Data Acquisition towards Defining a Multimodal Interaction Model for Human – Assistive Robot Communication

Evita-Stavroula Fotinea[1], Eleni Efthimiou[1], Athanasia-Lida Dimou[1], Theo Goulas[1], Panayotis Karioris[1], Angelika Peer[2], Petros Maragos[3], Costas Tzafestas[3], Iasonas Kokkinos[4], Klaus Hauer[5], Katja Mombaur[6], Ioannis Koumpouros[7], and Bartlomiej Stanczyk[8]

[1] Institute for Language and Speech Processing/ATHENA RC, Athens, Greece
{evita,eleni_e,ndimou,tgoulas,pkarior}@ilsp.gr
[2] Technische Universität München, Munich, Germany
Angelika.Peer@tum.de
[3] Institute of Communication and Computer Systems–NTUA, Athens, Greece
{petros.maragos,ktzaf}@cs.ntua.gr
[4] INRIA–Ecole Centrale Paris, France
iasonas.kokkinos@ecp.fr
[5] Agaplesion Bethanien Hospital/Geriatric Centre at the University of Heideberg, Germany
khauer@bethanien-heidelberg.de
[6] Ruprecht-KarlsUniversität Heidelberg, Germany
kmombaur@uni-hd.de
[7] Diaplasis Rehabilitation Center S.A, Kalamata, Greece
ykoump@teiath.gr
[8] ACCREA Engineering, Lublin, Poland
b.stanczyk@accrea.com

Abstract. We report on the procedures followed in order to acquire a multimodal sensory corpus that will become the primary source of data retrieval, data analysis and testing of mobility assistive robot prototypes in the European project MOBOT. Analysis of the same corpus with respect to all sensorial data will lead to the definition of the multimodal interaction model; gesture and audio data analysis is foreseen to be integrated into the platform in order to facilitate the communication channel between end users and the assistive robot prototypes expected to be the project's outcomes. In order to allow estimation of the whole range of sensorial data acquired, we will refer to the data acquisition scenarios followed in order to obtain the required multisensory data and to the initial post-processing outcomes currently available.

Keywords: assistive robot, natural HRI, multimodal communication model, multisensory data acquisition.

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part III, LNCS 8515, pp. 613–624, 2014.
© Springer International Publishing Switzerland 2014

1 Introduction

Mobility disabilities are prevalent in our ageing society and impede activities important for the independent living of elderly people and their quality of life. The MOBOT project[1] aims at supporting mobility and thus enforcing fitness and vitality by developing intelligent active mobility assistance robots for indoor environments that provide user-centred, context-adaptive and natural support. The driving concept here envisions cognitive robotic assistants that act (a) proactively by realizing an autonomous and context-specific monitoring of human activities and by subsequently reasoning on meaningful user behavioral patterns, as well as (b) adaptively and interactively, by analyzing multi-sensory and physiological signals related to gait and postural stability, and by performing adaptive compliance control for optimal physical support and active fall prevention.

Towards these targets, a multimodal action recognition system is currently under development which needs to monitor, analyze and predict user actions with a high level of accuracy and detail. The main thrust of our approach is the enhancement of computer vision techniques with modalities such as range sensor images, haptic information as well as command-level speech and gesture recognition. In the same framework, data-driven multimodal human behavior analysis will be conducted and behavioral patterns of elderly people will be extracted. Findings will be imported into a multimodal human-robot communication system, involving both verbal and nonverbal communication and will be conceptually and systemically synthesized into mobility assistance models taking into consideration safety critical requirements. All these modules will be incorporated in a behavior-based and context-aware robot control framework aiming at providing situation-adapted optimal assistance to users.

In this framework, end user data become a crucial starting point for the design and implementation of the robotic platforms and also for the definition of the foreseen communication model. The recording sessions for the acquisition of the MOBOT multimodal sensory corpora took place in the rehabilitation centre Agaplesion Bethanien Hospital/ Geriatric Centre at the University of Heidelberg.

In the field of human-action recognition several datasets with rich sets of activities, complex environments representing real-world scenarios have been published. Such datasets that combine video and mocap systems in a systematic way by collecting synchronized and calibrated data include the HumanEva I and II datasets [1]. The creation of HumanEva datasets were motivated mainly by the need for having ground truth that can be used for quantitative evaluation and comparison of both 2D and 3D pose estimation and tracking algorithms. Although the HumanEva datasets have been extensively used in establishing the state-of-the art in human action recognition, their application areas remain limited to evaluation of 2D and 3D motion and pose estimation based on video and mocap data only. There are a number of other multimodal datasets that enhance the standard mocap-video data with additional modalities, such as magnetic sensors or microphones. The TUM Kitchen Dataset [2], which consists of activities in a kitchen setting (i.e., subjects setting a table in different ways), for example includes also RFID tag and magnetic sensor readings in addition to the

---
[1] `www.mobot-project.eu/`

multi-view video and mocap data. Similarly, the CMU Multimodal Activity (CMU-MMAC) Dataset [3] contains multimodal measures captured from subjects performing tasks such as meal preparation and cooking. The set of modalities utilized in this dataset is rather comprehensive, consisting of video, audio, mocap, internal measurement units (i.e., accelerometers, gyroscopes and magnetometers) and wearable devices (i.e., BodyMedia and eWatch). These two datasets are the first examples of publicly available multimodal datasets with a rich selection of various modalities. Finally, the Berkeley Multimodal Human Action Database (MHAD) is currently the only to-date dataset that systematically combines multiple depth cameras with multi-view video and mocap that are geometrically calibrated and temporally synchronized with other modalities such as accelerometry and sound [4]. The specific dataset consists of multi-view video, depth and color data from multiple Kinect cameras, movement dynamics from wearable accelerometers and the accurate mocap data with the skeleton information. In addition, ambient sound during the action performance was recorded and synchronized to reveal discriminative cues for human motion analysis.

2 MOBOT Database: Sensors Used and Types of Data Collected

The MOBOT database was acquired by means of a sensorised passive rollator (Fig.1) comprising multimodal input from (i) laser range finder sensors, (ii) force/torque sensors, (iii) RGB and RGB-D cameras and (iv) microphones. In addition a motion capture system was used to record human limb movements as well as the rollator and subject's absolute positions in space.

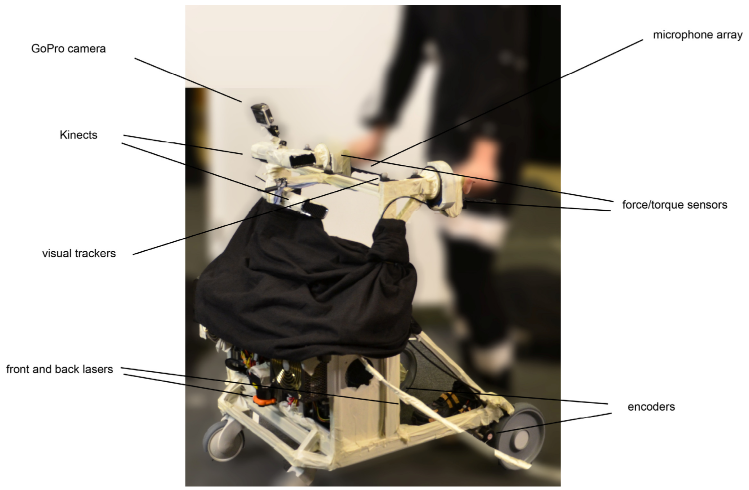

Fig. 1. Instrumented rollator

A diagram of the data acquisition setting implemented during the recording/measurement sessions that took place in Bethanien/Agaplesion Geriatric Hospital is presented in Fig. 2, while details on sensors employed are next provided.

2.1 Laser Range Finders

Two laser range-finder sensors were mounted on the passive rollator platform: One laser scanning range-finder sensor of type Hokuyo UTM-30LX was mounted at the front of the rollator platform facing towards the motion (normal walking) direction, to provide full scanning angle of the walking area. This sensor, also known as Hokuyo Top-URG, provides a 30m and 270° scanning range, and is suitable for robots with high moving speed because of its long range and fast response (~40 scans per second). One laser scanning range-finder sensor of type Hokuyo UBG-04LX-F01 was mounted at the back of the rollator platform facing the user legs, scanning a horizontal plane at the lower limbs (below knee) level, aiming to provide data on the gaiting of the user during typical walking operation with the rollator. This sensor, also known as Hokuyo Rapid-URG, provides a detection range of up to 5m with a 28 msec/scan time, and has a superior accuracy of +-10mm in short range (up to 1m) measurements.

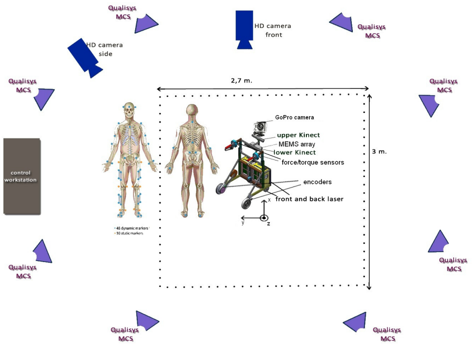

Fig. 2. Diagram of the MOBOT data acquisition setting in Bethanien Geriatric Hospital

2.2 Force/Torque Sensors

Two 6 DOF HR3 force/torque sensors of type JR3 45E15 were placed at the handles of the sensorised passive rollator. They are characterized by a force measurement range of 400N in x and y direction and 800N in z-direction as well as 50Nm around x and y and 500Nm around the z axis.

2.3 RGB and RGB-D Cameras

HD Cameras. For the recording of the scenarios and the required data collection four High Definition technology cameras with sensitive sensors were used. Three of the cameras were mounted on tripods and were static.

The central camera was placed as to record the patient when walking throughout the recording area. The aim was to record the whole body or the largest possible part of it. However, the presence of the rollator in most scenarios has hidden the posture of the torso and legs. So, it was necessary to place another camera in front and at an angle (side) to cover any optical gaps and provide further information of motion and posture of the patient as well as details of maneuvering and possible human interaction with a carer. Moreover, this camera could capture any difficulties in walking or causes in case of stumbling.

An HD camera (GoPro) was mounted on the passive rollator. The main criterion for the choice of this particular camera was the ability to record close and at a constant distance the patient's torso, arms and all its movements and in some cases also the head. Because of its small size and weight and its wide angle lens it was placed on top of the upper Kinect camera (for details see next subsection) mounted on the rollator. Due to the obligatory presence of the expert team in the field of view of the cameras during data capturing, a fourth HD camera was added to avoid or eliminate "visual noise". This camera was always on one side to supplement and give information that was missing or that was not visible from the other cameras. The location was not predefined and its position changed in-between the different scenarios so as to be placed in the optimum position in order to give the best viewing angle.

Kinect Cameras. Before data acquisition, the acquisition team experimented with different sensor choices and sensor placements, with the goal of placing the sensors in a configuration that achieves the broadest possible coverage of the human body, while being at a short distance from it. In particular position constraints did not allow placement of the sensors at a distance larger than 60 centimeters from the typical body position, as this would make the platform difficult to manipulate by elderly users. We therefore focused on solutions that would allow recording the human body from a short distance (Fig.3). We converged on using two Kinect-for-Windows (KFW) sensors, that are equipped with the ‘near mode’ option, which is absent in the more common Kinect-360 sensors. A considerable amount of effort was invested in order to obtain consistent data streams from the two KFW. We opted for placing the two Kinect sensors to be facing towards complementary directions so as to achieve broad coverage and to avoid interference between the two Kinects. The first sensor is facing horizontally towards the patient, aiming at capturing the area of the torso, waist, hips and the upper part of the limbs. The second sensor is facing downwards, capturing the lower limb motion information, with the aim of enabling the estimation of 3D limb positions and eventually also the analysis of gait abnormalities.

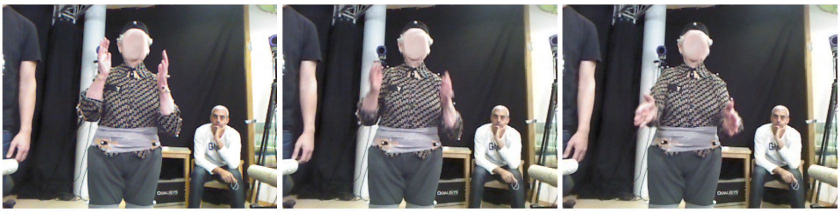

Fig. 3. RGB stream from Kinect

2.4 Microphone Array

As audio capturing device a microphone array has been chosen. Microphone arrays– instead of single sensors– are currently being explored in many different applications, most notably for sound source localization, beam-forming, and far-field speech enhancement and recognition. For data collection purposes, an 8-microphone MEMS array was mounted on the horizontal bar of the MOBOT rollator in a linear configuration (with a 4cm uniform spacing) in front of the user.

2.5 Motion Capture System

For motion capture a Qualisys system[2] was used, with eight cameras mounted on tripods around the recording area. Passive reflective markers were installed on the bodies of patient and carer to measure their human limb movements as depicted in Fig.4. Several limiting factors regarding the informants' population as well as supporting areas on the human body which should stay free of markers were taken into account. Two additional markers were placed on the head of the carer to identify his/her role. Visual markers were also placed onto objects like the door, the door frame, the rollator and the obstacle used in the recordings setting.

3 Multimodal Sensory Corpus Collection

3.1 Introduction: Aims, Scope

Data acquisition sessions were planned and organized to serve the corpus creation goal. In order to secure that the recording sessions would provide all necessary information for the post processing analysis, the scenarios to be recorded were carefully designed and executed. Scenarios include actions, gestures and other close to real life situations such as obstacle avoidance, interaction with other persons, simple everyday life operations (open/close door, switch on/off switch etc.) that needed to be reproduced by the participants in the recordings.

The MOBOT corpus creation was based on the grounds of human action recognition and context-aware robot control. The identified need was for data acquisition that would promote, as much as possible, natural interaction between (elderly) users and mobility aids along with the assistance or not, in some cases, of carers. In order to obtain the multi-sensory recordings, several recording scenarios had to be put in action and tested so that a maximum set of actions and movements would be well represented in the recordings and hence provide significant input from different modalities.

In the recordings, elderly human individuals (of varying age, gender, motor and cognitive abilities) performed a variety of assistance requiring cues in dialogues involving human carers and the passive rollator. Information on the individuals that participated in the recordings, the performed scenarios and their relation to the different modalities and combinations used in natural communication (language-based interaction, gesture-based interaction, multimodal instruction, but also silence in combination with specific body postures) as well as more technical aspects of the recordings such as settings and limitations, is provided in the following sections.

[2] http://www.qualisys.com/

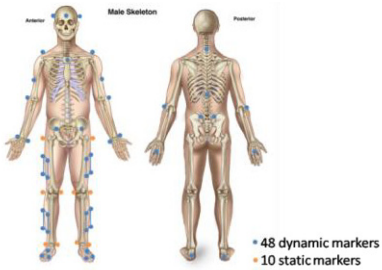

Fig. 4. Used marker set for recordings

3.2 Recording Scenarios

Aims and Scope. The recording scenarios are the outcome of careful design, initially based on an extended pre-described use case list of actions that a user might need to perform in real life situations [5]. The use cases list of actions defines goal-oriented sets of interactions between users and the system. In view of the recording sessions the use case list of actions had to be narrowed down as well as adapted to the recording environment. The envisioned goal was to incite a variety of actions, movements, gestures and functions within the restricted recording setting, in order to obtain enough and representative data as these would set the foundation for the technological development of the assistive device. Previous experience on the design of multimodal corpora [6-7], and multisensory acquisition data [8-9] was proven a valuable background asset, scientifically as well as methodologically.

Use Case Adaptation to the Actually Produced Scenarios. The use case list of actions that defined the goal oriented sets of interactions between users and the system functioned as a super set of real life tasks and/or situations that range from a single-short movement to repetitive long-lasting movements. The idea lying behind this list of action use cases is that mobility-assistants should support moving in different rooms, walking to various spaces as well as to stimulate activity through simplifying and reducing physical and cognitive load. Mainly, these action use cases were based on frequent situations while walking, partly including transfer situations, such as the sit-to-stand movement that is highly frequent. Critical situations or potential barriers due to insecurity, cognitive or physical impairment, as well as the risk of falling – which even though occurs rarely, it is crucial for the importance and usefulness of the device to the user– were taken into account.

A subset of this extensive action use case list [5] was implemented for the scenario creation that would be used in the recording sessions; the chosen action use cases were those that fulfilled all necessary criteria regarding the safety of the informants and were most suitable to technical specifications that the recording environment imposed. In some cases modifications in the course of the scenarios had to be performed leading to simpler acquisition versions of the scenarios than the extremely rich ones originally designed; this was done mainly in order to minimize the fatigue effect of the participants that were of advanced age and hence eliminate to a certain extend the risk of less participation or incomplete sets of recording sessions. For the above mentioned reasons, and in order to acquire data of actions important for all sensorial devices, several conventions were adopted:

1. Three types of variants in each scenario were proposed:

   a. Assisted by a carer
   b. Assisted by the passive rollator only &
   c. Unassisted

2. The number of trials per scenario and per variant presented variation, as there was a minimum set of repetitions required –one or two in most cases– and a maximum set, that was desired and was undertaken only in rare cases, depending on the patient's condition.

In all scenarios performed, data capturing took place from all sensors mounted to the passive rollator. However, in the design of each scenario a specific set of action data was targeted giving priority to different sensors at time. In the descriptions that follow, we focus only on the most important sensorial data with respect to the performed task's goal, while detailed descriptions (scenarios, variants, tasks contained, number of trials etc) can be found in [10].

Scenario 01 was designed to gather information regarding transfers "sit-to-stand" and "stand-to-sit", with or without the use of the sensorised passive rollator, enabling the acquisition of data from the force-torque sensors mounted on the handles of the rollator as well as the capturing data from the Kinect cameras.

Scenario 02 was designed to gather information regarding the gait patterns of the patients while they were walking with the rollator and the force-torque applied to it; the patient used the rollator while walking straight with a constant velocity and then was asked to maneuver while moving back to the start position. Data were captured from the sensors mounted on the handles, the high definition cameras, the Kinect cameras as well as from the infrared cameras placed around the area.

In **Scenario 03** the patient had to perform sit-to-stand transfers and maneuvers in order to avoid an obstacle placed in the testing area, while accelerating and decelerating his/her gait pace. Variants were included to incorporate several audio-gestural commands for communication with the rollator or with the carer, different maneuvering manners, and use or not of the sensorised passive rollator.

Scenario 04 was designed to capture the interaction of the patient while opening, closing a door and passing through a door passage with the use of the rollator. Data were captured by the high definition video cameras and by a camera that was placed on the side of the door to capture the patient's door passing. The total interaction of the patient with the rollator and door were captured by the Kinect cameras as well as the infrared cameras placed around the whole recording area.

Scenario 05 was designed to gather information by closely observing the patient performing a manipulation task that occurs in his daily life, like handling a switch (switching on/off). Data were captured by two separate high definition cameras to have an overview of how the patient behaves in a daily task and how he/she interacts with carer or the environment. The data captured by the Kinect cameras were targeting a. the gait of the patient (lower Kinect) and b. any performed gestures (upper Kinect camera). The microphone array was capturing the verbal commands. For this kind of scenario a manipulation task could only occur if the patient either was using only one hand to grasp the rollator or not holding the rollator at all. In all scenarios, the mounded sensors on the handles of the rollator were capturing force and torques that came from the patients. An indicative example of external and on rollator cameras' views of Scenario 05, can be viewed in Fig. 5.

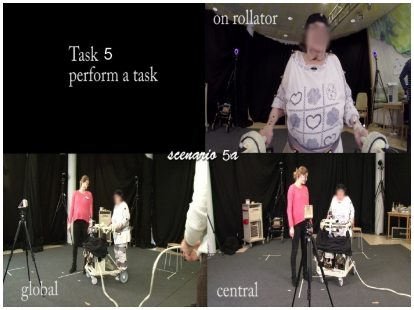

Fig. 5. Scenario Five: Performing a task. Views from external HD cameras and on rollator one.

Scenario 06 was designed to gather information from isolated gestural and verbal commands in order to train a model of human-robot communication. It included audio and gestural commands, uttered simultaneously, repeated a number of times, preferably having the patient in both sitting and standing position.

3.3 Patient-Subjects: Recruitment Strategy and Metadata

As regards patient-subjects' recruitment, the original goal was to recruit at least 10 patients, this being the lowest participation limit, yet opting for at best 15 patients. Due to high commitment and additional recruitment strategy procedures, our final subjects' number was raised to 18 patients. This was the result of a recruitment strategy which consisted into increasing the recruitment potential of the rehabilitation wards by including an associate sport club and by screening patients in acute care geriatric hospital. All participants met the inclusion criteria as all were acute patients in the rehabilitation wards either in the past or will be admitted in the rehabilitation centre in the near future. Inclusion criteria can be found in [5]. In actual figures, altogether 354 persons were screened and contacted within a two week assessment period. The screening process included a personal contact, consulting the patient's charts and contacting care or therapeutic personnel. In case a patient met the inclusion criteria, he/she was informed on the testing and a written consent was asked on his/her behalf, while a personal contact to relatives was also necessary. Participants' metadata consist of information about the gender, age, height, weight and knee height of each subject as well as his/hers cognitive and mobility score (with the aid of the diagnostic tool MMSE

[11-12] and their subsequent classification into a cognitive and mobility category) as detailed in [10]. We briefly report here that both genders have participated in the recordings, 5 male and 12 female subjects, the age range being 74-87 years old.

3.4 Corpus Quantitative Data

Synchronization of multimodal data streams was achieved by recording all the data to a ROS-Robot Operating System[3] bag. For this purpose ROS nodes were programmed for the two laser range finders, the two Kinects and the microphone array. Since the two force/torque sensors and the two wheel encoders of the rollator will be part of the real-time code of the overall control architecture to be realized, and given that low-level drivers for reading these sensor signals from the respective I/O cards were already available as blocks in Matlab/Simulink, a slightly different procedure was adopted for the recording of this type of data: An additional Simulink s-function block was programmed based on roscpp and tlc code to publish ROS topics directly from Simulink after having compiled the diagram using the prt.tlc target implementing a Preempt-RT real-time executable. This allowed seamless integration of the force/torque sensor and encoder data into the overall ROS bag. Motion capture data from Qualisys was recorded in raw format as it requires significant post-processing. Data recordings of this data type were synchronized with the rest of the recordings via a digital synchronization signal output in TTL format from the Qualisys system. In doing so, a cable was used to connect the synchronization output of one of the Qualisys cameras to the I/O card of the rollator. The synchronization signal was read using the respective Simulink block and passed further to the overall ROS bag via the implemented publisher block. Motion capture data will be post-hoc added to the ROS bag as soon as data has been post-processed.

A special launch file was written in ROS to start the different ROS nodes and thus, also the publishing of sensor data. In order to start recording from all the available sensor data topics, a special program was written, which first opens the respective ROS bag, subscribes to the different ROS topics and then waits for a trigger signal to start recording to the ROS bag. This was necessary as opening a ROS bag and subscribing to topics takes several seconds of time, which would have impeded an immediate storage of the data when the trigger signal arrives. The recording was started with a trigger button, which activated the Qualisys recording. As at the same time a synchronization signal was send from Qualisys to the ROS system, also the recording of the ROS bag was triggered and started at the same time.

The amount of ROS bag data collected altogether summarises up to 1.3 TB. The amount of data collected from the two HD cameras and the GoPro camera mounted on top of the passive rollator is approximately 250 GB (437 files), whereas a rough estimation of video data duration is 10 hours for global video duration and 8 hours for useful one (not containing preparations time, meaningless poses etc). Exact quantitative data will be available after post-processing of the acquired corpus, which is required prior to annotation.

---

[3] www.ros.org/

4 Data Post Processing

In order to end up with a corpus of properly annotated video data, the audiovisual materials from the cameras are initially being post-processed to ensure synchronization of video streams from the different capturing devices, prepare data for storage, and also eliminate possible problems or recover defects created during capturing.

Synchronization involves manual video editing to rework the streams and define their common starting point. The streams are being rendered independently and together in a single stream "picture in picture" (PiP) to have all the information accumulated (Fig. 5). Video processing consists mainly in adjusting brightness, especially regarding the camera on rollator and muting noise and external sounds, while amplifying and normalizing useful sounds and speech interactions. There is also an initial need for cleaning the data from visual artifacts due to reflections prior to annotating the markers in Qualisys, followed by the actual post-processing procedure that will allow for the creation of the human biomechanical model in Visual3D. Finally, a common naming convention of the individual files has been adopted to allow for a concise organization of the acquired corpus.

For the annotation process compressed files to mp4 will be used. The annotation of the visual data will be performed in the ELAN environment (ELAN 4.6.2[4]), being an annotation environment specifically designed for the processing of multi-modal resources [13]. Annotation is time aligned; each channel of information will be annotated into a separate annotation tier which may consist of several sub tiers according to the level of fine-grained information that is needed. The output of the annotation procedure is exported into .xml files. A preliminary inspection of the audio-visual data made necessary for the creation of at least 5 different major annotation tiers describing the scenario, the predefined tasks in each scenario, the actually performed actions, information from the audio channel (noise, oral commands), information form the visual channel (noise, gestures, pauses, stumbling etc).

5 Conclusion

Multi-modal sensorial data is a fundamental prerequisite for defining an effective human-robot communication model when developing a multimodal action recognition system. Here we reported on creation of the MOBOT dataset, which enhances the state-of-the art in the field of human-action recognition dataset creation with rich sets of activities in complex environments representing real-world scenarios.

Acknowledgements. The work leading to these results has received funding from the European Union under grant agreement n° 600796. Special thanks to the volunteers Bethanien patients for their participation in the recording sessions and also to the in

[4] `http://tla.mpi.nl/tools/tla-tools/elan/`, Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands.

situ recording team, namely Phoebe Kopp-BETHANIEN, Panagiotis Karioris-ATHENA RC, Theodore Goulas-ATHENA RC, Milad Geravand-TUM, Khai-Long Ho Hoang-UHEI, and Davide Corradi-UHEI.

References

1. Sigal, L., Balan, A., Black, M.: HumanEva: Synchronized video and motion capture data-set and baseline algorithm for evaluation of articulated human motion. Int. Journal of Comput. Vis. 87, 4–27 (2010)
2. Tenorth, M., Bandouch, J., Beetz, M.: The TUM Kitchen Data Set of everyday manipulation activities for motion tracking and action recognition. In: IEEE Int. Conf. on Computer Vision Workshops (ICCVW), pp. 1089–1096 (2009)
3. De La Torre, F., Hodgins, J., Montano, J., Valcarcel, S., Forcada, R.: Macey. J.: Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) database. Technical Report CMU-RI-TR-08-22, Robotics Institute, Carnegie Mellon University (2009)
4. Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., Bajcsy, R.: MHAD: A Comprehensive Multimodal Human Action Database. In: Workshop on the Applications of Computer Vision, pp. 53–60 (2013)
5. MOBOT's Deliverable D5.1- Preliminary report on use cases and user needs
6. Matthes, S., Hanke, T., Regen, A., Storz, J., Worseck, S., Efthimiou, E., Dimou, A.-L., Braffort, A., Glauert, J., Safar, E.: Dicta-Sign – Building a Multilingual Sign Language Corpus. In: Proc. of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon (LREC 2012), Istanbul, Turkey (2012)
7. Matthes, S., Hanke, T., Storz, J., Efthimiou, E., Dimou, A.-L., Karioris, P., Braffort, A., Choisier, A., Pelhate, J., Safar, E.: Elicitation Tasks and Materials designed for Dicta-Sign's Multi-lingual Corpus. In: Proc. of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon (LREC 2012), Istanbul, Turkey (2012)
8. Wallraven, C., Schultze, M., Mohler, B., Vatakis, A., Pastra, K.: The POETICON enacted scenario corpus - a tool for human and computational experiments on action understanding. In: Proc. 9th IEEE Conference on Automatic Face and Gesture Recognition (2011)
9. Pastra, K., Wallraven, C., Schultze, M., Vatakis, A., Kaulard, K.: The POETICON Corpus: Capturing Language Use and Sensorimotor Experience in Everyday Interaction. In: Proc. of LREC 2010, 7th Int'l Conference on Language Resources and Evaluation (2010)
10. MOBOT's Deliverable D2.1- Data acquisition and multimodal sensory corpora collection
11. Folstein, M.F., Folstein, S.E., McHugh, P.R.: Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. J. Psychiatry Res. 12(3), S.89–S.98 (1975)
12. Woodford, H.J., George, J.: Cognitive assessment in the elderly: a review of clinical methods. Q. J. Med. 100, 469–484 (2007)
13. Brugman, H., Russel, A.: Annotating Multimedia/ Multi-modal resources with ELAN. In: Proc. of LREC 2004, 4th Int'l Conference on Language Resources and Evaluation (2004)