 20% of the participants preferred gestures generated by the trained model.

3 SUMMARY AND FUTURE WORK

In this paper, we introduced a 1-year old project called AIR to build artificial intelligence for human-care service robots. With this project, we plan to build large-scale datasets with elderly people that can leverage advancement of the elderly-care technologies, and also machine learning-based techniques to enable more natural social interactions with robots will be explored. The main results of AIR shall all be open-sourced for the benefit of the community to solve the problem of aging society together.

ACKNOWLEDGMENTS

This work was supported by the ICT R&D program of MSIP/IITP. [2017-0-00162, Development of Human-care Robot Technology for Aging Society]

REFERENCES

[1] Henny Admoni and Brian Scassellati. 2014. Data-driven model of nonverbal behavior for socially assistive human-robot interactions. In *Proceedings of the 16th international conference on multimodal interaction*. ACM, 196–199.

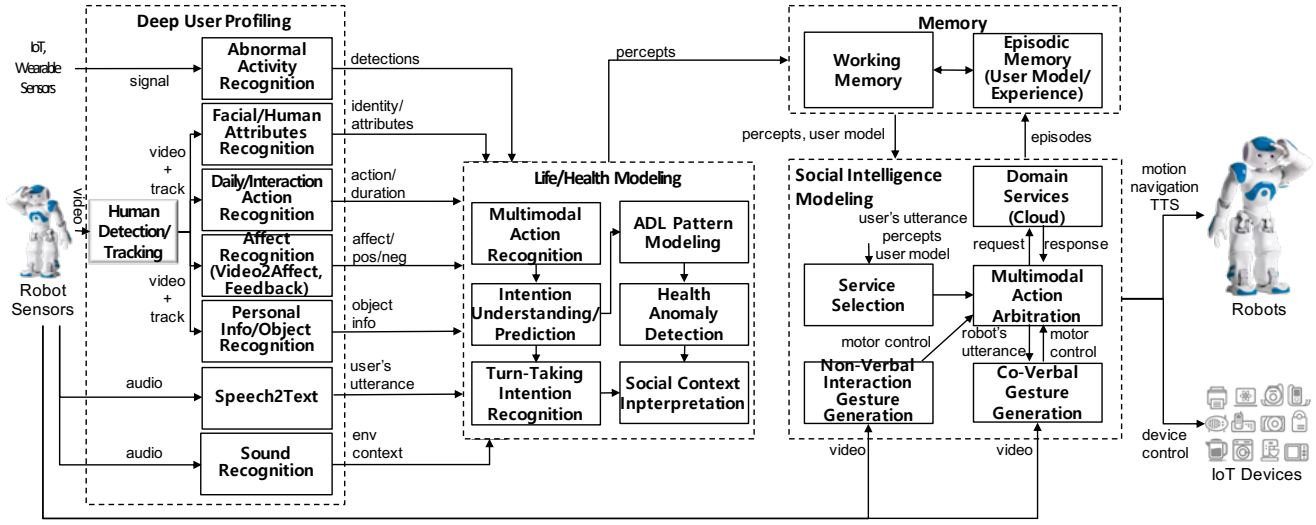

Figure 2: Core Modules and Their Interactions in the AIR framework

[2] Chien-Ming Huang and Bilge Mutlu. 2014. Learning-based modeling of multimodal behaviors for humanlike robots. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 57–64.

[3] Minsu Jang and Jaehong Kim. 2015. A relationship-based model of robot service personalization. In *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on*. IEEE, 192–193.

[4] Mark L Knapp, Anita L Vangelisti, and John P Caughlin. 2014. *Interpersonal communication and human relationships*. Pearson Higher Ed.

[5] Shingo Murata, Hiroaki Arie, Tetsuya Ogata, Shigeki Sugano, and Jun Tani. 2014. Learning to generate proactive and reactive behavior using a dynamic neural network model with time-varying variance prediction mechanism. *Advanced Robotics* 28, 17 (2014), 1189–1203.

[6] Kuniaki Noda, Hiroaki Arie, Yuki Suga, and Tetsuya Ogata. 2014. Multimodal integration learning of robot behavior using deep neural networks. *Robotics and Autonomous Systems* 62, 6 (2014), 721–736.

[7] Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2017. Show, attend and interact: Perceivable human-robot social interaction through neural attention Q-network. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 1639–1645.

[8] Joseph Redmon and Ali Farhadi. 2016. YOLO9000: Better, Faster, Stronger. *arXiv preprint arXiv:1612.08242* (2016).

[9] Isabella Steger. 2017. South Korea is aging faster than any other developed country. *Quartz Media LLC* (August 2017).

[10] Xiang Wu, Ran He, and Zhenan Sun. 2015. A lightened cnn for deep face representation. In *2015 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR)*, Vol. 4.