

Getting to Know your Robot Customers: Automated Analysis of User Identity and Demographics for Robots in the Wild

Roberto Pinillos Herrero¹, Jaime Pulido Fentanes², Marc Hanheide²

Abstract—Long-term studies with autonomous robots “in the wild” (deployed in real-world human-inhabited environments) are among the most laborious and resource-intensive endeavors in Human-Robot Interaction. Even if a robot system itself is robust and well-working, the analysis of the vast amounts of user data one aims to collect and analyse poses a significant challenge. This paper proposes an automated processing pipeline, using state-of-the-art computer-vision technology to estimate demographic factors from users’ faces and re-identify them to establish usage patterns. It overcomes the problem of explicitly recruiting participants and having them fill questionnaires about their demographic background, and allows to study completely unsolicited and non-primed interactions over long periods of time. The paper offers a comprehensive assessment of the performance of the automated analysis with data from 68 days of continuous deployment of a robot in a care home, and also presents a set of findings obtained through the analysis, underpinning the viability of the approach.

I. INTRODUCTION

Despite impressive progress and a stronger focus on real-world application Human-Robot Interaction (HRI) studies with robots “in the wild” are still rare and pose significant challenges in terms of the analysis and conclusions to be drawn from them. These challenges stem quite often from the lack of information about users that are not recruited in the context of a specifically designed study, but are spontaneously interacting with a robot that is deployed and available for them. When robots are deployed for long periods of time in public or semi-public spaces, the subjects are usually unknown up front. But, it is often precisely those users that spontaneously and without explicit encouragement interact with a deployed autonomous robotics system, that could provide the most insights into the usability and the specific usage patterns of such robots. A key factor to systematically analyse the interaction patterns of a deployed robotic system are the demographics of the users interacting with the system and bringing these in relation to the way they interact with the robot. However, explicitly asking these data from interacting users can be cumbersome or very expensive. E.g., when a robot is indeed deployed 24/7 in an environment for many weeks, observational studies are very uneconomic and acquiring data from users after they have spontaneously

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 732737 (IL-IAD) and the Spanish Ministerio de Economía, Industria y Competitividad, fundamental research project ref. DPI2014-56500-R.

¹Roberto Pinillos is with CARTIF Foundation, Valladolid, Spain robertopinillos@astibot.es

²Jaime Pulido and Marc Hanheide are with Lincoln Centre for Autonomous Systems, School of Computer Science, University of Lincoln, UK {jpulidofentanes, mhanheide}@lincoln.ac.uk



Fig. 1. SCITOS G5 mobile robotic platform at the care home.

interacted with a robot near impossible. In many cases, ethical and data protection constraints can also make such a data acquisition exercise impossible, which would require individual consent from each participant. However, general approval to record limited data in a given environment from a robot’s perspective, stored for a restricted period of time and accessible to a defined set of experimenters is often much more easily obtained. The resulting datasets are usually huge, relatively unstructured, and expensive or impossible to make sense of when it comes to a systematic analysis.

With recent advances in data science and image understanding, however, an automated analysis of large-scale datasets in HRI can provide an alternative to laborious manual annotation and analysis, and can even provide information about users that is not easily obtained in a post-hoc analysis from recorded datasets.

Addressing this challenge and exploring the opportunities it provides, in this paper, a novel processing pipeline is proposed, using state-of-the-art face analysis technology to estimate demographic information of its users. The proposed processing pipeline comprises a person detection and tracking system that runs online on an autonomous robot in order to identify situations where users are approaching the robot to interact. This is followed by automated facial analysis to extract the identity of re-occurring users and demographic information such as age and gender fully automatically.

a) Long-term deployment of an info-terminal robot:

The proposed approach is analysed in the context of a mobile robotic system (see Fig. 1) that has been deployed for a total of 127 days in a care home for older adults, readily available for any person wishing to interact with it, without an experimenter or technician on site. The analysis in this

paper is focused on data gathered over a period of 68 days for which data recording was enabled in the system, comprising trajectories of people in the vicinity of the robot and video recording of persons standing in front of the robot while potentially interacting with it via a touchscreen interface. The robot (among other tasks) acted as a mobile info-terminal, providing potential users with information such as the menu of the canteen, the current weather and local news [1]. The automated analysis is based on a total of more than 24000 detections of persons in the robot’s vicinity, leading to an analysis of 600 successful interactions attributed to 334 different persons, all of whom where unsolicited interactions that occurred spontaneously as a result of the presence of the robot in the environment.

b) *Contributions:* This paper makes three distinct contributions to the field of Long-Term HRI:

- 1) A comprehensive critical appraisal of automatic estimation of demographic and user identification using the proposed pipeline based on state-of-the-art, readily available facial analysis technology. The proposed pipeline is available as open-source, integrated with the ROS framework, and released to the community as binary Ubuntu packages¹.
- 2) Identification and analysis of unsolicited usage patterns and demographic data of users automatically obtained for the long-term dataset recorded in a care home environment, providing insights into the use of such a system, covering interactions that occurred over a period of 68 days as a case study.
- 3) The underpinning dataset is released², for the community to further analyse the data and draw conclusions.

The paper is structured as follows: Sec. II contextualises the presented work in the scientific context. In Sec. III, the technology underpinning of the mobile robotic system and the dataset originating from its long-term deployment are being presented. This is followed by the presentation of the proposed processing and analysis pipeline in Sec. IV. The pipeline is critically appraised in Sec. IV-D, and then used to gain insights into the use of the long-term deployed system in Sec. IV-E.

II. SCIENTIFIC CONTEXT

This paper is contextualised in the area of long-term interaction studies and proposed to use automated analysis to obtain demographic and identity information from video data.

a) *Long-Term Interaction with Autonomous Robots “in the wild”:* Studies in long-term interaction with autonomous robots have a long-standing tradition in Human-Robot Interaction (HRI). Ranging from museum robots [2] studied in the last century, over robots deployed in schools [3] and offices [4], to robots in care homes [1], the long-term engagement patterns have been in the focus of interest.

¹https://github.com/strands-project/aaf_deployment/tree/indigo-devel/face_analysis

²in SPSS and CSV formats at <https://figshare.com/s/6b2f65c749bba8495b79>

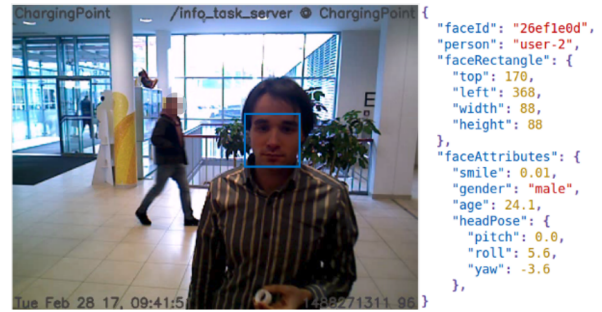


Fig. 2. Exemplary output of the MicroSoft Face API for a photo captured by the robot.

Recently, [5] placed a total of 70 robots in people’s homes to analyse user experience over a period of 6 months. Among the interesting observations often studied in HRI is the so-called novelty effect (e.g. also in [3]), which inherently required to (re-)identify users when they return to the robot for further interaction after an initial customisation phase. For “robots in the wild” this re-identification can pose quite a challenge when hundreds of users are interacting with a robot; a challenge addressed in this paper.

One of the most explored applications in human-populated environments is using mobile robots to provide information services. The reason this application is so popular is that mobile robots in contrast to fixed information terminals, can move around their environments to provide information and actively approach potential users [1], [6] where needed.

While most studies in literature either explicitly use questionnaires to evaluate user experience or acquire observational data, in this paper we take a more opportunistic approach: We *offered* the robot to users in a care home, but did not explicitly elicit or even encourage interactions, nor did we ask for any subjective assessments. Rather we present a novel approach to relate automatically obtained interaction patterns with automatically obtained biometric information during the long-term deployment of the autonomous robots, without the presence of any experimenters or technical support staff during the deployment.

b) *Biometrics and Demographics Analysis:* Age estimation from faces has seen most impressive progress in recent years, mostly due to the application of deep learning methodology. [7] provides a concise overview and [8] can be seen as one of the most recent state-of-the-art solutions. Overall, state-of-the-art technology appears to be well capable of estimating the age of people from their faces in the range a Mean Absolute Error (MAE) of less than 3 years on publicly available datasets. Also, see [9] for a more detailed survey comparing performances of machine perception versus human perception of age and gender. These results are encouraging in the sense that they indicate that robots could be able to on the fly estimate the age of an interacting person. Obviously, robots might not only be interested in a person’s age but – in many cases – even more important is the ability to re-identify a returning person. Again, recent advances in face recognition technology render

the automatic identification well suitable to be employed in HRI scenarios. A survey by [10] compared more than 13 different approaches for pose-invariant face recognition and reports the best to perform significantly above 95% recognition accuracy on most standard datasets.

In this work, we present results obtained from processing faces with the Microsoft Face API [11] (see Fig 2 for exemplary output). The service is provided through a Web API, and no documentation about this proprietary approach is publicly available, but it is deemed to employ state-of-the-art technology based on the work by [12], comparable to published performance results.

III. METHODS & MATERIALS

The data analysed for this work was collected during months of robot deployment in a care home and is focused on the user interactions with the robot whilst it was providing an info-terminal service. The robot was deployed in the care home for a total of 127 days, 95 of which the robot was actively providing its services.

The robot operated daily from 6 am until 9 pm, however because of safety concerns it was only allowed to work away from its charging station on weekdays from 9 am to 6 pm, out of these hours it provided the info-terminal service at the charging station.

A. System Description

The mobile platform employed is a 1.76m tall SCITOS G5 mobile robotic platform; it can reach 7 hours of autonomous, uninterrupted operation without a recharge. It has a touch-screen on the back which is the focal point for interaction. Fig. 1 shows the physical appearance of the robot. For perception, the robot is equipped with a Kinect camera on top of its head for people perception and a SICK S300 laser range finder for obstacle avoidance, localisation and further leg detection.

Its software system is based on the ROS middleware. For navigation, the ROS standard navigation “move-base” stack³ is enhanced with a topological layer that allows route planning using previous robot experience [13]. For people perception, the robot uses an upper body detection module based on 3D images acquired live from the head camera, and a leg detection component analysing 2d laser scans. A Bayesian tracker combines the modalities to produce robust people trajectories [14].

The robot offers different services in the care home, including an info-terminal service, a robotic bellboy and a walking group physical therapy behaviour [15], [16].

B. The Info-Terminal Service

This service is a mobile information point with a touch-screen based interface that allows users to request information or be entertained on the screen of the mobile platform. Hence, the system is similar to static information points often available in care homes but with the added ability

to provide the information at varying places in the overall environment [1].

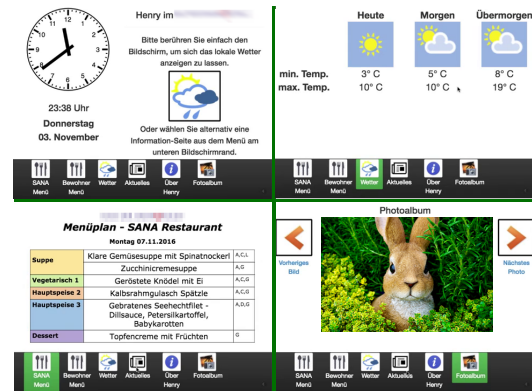


Fig. 3. Info-terminal Interface.

The info-terminal interface is designed considering older adults needs since the information service was aimed mainly (but not limited) at them. Also, the touch-screen of the robot is accessible to wheelchair users and walking people alike. The interface is implemented in HTML5 and JavaScript and served via a web browser operating in full-screen mode, Figure 3 shows some screenshots of the info-terminal. The information screens SANA accessible are:

- *Home screen*: The home screen is set automatically when the robot starts offering the service and it shows the current time and date as well as the local weather.
- *Weather*: Displays the local weather forecast for the current day and its two successive days.
- *News*: Shows national news and news from the care home, They are automatically updated from a news feed and the care home facebook page, respectively.
- *Restaurant Menus*: Displays the menus of the restaurant and the choices available for the residents.
- *Photo Album*: This features a selection of photographs from natural environments and the care home itself.
- *About Henry*: A page with an informative video about the robot and why it was there.

During its deployment, the robot would offer its services at selected locations around the care home; every ten minutes the robot would move to a new location chosen by the approach presented in [1].

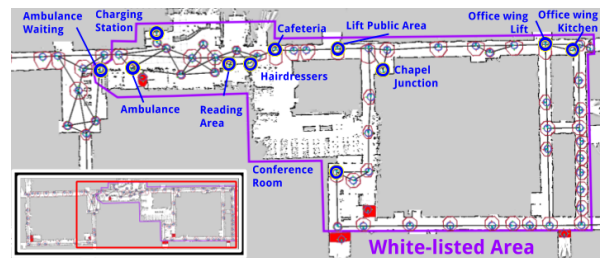


Fig. 4. Care home topological map. Labeled way-points are those where info-terminal was provided, purple box covers white-listed way-points.

³A detailed description of the ROS navigation stack can be found at: http://wiki.ros.org/move_base

Every time the robot reaches a new location it positions itself so its touchscreen is easily accessible by users, and displays the main menu of the info-terminal, ready for any users to interact with it. The robot will not move to a new location in the middle of an interaction even if the 10 minutes have ended, however, it will move 1 minute after the last interaction was detected.

C. Data Collection

The user data analysed for this paper was collected over a 68 day period, from the 1st of January to the 9th of March 2017. During this period the robot was fully operational for 58 days, while it was offline or constraint to its charging station during the remaining 10 days due to local constraints (public holidays, weekends and maintenance). We shall discuss the population and user base to be encountered by the robot in this environment in the context of the results in Sec. IV-E. The study has been conducted with the approval of the care home Ethics board and the research institution’s research ethics committee. The data collection was confined to the public (non-residential) section of the building, with recording specifically prohibited in care areas, and only permitted in “whitelisted” areas as detailed in Fig. 4.

The source of data for identifying users is timestamped videos recorded from the robots’ head camera. These videos captured the whole stream of robot operation at one frame per second, however, because of data protection reasons, especially in the hospital area, videos were only recorded in *whitelisted* areas of the environment as labelled in the topological map used for localisation. No video was recorded in other (blacklisted) way-points.

During the deployment several data streams from the robot were recorded, however, for this work, only the following data has been utilised:

- Touch events on the touchscreen,
- the currently displayed info screen,
- the video of the robot’s head camera at 1 Hz (only at whitelisted way-points), recording the interacting user,
- trajectories of people in the robot’s vicinity, acquired by employed tracking system [14], and the
- topological localisation of the robot.

IV. AUTOMATED ANALYSIS

To facilitate analysis of users of the system, an automated analysis pipeline is proposed, that brings together the information about detected people with the recorded videos. The automated analysis consists of a pipeline that in most parts can also be employed in an online manner to allow a robot to analyse interacting users in real time. However, in this paper, this pipeline is used in a post-mortem analysis based on the data collection outlined in Sec. III.

A. Interaction Types

Due to the nature of the info-terminal service provided by the robot, users can access it with active or passive intent. Hence, before outlining the technical details of the pipeline,

two different kinds of interaction that are to be considered in this paper shall be defined:

- *Active Interactions*: refer to interactions of users using the touchscreen of the robot, actively engaging with it and requesting specific information, e.g., weather.
- *Passive Interaction*: are those interactions where users approach the robot and clearly look at its screen, but without touching it.

An example of each type of interaction (and also a case of passing-by, ie. not interacting) is shown in Fig. 5(b) and Fig. 5(a), respectively. In this context, the reader is referred to table II which summarises the total number of identified users, classifying them into entirely passive, entirely active, or both. Note, that this table already requires users to be identified, as will be discussed in the following subsection.

B. Face Analysis

As discussed in Sec. II impressive progress has been made on applications extracting biometric, demographic or identity information from faces. It is these advances that nowadays make it possible for companies to offer products using this technology as online services to their customers. The majority of these services provide face detection and some demographic attributes, but few of them also provide face recognition and identification, which is essential for the analysis proposed in this paper. The proposed system employs the Microsoft Face API (originally based on work of [12]) through its web API providing face (re-)identification, and estimation of demographic attributes such as age, gender, hair colour, accessories, and glasses among others. As part of the work presented in this paper, the service has been made available as a ROS component for the automated demographics analysis of robot users.

C. Processing Pipeline

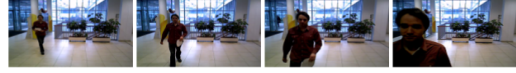
Figure 6 sketches the different processing steps to identify and classify user interactions. For the post-mortem analysis of the info-terminal two different pathways are proposed, attributed to the very different nature of the two types of interaction (*active* and *passive*). The entire pipeline is governed

TABLE I
DATASET OVERVIEW

Date range analysed	2017/1/1 - 2017/3/9
Total days in range	68
Total days with video	58
Total time video recorded (hours)	1014
Total people detected (incl blacklists)	24583
Total people detected at white-listed way-point	17516
Total people detected after filter (out of camera range, etc)	12364
Total passive interactions automatically	406
Total passive interactions manually	359
Total passive interactions with user associated	210
Total passive interactions with user duplicates removed	172
Total active interactions with video (incl black-list)	7352
Total active interactions at white-listed way-point	3813
Total active interactions with associated faces	1120
Total active interactions with user associated	616
Total active interactions with duplicates removed	428
Total interactions analysed	600
Total different users recognized automatically	311
Total different users recognized manually	344



(a) Passive Interaction



(b) Active interaction.



(c) No interaction.

Fig. 5. Interaction types

TABLE II

NUMBERS OF INDIVIDUAL USERS, PASSIVE AND ACTIVE, OR BOTH

Users only ever been passive	103	29.9%
Users only ever been active	179	52.0%
Users that were active and passive	52	15.1%
TOTAL	344	100%

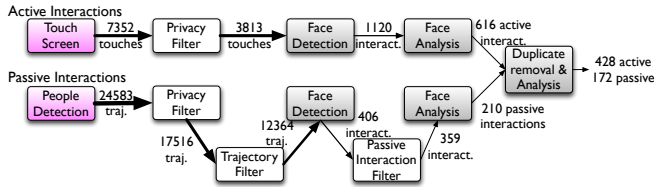


Fig. 6. Overview of the processing pipeline for active and passive interactions.

by an automated detection module (either “Touchscreen” or “People detection”, respectively), which is essential to provide an initial filter for the vast amount of data obtained in long-term deployments which are not economically accessible to extensive manual annotation. The pipelines then filter this initial set of detected potential interaction, with numbers of interaction cases that are processed through the pipeline and filtered out during the process being indicated in Fig. 6.

a) Active Interaction: Active interactions are first extracted from the data logged during the deployment. For each touchscreen event, a tuple containing a timestamp, screen-name, and the current way-point was logged. The location stored for each interaction is then used to exclude all the interactions that took place in all blacklisted areas. As a result of this filter, the number of touches to be analysed decreases from 7352 to 3813.

In next stage, the video frames corresponding to the specific timestamp of each remaining interactions are extracted, and a simple face detection method based on OpenCV libraries is used to detect faces in the image. Faces might well be absent in the image, mainly due to two causes: (i) users are too close to the robot when touching the screen and the faces are partially/entirely out of the image, or (ii) the low rate of 1 Hz sometimes might miss the best frame to see the user. Another factor is sometimes motion blur due to bad lighting conditions. A time window of $\pm 2sec$ is analysed around the interaction time to find the best face

detection and is associated with that interaction. For the current dataset, from the remaining 3813 touches, 1120 are successfully associated with a face.

The last step consists of sending each frame with detected faces to the Microsoft Face API to possibly re-identify previous users, or training new users automatically if they are not yet in the dataset. Also, if there is more than one face detected in the frame, these users are considered as a group and the biggest or most centred face is assigned to the user who touches to be considered the “primary-user”, or in other words the one that is actively interacting with the robot. As a result of this filter, 616 different interactions with identified users are obtained.

b) Passive interactions: Unlike the active interactions, there is no screen interaction information that can be used as a reference. Hence, the pipeline starts directly from the robots’ people perception system. It provides the trajectory of each detected person using the system by [14]. However, with the screen being placed on the back of the robot, opposite to the 2d laser, it cannot rely on the leg detector to detect people interacting with the touchscreen. Hence, on this side, the data from people tracker is only provided by an upper body detector based on 3D images from the directional head camera.

Due to this fact, after the first “Privacy Filter”, there are 17516 trajectories that might correspond to passive interactions. This makes it necessary to filter out detections that are not in the camera range. On this step we also filter out false positives (for example, a few objects detected as persons). This process is called “Trajectory Filter” and discards interactions when: (i) the robot is in motion, as the head camera is heading forwards for navigation and it is not possible to see who is looking at the screen, or (ii) all of the trajectory is out of the range of the camera (± 65 degrees from the back of the robot), or (iii) the detection time is shorter than 2 seconds, or (iv) if during the whole tracking, the person’s position changes by less than 0.1 meters (likely false positive).

After this process, the number of possible interactions decreases from 17516 to 12364. Following on, for each person trajectory remaining every corresponding video frame is analysed, searching for face detections related to the person position in that same timestamp. If there are face detections in at least 2 consecutive frames and the distance to the robot is lower than 3m (a generous range to look at information

TABLE III

COUNTS FOR ACTIVE AND PASSIVE INTERACTIONS IN RELATION TO SPECIFIC INFO SCREENS WHEN THE USERS' FACES WERE DETECTED.

Screen looked at	Type of Interaction		Total
	active	passive	
Init	40	115	155
Robot Info	50	2	52
Cafeteria Menu	41	8	49
Resident Menu	56	13	69
News	48	13	61
Photos	110	13	123
Weather	83	8	91
TOTAL	428	172	600

on the screen), it is considered that this user could have looked at the info-terminal and this is considered a passive interaction. The remaining 406 possible passive interactions have been hand annotated only for analysis in this paper and only 47 of them were non-interactions (remaining false positives). These mostly correspond to people, normally in a group, talking or standing in front of the robot but without showing interest in it.

The final stage is to run the detected faces through the Microsoft Face API in the same way as for active interactions to obtain robust face detection and recognize previous users. As result of all the process, there are 210 passive interactions with associated users. A summary of the counts in our dataset is presented in table I.

The data gathered from the deployment of the robot as described in Sec. III has been used to first evaluate the potential of the proposed pipeline shown in Fig. 6. As part of the analysis, the raw results had to be post-processed to remove duplicates of active interactions, which resulted from users pressing several times during the very same interaction window. Also, passive interactions that were co-occurring as active interactions during the same interaction period were removed as duplicates. This processing step yielded the 428 active interactions and 172 passive interactions further analysed in this section. The outcome of the automated analysis has then been studied to gain insights into the use of the info-terminal application.

D. Critical Appraisal of Automated Analysis

Table III provides an overview of the different types of interactions that occurred during the provision of the info terminal service. In total, there were 600 analysed interactions, with an *interaction* being defined as either *active interaction*, or *passive interaction*, as detailed above.

a) Robustness of Face Identification: In total, among the 600 interactions, 334 different users had been identified manually by two annotators and each interaction first individually coded by the two annotators with a very high agreement (Cohen's $\kappa = 0.987$). In a moderation process, both annotators agreed on a jointly consolidated annotation to be used as ground-truth for the assessment of the automated identification using face identification.

The number of manually identified persons is higher than

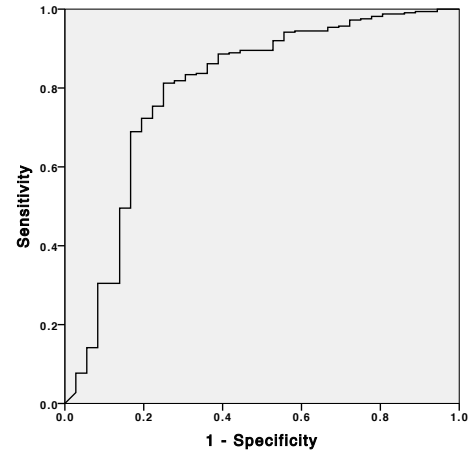


Fig. 7. ROC curve for face identification. The area under the curve is $AUC = 0.795$.

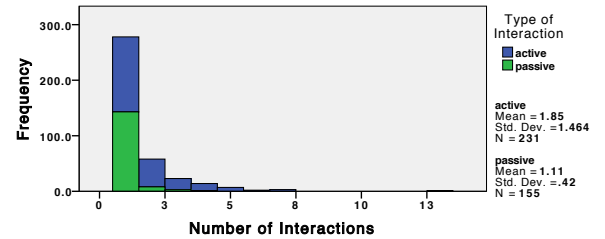


Fig. 8. Histogram of number of interactions for individual users.

the 311(93%) different users identified by the automatic analysis. Hence, it is evident that the system confused a number of users as already known, while in some cases it also separated the same person into two different classes. Consequently, as part of the manual annotation, 31 users were separated by the manual annotators, 12 users were merged, and 21 were modified. If considering the automatic and the manual annotations as two independent annotators, the inter-rater agreement between the manual and the automatic identification is Cohen's $\kappa = 0.927$ (95% Confidence Interval: ± 0.021), indicating still a very high agreement.

The overall accuracy of the automatic identification is $92.8\% \pm 1.1\%$ (Standard Error SE) in our dataset. However, the rate can be tweaked if one makes use of the confidence measure that is provided by the employed face identification approach. This confidence measure ranges from 0.0 to 1.0 ($\mu = 0.7013 \pm 007$) and is indicative of the certainty that two faces belong to the same person. Rejecting uncertain recognition results according to this measure can increase the specificity, making sure faces are correctly identified, at the cost of not identifying some at all. The best trade-off obviously is application-dependent. A best operational point can be determined using the ROC-curve shown in Fig. 7.

b) Robustness of Age and Gender Estimation: As interactions with the robot were spontaneous in our study without and ground truth demographic data available, an assessment of the age and gender estimation can only be done statistically. The absence of ground truth data, and the

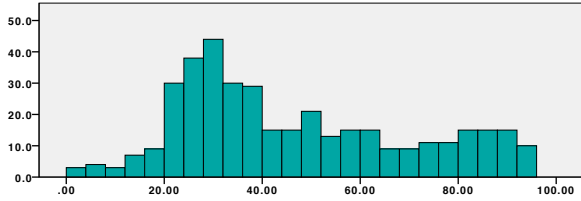


Fig. 9. Age distribution among all users.

fact that studies have shown (e.g. in [9], [10]) that automatic estimation of age and gender nowadays often outperforming humans on many dataset, we deem a further costly comparison of the estimation performance to human performance as not providing additional relevant information for our case. Starting from the manual annotation of people’s identity, the analysis focused on recurring people and analyses the dispersion of the age estimation for those returning users. Looking at Fig. 8, it can be seen that a total of 322(54%) of the 600 analysed interactions were interactions with returning users. In order to provide an assessment of the robustness of the age estimation, the average standard deviation of the age estimation of the very same user can be useful. In the presented dataset the mean of the standard deviation is $\bar{\sigma} = 5.113 \pm 0.46$, indicating that the noise in the estimation of age is quite a bit higher than the accuracy reported in state-of-the-art algorithms discussed in Sec. II, which indicates that age estimation can be as accurate with 3 years on average. However, those reported results are obtained in controlled datasets, sometimes with a number of samples of the same face, while in our study only one image is being analysed for each individual interaction to estimate the age.

A similar analysis of the gender classification shows that the gender classification is consistent (the same gender classification in all occurrences of the same user) in 301(95.7%) of the 322 interactions of returning users, again indicating a robustness in the same order of magnitude expectable from state-of-the-art technology.

E. Usage Patterns and Demographics Analysis

Having confirmed a reasonable robust estimation of demographic information from the automated analysis, some insights into the actual data acquired in the context of the long-term deployment can be gathered. This analysis is mostly descriptive and explorative by nature, as in this post-mortem analysis no dedicated conditions for hypothesis testing had been defined. This analysis should mostly be seen as a use case for the automated analysis, although some interesting patterns of use can be derived that may have wider insights into the deployment of robots in care homes.

While the proposed pipeline, in general, also allows estimating affective states of the user (e.g. the average happiness across all interactions was measured as 19.7% as opposed to 2.2% anger), in this paper the emphasis is on demographics. As another example of such an automated analysis, we consider the relationship between age groups and the usage pattern, ie., whether individual people are returning users or

TABLE IV
USAGE PATTERNS IN RELATION TO AGE GROUPS

Age Group		One time User	Returning User	Regular User	TOTAL
Child	Count	8	2	1	11
	Expected Count	6.4	3.3	1.2	11.0
Working Age	Count	100	57	15	172
	Expected Count	100.5	52.1	19.4	172.0
Pensioner	Count	27	11	10	48
	Expected Count	28.1	14.5	5.4	48.0
TOTAL	Count	135	70	26	231
	Expected Count	135.0	70.0	26.0	231.0

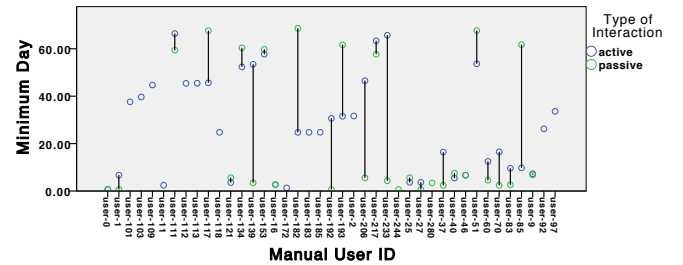


Fig. 10. Drop-line plot of first active and first passive interaction for all users that had both.

only interact with the robot once. Fig. 9 provides an overview of the estimated age distribution, which may be typical for an environment providing a home for approximately 350 older adults and is the workplace of about 465 staff. It shall be noted that the institution also includes a nursery for about 20-25 pre-school children. For further analysis, the age distribution, attributed to the fact that age estimation must be considered quite noisy (see beginning of this section), is divided into three bins: Children (younger than 18), working-age people (between 18 and 65) and Pensioners (older than 65, usual pension age in Austria). Table IV summarises the observations regarding the demographics, and classifies *active* users into three categories, namely, “One time User” (single interaction), “Returning User” (2-3 interactions), and “Regular User” (more than 3 interactions). Pearson’s χ^2 measure does not indicate a significant correlation ($p = 0.125$) between these factors among the 231 recognised active users of the system.

F. Returning Users

A most interesting analysis possible by automatically re-identifying users is to investigate returning users. As Fig.8 already indicated, the majority of users engaged with the robot only ones, and as Tab. IV highlighted, 135 (58.4%) of the active users only ever engaged with the robot once, while some even stay entirely passive (see Tab. II).

In order to gain a deeper understanding into the typical long-term usage patterns, 42 users who have had at least categorised as “Returning Users” and whose interactions spread over more than a day where analysed in Fig. 10. The plot shows the first passive and the first active interactions over the course of the data collection, respectively, and highlights the period of non-engagement in between. For most (23) users, there was little time difference, and they

were passively and actively active on the same day. However, 12 users first engaged only passively with the system (first took a look), and only after a significant amount of time (more than a day) returned to become active. On the opposite, 7 users returned after an active engagement to then engage only passively. The plot does not show any interactions that happened in between, though, and focuses only on the earliest interactions of each type. It can be utilised to investigate designed active engagement behaviour in future versions of the robot.

G. Discussion

The results regarding the performance of the automated analyses are encouraging. Even though only a sample of the overall interactions (600) could be automatically analysed, the performance in identification is very good comparing it to human classification with a very high inter-rater agreement. Hence, we draw the conclusion that our system can indeed be utilised to form statistical statements of the identity of users in scenarios as the one presented. The age and gender estimation is deemed sufficient to categorise users into dedicated groups and perform analysis with the demographics obtained in this way.

Due to space constraint, we only presented a small subset of mostly descriptive statistics and analyses of the obtained dataset. We refer the reader to the dataset to investigate further factors or specific hypotheses.

V. CONCLUSION

a) Summary of Contribution: This paper proposed a processing pipeline, both suitable for post-mortem analysis as well as *in-situ* classification of interactions and obtains the demographics and identity of returning users with high accuracy. This allows the processing of large amounts of recorded data and preparing it for analyses with ease. We have shown that the performance of the estimation is sufficient to employ it studies with robots in the wild. An extensive dataset acquired from the long-term deployment of an autonomous mobile info-terminal robot has been presented, and – using the proposed approach – analysed to gain insights into the demographics and usage patterns of the info-terminal robot. All technical contributions, software and dataset, are freely available to the community with our system being able to also run online on a robot, it offers the opportunity to make use of such acquired knowledge *in-situ*, e.g. to improve adaptive behaviour through learning of individualised usage patterns, to then anticipate user’s intentions and act accordingly, e.g., as presented in [1].

b) Limitations and Future Work: One of the most obvious limitations of the approach is the limited view of the interacting user’s face, required in order to be able to process it. Indeed, it was this limitation that reduced the number of interactions that could be analysed down to 428 active interactions (out of the 3813 theoretically possible one). Very often, the employed camera would only see half the face or none at all during the interaction. Nevertheless, we consider the results very encouraging to equip our robots

with the ability to automatically analyse and, in consequence, also learn about its users, given better image acquisition equipment is deployed in our future deployment, e.g. in a local museum a project that is ongoing.

REFERENCES

- [1] M. Hanheide, D. Hebesberger, T. Krajník, and others, “The When, Where, and How: An Adaptive Robotic Info-Terminal for Care Home Residents – A long-term Study,” in *Proc ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, ACM/IEEE. Vienna: ACM/IEEE, 2017.
- [2] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, “MINERVA: a second-generation museum tour-guide robot,” in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, vol. 3. IEEE, 1999, pp. 1999–2005. [Online]. Available: <http://ieeexplore.ieee.org/document/770401/>
- [3] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, “A Two-Month Field Trial in an Elementary School for Long-Term Human-Robot Interaction,” *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 962–971, 10 2007. [Online]. Available: <http://ieeexplore.ieee.org/document/4339534/>
- [4] J. Biswas and M. Veloso, “The 1,000-km Challenge: Insights and Quantitative and Qualitative Results,” *IEEE Intelligent Systems*, vol. 31, no. 3, pp. 86–96, 5 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7478588/>
- [5] M. M. de Graaf, S. Ben Allouch, and J. A. van Dijk, “Long-term evaluation of a social robot in real homes,” *Interaction Studies*, vol. 17, no. 3, pp. 461–490, 3 2016. [Online]. Available: <http://www.jbe-platform.com/content/journals/10.1075/is.17.3.08deg>
- [6] R. Pinillos, S. Marcos, R. Feliz, E. Zalama, and J. Gómez-García-Bermejo, “Long-term assessment of a service robot in a hotel environment,” *Robotics and Autonomous Systems*, vol. 79, 2016.
- [7] S. Soman and A. Austine, “A Survey on Age Estimation Techniques,” *International Journal of Computer Applications*, vol. 161, no. 4, pp. 26–28, 3 2017. [Online]. Available: <http://www.ijcaonline.org/archives/volume161/number4/soman-2017-ijca-913176.pdf>
- [8] C.-C. Ng, Y.-T. Cheng, G.-S. Hsu, and M. H. Yap, “Multi-layer age regression for face age estimation,” in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 5 2017, pp. 294–297. [Online]. Available: <http://ieeexplore.ieee.org/document/7986859/>
- [9] H. Han, C. Otto, X. Liu, and A. K. Jain, “Demographic estimation from face images: Human vs. machine performance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148–1161, 2015.
- [10] C. Ding and D. Tao, “A Comprehensive Survey on Pose-Invariant Face Recognition,” *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 3, pp. 1–42, 2 2016. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2885506.2845089>
- [11] Microsoft, “Microsoft Face API,” 2017. [Online]. Available: <https://docs.microsoft.com/en-gb/azure/cognitive-services/face/overview>
- [12] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” in *Proceedings of the British Machine Vision Conference 2015*, 2015, pp. 1–41. [Online]. Available: <http://www.bmva.org/bmvc/2015/papers/paper041/index.html>
- [13] J. Pulido Fentanes, B. Lacerda, T. Krajník, N. Hawes, and M. Hanheide, “Now or later? predicting and maximising success of navigation actions from long-term experience,” in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015.
- [14] C. Dondrup, N. Bellotto, F. Jovan, M. Hanheide *et al.*, “Real-time multisensor people tracking for human-robot spatial interaction,” 2015.
- [15] D. Hebesberger, C. Dondrup, T. Koertner, C. Gisinger, and J. Pripfl, “Lessons learned from the deployment of a long-term autonomous robot as companion in physical therapy for older adults with dementia: A mixed methods study,” in *Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI ’16. IEEE, 2016.
- [16] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Murova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner *et al.*, “The strands project: Long-term autonomy in everyday environments,” *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017.