Social Human-Robot Interaction for the Elderly: Two Real-life Use Cases *

Athanasia Zlatintsi
nzlat@cs.ntua.gr

Isidoros Rodomagoulakis
irodoma@cs.ntua.gr

Vassilis Pitsikalis
vpitsik@cs.ntua.gr

Petros Koutras
pkoutras@cs.ntua.gr

Nikolaos Kardaris
nick.kardaris@gmail.com

Xanthi Papageorgiou
xpapag@mail.ntua.gr

Costas Tzafestas
ktzaf@cs.ntua.gr

Petros Maragos
maragos@cs.ntua.gr

School of Electrical and Computer Engineering
National Technical University of Athens, 15773 Greece

ABSTRACT

We explore new aspects on assistive living via smart social human-robot interaction (HRI) involving automatic recognition of multimodal gestures and speech in a natural interface, providing social features in HRI. We discuss a whole framework of resources, including datasets and tools, briefly shown in two real-life use cases for elderly subjects: a multimodal interface of an assistive robotic rollator and an assistive bathing robot. We discuss these domain specific tasks, and open source tools, which can be used to build such HRI systems, as well as indicative results. Sharing such resources can open new perspectives in assistive HRI.

Keywords

assistive HRI, multimodal audio-gestural recognition

1. INTRODUCTION

Social human-robot interaction is under an abrupt emerging reinvention. This is evident (1) by facts in the market e.g., 1.5bn in global robotics for 2019; (2) the research spanning a great range of cases e.g., social companions [1], to ones that deal with dementia [8] and disorders [3]; (3) as a consequence of the core machine learning advancements, with deep learning in natural language processing, computer vision and speech recognition, where rates have been doubled in tough tasks within a few years [7]. What is crucial for the above is the *data*. For recent deep learning approaches, domain specific datasets are the cornerstone for training.

Herein, we show our perspective on social human-robot interaction via a rich HRI set of resources including domain specific datasets and automatic machine learning tools. This refers to multimodal communication with speech and gestures, as applied on the assistive service robots for the elderly, in two real-life use cases: a) a robotic platform that supports the mobility and thus enforces fitness and vitality, see Fig. 1 (top), and b) an assistive bathing robot, which helps to perform and complete bathings tasks identified as difficult and stressful, see Fig. 1 (bottom). Both cases shall assist towards independent living for the elderly to improve their life quality. The automatic multimodal recognition on both cases is based on state-of-the-art algorithms and a suite of tools that can train audio-visual models and recognize, in an online manner, gestures.

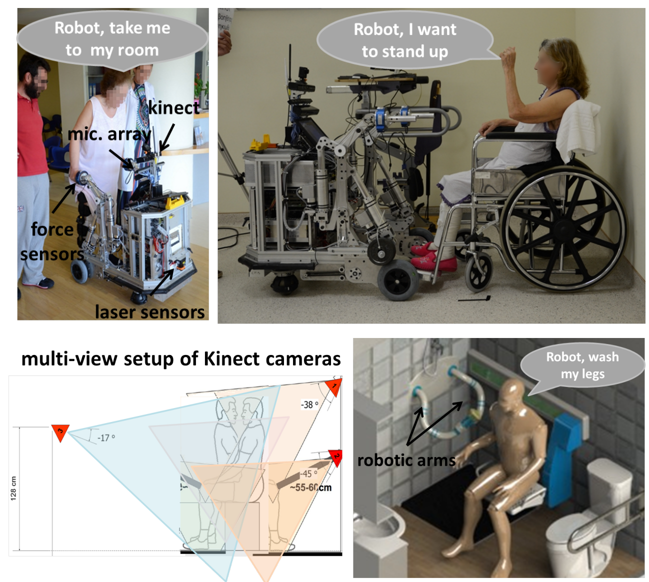

Figure 1: Interacting with smart robotic assistants for everyday activities: assistive robotic rollator (top); automated bathing environment (bottom).

The technological advances in assistive living have led HRI research to extend into various areas [4], such as multimodal interfaces. Even though, our everyday communication is a blend of various modalities, e.g., speech, gestures and eye gaze, co-speech gesturing is still quite limited in HRI [2].

*This research work was supported by the EU grants FP7-ICT-2011.2.1-600796 and H2020-643666.

HRI '17, March 6–9, 2017, Vienna, Austria.

ACM ISBN 978-1-4503-4336-7/17/03..

DOI: http://dx.doi.org/10.1145/3029798.3038400

Figure 2: Sample gestures for the two HRI tasks. Left: assistive robotic rollator; right: bathing task.

Our goal is to enhance the communication, making it natural, intuitive and easy to use, in other words enhancing it wrt social aspects. For a review, we refer the reader to [6].

2. TWO REAL-LIFE USE CASES

MOBOT[1]: The first use case includes an active mobility assistance robot for indoor environments aiming to support mobility and thus enforce fitness and vitality providing user-centered, context-adaptive and natural support. The developed experimental prototype shown in Fig. 1 (top) consists of a robotic rollator equipped with sensors such as: laser range sensors scanning the walking area for environment mapping and obstacle detection, detecting also lower limbs movement at the back; force/torque handle sensors; two Kinect sensors to record users' upper body movements and the lower limbs and an array of 8-microphone MEMS for audio capturing.

I-Support[2]: In the second use case the goal is to develop a robotic shower system in order to enable independent living for elderly so as to improve their life quality. The core system functionalities identified as important from a clinical perspective (taking into account impairments, limitations and user requirements) are the tasks for bathing the distal region and the back region [9]. The experimental prototype in this case includes three Kinect sensors, as shown in Fig. 1 (bottom), that reconstruct the 3D pose of the human and the robot, recognizing also user gestures and an audio system including 8 distributed condenser microphones.

Kinect cameras and microphones, in both cases, are used to record, challenging at times, audio-gestural data (i.e., actions for modeling the interaction between the users and the robot); keeping however in mind that elderly have to be able to perform and remember them regardless their cultural background. The collected domain specific audio-gestural commands are related to the specific tasks, including also actions for emergency situations. An online multimodal action recognition system has been used [5] to monitor, analyze and predict user actions in those challenging conditions, giving emphasis to the command-level speech and gesture recognition. Always-listening recognition, using state-of-the-art methods, is applied separately for spoken commands and gestures with their results combined at a second fusion phase. The output is fed to the robot's controller and the predicted action or task is executed.

For the first use case, experiments have been conducted on challenging data acquired with elderly users while interacting with the platform, using 8 gestures and German spoken commands (see Fig. 2), obtaining accuracies of 84.1%, 57% and 90.2% for the audio, visual modality, and their fusion, respectively. For the experiments of the two bathing tasks (i.e., washing the back and the legs) a small audio-gestural dataset was used, including 28 gestures and German spoken commands (see Fig. 2), accomplishing average accuracy of ca. 82% for gesture recognition and ca. 75% for spoken command recognition (see Table 1).

Modality	"washing the back"	"washing the legs"
Visual	83.9%	81.1%
Audio	75.3%	81.6%

Table 1: Recognition results for the two bathing tasks.

3. CONCLUSIONS

We presented two real-life use cases, tools and data. Such resources can be employed to develop natural interfaces for multimodal interaction. Our intention is to further investigate how the communication will be as intuitive as possible using co-speech gesturing, which is the most natural way for human-human communication, while also enhancing the recognition, in cases of speech dysfluencies or kinetic problems. In addition, multiple domain data could be combined to build a generalized dataset which in the future could be used to tackle challenging tasks were multimodality in interaction is in question. Finally, by sharing such resources, we aim to build a public crowdsourced library that shall open new perspectives in smart assistive HRI.

4. REFERENCES

[1] J. Broekens, M. Heerink, and H. Rosendal. Assistive social robots in elderly care: A review. *Gerontechnology*, 8(2), 2009.

[2] J. Cassell. *Computer vision in human-machine interaction, A framework for gesture generation and interpretation.* Cambridge Univ. Press, 1998.

[3] D. Feil-Seifer and M. J. Mataric. Automated detection and classification of positive versus negative robot interactions with children with autism using distance-based features. In *Proc. 6th ACM/IEEE Int'l Conf. Hum.-Robot Interact. (HRI-11)*, 2011.

[4] M. A. Goodrich and A. C. Schultz. Human-robot interaction: A survey. *Found. trends human-computer Interact.*, 1(3):203–275, 2007.

[5] N. Kardaris, I. Rodomagoulakis, V. Pitsikalis, A. Arvanitakis, and P. Maragos. A platform for building new human-computer interface systems that support online automatic recognition of audio-gestural commands. In *Proc. ACM on Multimedia Conf.*, 1169–1173, 2016.

[6] A. Kotteritzsch and B. Weyers. Assistive technologies for older adults in urban areas: A literature review. *Cognitive Computation*, 8:299–317, 2016.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. Adv. in Neural Inform. Process. Syst.*, 2012.

[8] E. Mordoch, A. Osterreicher, L. Guse, K. Roger, and G. Thompson. Use of social commitment robots in the care of elderly people with dementia: A literature review. *Maturitas*, 74:14–20, 2013.

[9] J. Werle and K. Hauer. Design of a bath robot system-user definition and user requirements based on international classification of functioning disability and health (ICF). In *Proc. RO-MAN*, 2016.

[1] http://www.mobot-project.eu/

[2] http://www.i-support-project.eu/