Social human-robot interaction for the elderly: Two real-life use cases

Athanasia Zlatintsi, Isidoros Rodomagoulakis, Petros Koutras,
Vassilis Pitsikalis, Nikolaos Kardaris, Xanthi Papageorgiou, Costas Tzafestas, Petros Maragos
ICCS, School of ECE, National Technical University of Athens, Greece

Social Assistive HRI

❑ Social human-robot interaction is under an abrupt emerging reinvention.

❑ 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.

❑ 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.

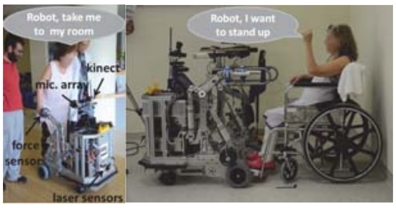
❑ This refers to multimodal communication with speech and gestures, as applied on the assistive service robots for the elderly, in two real-life use cases:

1. A robotic platform that supports the mobility and thus enforces fitness and vitality and
2. an assistive bathing robot, which helps to perform and complete bathings tasks identified as difficult and stressful.

❑ 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.

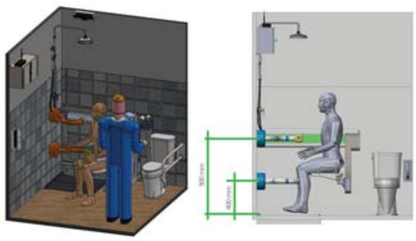
The MOBOT use-case

❑ The MOBOT 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 bellow 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



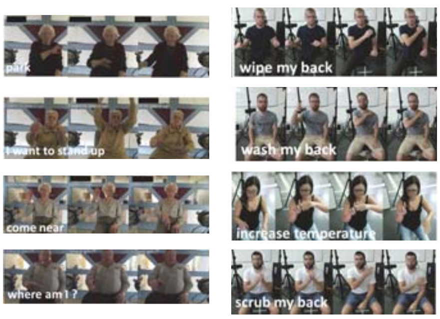
The I-Support use case

❑ 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 bellow, that reconstruct the 3D pose of the human and the robot, recognizing also user gestures and an audio system including 8 distributed condenser microphones.



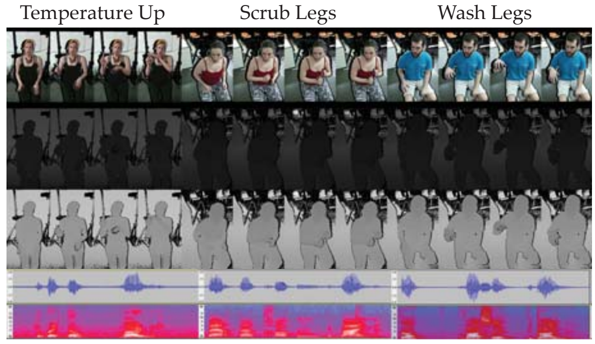
Data Examples and Vocabulary

MOBOT | I-Support



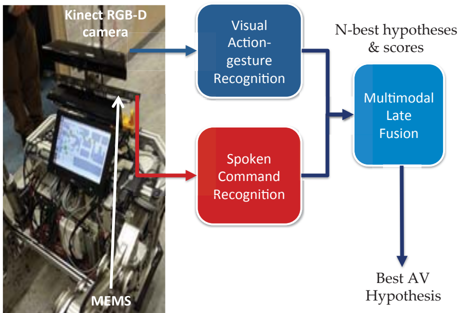
Sample gestures for the two HRI tasks. Left: assistive robotic rollator; right: bathing task.

First robotic prototype of the bathing robot arm.



Data streams acquired by sensors #1 and #2: RGB (top), depth (2nd row) and log-depth (3rd row) frames from a selection of gestures ("Temperature Up", "Scrub Legs", "Wash Legs", "Wash Back") accompanied by the corresponding German spoken commands (bottom) "Warmer", "Trockne die Beine", "Wasche die Beine", all preceded by "Roberta", which is the name of the assistive robot.

Overall system: audio-gestural command recognition

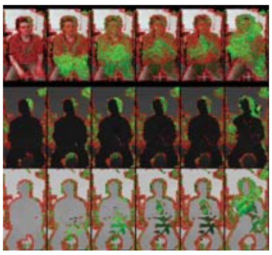


MOBOT robotic platform

Overview scheme of our multimodal interface

Visual processing for gesture classification

❑ Employ Dense Trajectories features along with the popular Bag-of-Visual-Words (BoVW) framework. The Dense Trajectories method [17] main concept consists in sampling feature points from each video frame on a regular grid and tracking them through time based on optical flow.



❑The employed descriptors are: the Trajectory descriptor, HOG [18], HOF [19] and Motion Boundary Histograms (MBH) [18]. A non-linear transformation of depth using logarithm (log-depth) enhances edges related to hand movements and leads to richer dense trajectories on the regions of interest, close to the result obtained using the RGB stream.

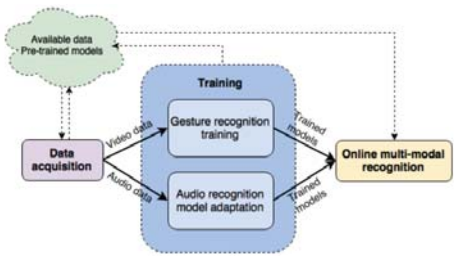
Comparison of dense trajectories extraction over the RGB (top), depth (middle) and log-depth (bottom) clips of gesture "Scrub Back".

How to build new multimodal gesture recognition systems

http://robotics.ntua.gr/projects/building-multimodal-interfaces
https://bitbucket.org/nkardraris/building_interfaces_multimodal_interaction

❑ Software platform for building interfaces that recognizes natural user input, as audio commands, manual gestures, captured by sensors such as kinect. Includes a component for acquiring multimodal user data, used as input to a training module. These models are employed in automatic online recognition recognition.

❑ The overall framework is demonstrated by a working system use case. Demonstrating the potential of the software platform, on building other new human-computer interaction systems. Users may populate libraries of models and/or data, that can be shared in the network. In this way users may re-use or extend existing systems.



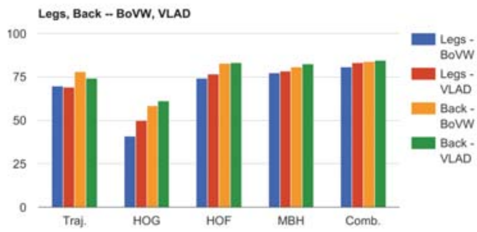
N. Kardaris, I. Rodomagoulakis, V. Pitsikalis, A. Arvanitakis and P. Maragos, A platform for building new human-computer interface systems that support online automatic recognition of audio-gestural commands, Proc. ACM Multimedia 2016, Amsterdam, The Netherlands, 15-19 Oct. 2016

Recognition results

❑ MOBOT: 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, obtaining accuracies of 84.1%, 57% and 90.2% for the audio, visual modality, and their fusion, respectively.

❑ I-SUPPORT: 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, accomplishing average accuracy of ca. 82% for gesture recognition and ca. 75% for spoken command recognition.

Modality	Washing the back	Washing the legs
Visual	83.9%	81.1%
Audio	67.7%	76.7%



Average classification accuracy (%) for the pre-defined gestures performed by 23 subjects. Results for the five different features using the two different encodings are shown for RGB data for the two tasks, i.e., "washing the legs" and "washing the back".

Conclusions

❑ We present 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.

❑ 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 crowd-sourced library that shall open new perspectives in smart assistive HRI.

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., 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.

Acknowledgments and further information

This research work was supported by the EU under the research program Mobot with grant FP7-ICT-2011-9 2.1–600796 and in part under the EU project I-SUPPORT with grant H2020-643666. Please contact [vpitsik, maragos,nzlat,pkoutras,ktzaf]@cs.ntua.gr, xanthi.papageorgiou@mail.ntua.gr, nick.kardaris@gmail.com. More information can be found at http://cvsp.cs.ntua.gr, http://robotics.ntua.gr, http://www.mobot-project.eu.