

SmartLobby: A 24/7 Human-Machine-Interaction Space within an Office Environment

Stefan Fuchs

Honda Research Institute Europe
Offenbach, Germany
stefan.fuchs@honda-ri.de

Manuel Mühlig

Honda Research Institute Europe
Offenbach, Germany

Nils Einecke

Honda Research Institute Europe
Offenbach, Germany
nils.einecke@honda-ri.de

Bram Bolder

Honda Research Institute Europe
Offenbach, Germany

Fabian Eisele

Technical University of Darmstadt
Darmstadt, Germany

Abstract

In this work, we present the SmartLobby, an intelligent environment system integrated into the lobby of a research institute. The SmartLobby is running 24/7, i.e. it can be used any time by anyone without any preparations. The goal of the system is to conduct research in the domain of human machine cooperation. One important first step towards this goal is a detailed human state modeling and estimation. As the system is built into the lobby area, also people having no scientific background are interacting with the system which strongly reduces the bias in the collected data. On software side the SmartLobby mainly integrates state-of-the-art algorithms that enable a thorough analysis of human behavior and state. These algorithms constitute the fundamental basis for the development of higher level system components. Here, we present our system with its various hardware and software components. Thereby, we focus on the realization of a non-intrusive head-eye-tracking as a key component to continuously observe persons using the system. A first multi-week lasting experiment utilizes this capability to investigate a method for content personalization, which demonstrates the effectiveness of the whole system.

1 INTRODUCTION

This paper presents the SmartLobby, an intelligent environment system, built into the entrance area of a research institute. It allows to present the institute's ongoing research and to perform research at the same time. The SmartLobby is designed for the institute's major research topic: cooperative intelligence, as the next level of artificial intelligence (AI). That is, AI systems cooperating with humans in an adaptive fashion and having a sense for responsibilities in the cooperation. To achieve this goal, it is necessary to research on anthropomorphic

Copyright © by the paper's authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In: E. Strinati, D. Charitos, I. Chatzigiannakis, P. Ciampolini, F. Cuomo, P. Di Lorenzo, D. Gavalas, S. Hanke, A. Komninos, G. Mylonas (eds.): Proceedings of the Poster and Workshop Sessions of AmI-2019, the 2019 European Conference on Ambient Intelligence, Rome, Italy, November 13-15, 2019, published at <http://ceur-ws.org>



Figure 1: View into the SmartLobby area.

interaction mechanisms as well as on innovative interactions specialized to specific tasks. It is also important to equip intelligent systems with competences that adapt the interaction with the human depending on the situation. An essential prerequisite is the ability to estimate the internal and observable human states. Although this topic list is not comprehensive it highlights that this kind of research cannot be conducted isolated inside a lab. Especially, online learning, adaptation and personalization rely on a continuous feed with data from human interaction.

The SmartLobby presented in this paper comes with an unobtrusive appearance, which makes the ongoing research a tangible and natural experience for visitors and employees. At the same time, the SmartLobby features properties that support versatility and flexibility as these are requirements that come along with this very agile research domain.

2 RELATED WORK

The SmartLobby is following the concept of living lab that was coined at the MIT by Mitchell, Larson and Pentland in 2003 [12]. The first living labs have been realized in smart environments to observe humans in their usage of emerging technologies (microscopic level). Especially with the transformation from ubiquitous computing [37] towards intelligent environments [5, 17, 34] the importance of living labs has grown. Intelligent environments are considered to support and collaborate with humans in a non-intrusive intelligent way and to satisfy their experience in the environment. Hence, the development of such systems requires a user-centered research methodology in order to continuously take into account the needs and wishes of potential users as well as the evolving real life context. Meanwhile, the concept of living labs has been tailored more generally (macroscopic level) to enhance innovation and usability of information technology driven applications in the society. This trend intensified even more with the recent progress in the field of artificial intelligence where seeing the potential users as competent partners in transparent development processes may reduce qualms regarding this technology.

We see intelligent-cyber-physical-systems (ICPS) [25] at the intersection of cooperative intelligence, smart environments and robotics. Augusto et al. [5] describe intelligent environments as a synthesis of ubiquitous computing, smart environments and artificial intelligence. Saffiotti et al. [34] went a step further and introduced physically-embedded-intelligent-systems (PEIS-ecology) as the intersection of ubiquitous computing, AI and robotics. In this context, artificial intelligence is an essential prerequisite to gain autonomous self-directed agents that can follow own goals and draw related decisions. We, however, aim at enriching the system with a sense for the human state as well as a sense for responsibilities in the cooperation with the human. Thus, we

refer to ICPS as an intelligent environment that adheres to the principles of cooperation (see [22]).

ICPS systems target at applications in various domains from health, hospitals over manufacturing and emergency services to education. A very common application are smart homes equipped with health features also called: ambient assisted living (AAL) [36, 32]. AAL is mainly focusing on helping elderly people to keep their self-determination. The basic idea is to augment the home environment of elderly people with intelligent devices and offering remote care services that assist the elderly people in their daily live. This field has grown substantially due to advancements in smart sensors and devices. A survey [7] of AAL papers between the years 2004 and 2013 found that most work is concentrating on activity recognition [10, 23, 27] and monitoring of the vital status [4, 3, 33]. Further reviews [30, 9] show that the most potential sensors in current AAL projects are cameras, in particular RGB-D cameras. Cameras allow for analysis that is difficult to achieve with other sensors like human behavior analysis, fall detection or emotion recognition. On the downside, the information richness of cameras also constitute a critical privacy issue. In particular, cameras installed in sensitive areas like bathrooms are very critical and result in low acceptance. Hence, privacy preservation is also an important focus of research. For example [8] tackle the privacy issue by replacing the real camera images with a pre-stored background image plus stick-figures of the current people visible. Furthermore, all necessary data for computing behavior and situation analysis is stored locally.

Another close area to our work is research on improving quality of life in working environments. However, the setup is often restricted to sensors directly installed at each desktop [20, 24, 3] and not in an ambient fashion as in AAL. Going more in direction of ambient assisted working is the idea of a virtual secretary presented in [11]. Several offices were equipped with cameras and microphones in order to recognize people and their activities. In case a meeting situation is detected, visitors trying to enter the door or callers trying to phone someone in the room are automatically informed. Unfortunately, false detection rates are high because of the simple heuristics used. Furthermore, the system was mainly tailored around detecting presence and availability of persons in the offices.

Finally, [31] analyzed two prominent AAL middleware platforms and derived a set of recommendations for future AAL systems from this. Two major insights are that AAL systems should favor GUIs over text-based interfaces and that the system should have a good balance between standardization and customizability. Our system [14] adheres to these suggestions by having a graphical touch interface and by using ROS [29] as underlying middle-ware platform.

3 SMARTLOBBY SYSTEM

The following section describes the requirements and the realization of the the overall system with its structural concept, selected sensors, effectors and software.

3.1 General Requirements

Our goal was to design a room that has both a representative character and an intelligent behavior according to the concept of ICPS. The intelligence of the SmartLobby is not meant to be a static demonstration of research results but rather to be continuously developed further by interaction. That is, there are challenges arising from the users with their heterogeneous background and privacy considerations as well as regarding robustness, operability and versatility. The daily usage by expert and non-expert (e.g. administration or guests) users in a standard office environment provides the necessary insights to actually generate intelligent systems with a natural and intuitive human-machine-interaction.

The overall impression of the room should avoid usual lab impressions such as cables, extensive sensor rigs or prototypical electronics. A presentation software has to present the institute with its structure, employees, connections to other institutes and research projects. Additionally, the presentation software should welcome visitors or announce talks embedded into a weekly press review. The presentation software should be intuitive and self-explaining in order to encourage a usage by guests and novices.

One major target is the 24/7 operation of the system's basic functionalities, i.e. the presentation software, people tracking, live feedback, and head-eye-tracking have to run continuously. Furthermore, a fundamental requirement is versatility and extensibility of the hardware installation to accommodate to new research approaches.

3.2 Structural Concept

The location we selected for the SmartLobby is an essential junction of the institute’s traveling paths, because it connects two building parts with the employee offices and laboratories. With administration and management offices, a coffee kitchen and two meeting rooms directly connected, this area is frequently visited by employees and visitors alike. Fig. 1 gives an image of the office lobby.

The room has a size of 6 m \times 5.5 m and features five doors and two open gateways. A large bar table extends from the coffee kitchen’s service hatch into the room. The bar table invites to meet, interact with the system and operate the touch screens that are embedded into it.

The wooden material of the walls generates a comfortable atmosphere while the lamella design maintains a more technical touch. In the upper part the lamella distance is increased to make them more transparent for light from the adjacent rooms.

Sensors and actors are either directly built into the wooden wall or attached to it. Power and network cables are hidden inside the wall to support a living room atmosphere. In order to keep the room flexible, the walls feature various hidden extension slots that can house additional sensors, effectors or computation hardware. However, major computing hardware is installed remotely in a server room.

3.3 Sensors and Effectors

Fig. 3a shows a top view of the SmartLobby and its sensors. The room is equipped with seven imaging devices. There are three Kinects (XBox One) for providing depth measurements, three ceiling-mounted RGB cameras (IDS UI-5250CP-C-HQ, lens 6 mm) in the corners for a complete room overview, and a pan-tilt-zoom-(PTZ)-camera (Axis Q6128, lens 3.9 mm to 46.8 mm) for high fidelity scans. At a distance of 6 m the PTZ-camera still captures images with a spatial resolution of 5 pix/mm. The room is also equipped with two microphone arrays (depicted by small green circles in Fig. 3a) each consisting of eight pressure zone microphones (AKG C 562 CM, Beyerdynamic Classis BM32W). Their arrangement enables sound localization and highly sensitive speech recognition.

The most straightforward interaction with the system is the usage of the displays. At the heart of the interaction is a multi-touch-table (MTT) that comprises two touch displays (scape tangible 55) each having a UHD resolution (3840 pix \times 2160 pix) at 55 inch screen diagonal. The left touch display runs a proprietary software for interactively presenting the research institute. It can be operated not only by finger touches but also by physical passive and active tokens that trigger certain information. There is one wall mounted display close to the MTT with a 84 inch screen diagonal. The wall along the main passage features two ultra-wide displays each with a size of 2.16 m \times 0.35 m at a resolution of 3840 pix \times 600 pix. This installation enables the interactive control of the displays as a feedback for persons walking by. Two speakers beside the large display allow for playing back audio and video data as well as performing speech-based interaction.

A strong focus was also put on the light installation. The lamellas in the walls are partially replaced by multi-color LED stripes. These LED stripes are grouped into eight segments (two for each wall), which can be controlled independently. The ceiling lights provide white light with controllable intensity. The light is driven by DMX protocol and accessible via Ethernet by a LANBox-CLX device.

3.4 Mobile Robot

The SmartLobby covers only a small area of the institute. As a straightforward extension further research projects involve mobile robots that are permanently active in the whole office area, e.g. for guiding people. Together with these robots the SmartLobby connects to a multi-entity system that shares information among the different entities. For example the robots act as mobile extended sensors for the SmartLobby. In return the SmartLobby feeds user information gathered in user interactions back to the robots and potentially other entities. A more detailed description of the mobile robots and their usage is described in [16, 13].

3.5 Software

The large variety of sensors and the generated data cannot be processed on a single computer. Thus, there is the need for a versatile middleware that allows for distributed processing as well as fast integration of new sensors and algorithms. We decided to use ROS [29], which comes with a lot of available sensor packages (e.g. `iai_kinect2`, `uEye_cam`) and deployable applications (e.g. `OpenPTrack`). Also ROS provides means for web-based usage that

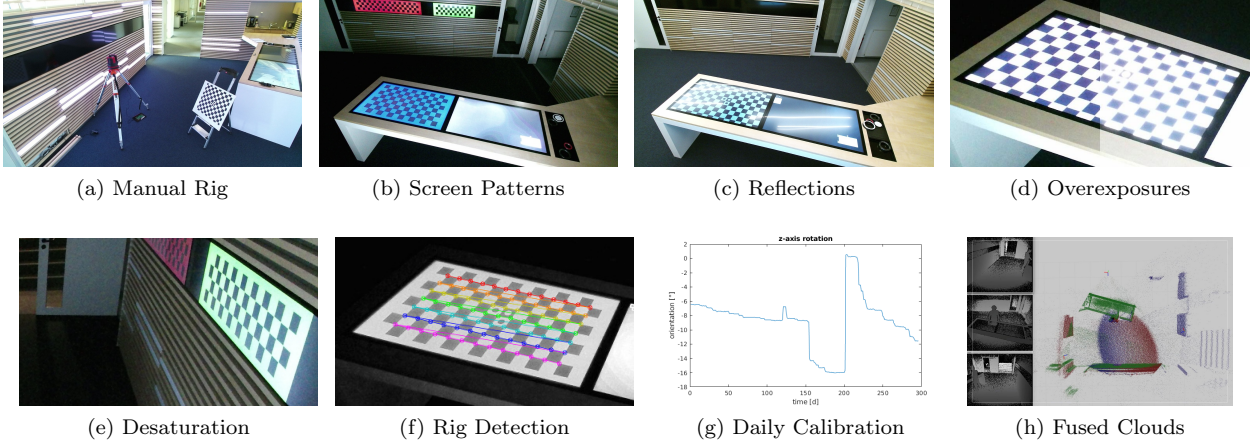


Figure 2: (a) Camera calibration using a rig. (b) Camera calibration using differently colored patterns on different screens. (c) Reflection of the ceiling light overcasts calibration pattern. (d) At night, the screen patterns might overexpose if display brightness is set to high. (e) Steep viewing angles can lead to color desaturation. (f) OpenCV 2D chessboard point detections. (g) Daily calibration results over the course of 296 days. (h) Combined point cloud of the three Kinects after automatic extrinsic calibration.

is independent of the operating systems involved (`ros_bridge`). Hence, ROS can be easily integrated with any kind of hardware as long as an HTML5-renderer is available.

In order to assure robustness of our multi-entity system, the mobile robots and the smart lobby are organized in separate ROS systems with separate ROS masters. This ensures that the robots can disconnect and reconnect at any time without the need of restarting the overall system. To share specific topics and services across the multiple ROS systems, the `multimaster_fkie` package [1] is used.

4 SMARTLOBBY MODULES

The variety of ROS software modules deployed in the SmartLobby can be categorized into: infrastructure modules to control the sensors and effectors, maintenance modules to guarantee the 24/7-operation, and application modules. The following sections describe selected software modules from the latter two categories. Thereby, we focus on modules that enable and foster accurate human monitoring to realize a personalization application.

4.1 Automatic Camera Calibration

After the installation of the cameras an extrinsic calibration is necessary to unify the coordinate systems to merge (see Fig. 2h) or to transfer information amongst the sensors. In a first attempt, we calibrated the cameras with a standard checkerboard calibration rig (see Fig. 2a). Unfortunately, the cameras are subject to slight movements, due to ball joints in the camera mount and lightweight dry-walls which vibrate slightly when a door is closed. A regular recalibration would be necessary but is very time consuming when using a manually placed rig, especially because multiple rigs are necessary in order to cover all cameras. Hence, we devised an automatic calibration, which makes use of the installed monitors.

The basic idea is displaying the calibration patterns on the screens. Since the screen positions are very stable their exact position needs to be measured only once. In order to easily distinguish the screens we use different pattern colors as shown in Fig. 2b. There are, however, some details that need to be considered. Firstly, the calibration patterns shouldn't be occluded. For this reason, we decided to trigger the calibration at 1 am, when people are rarely present. Secondly, there is a problem with the monitors' reflective properties. The cameras might see a reflection from the lighting which overcasts the calibration pattern (see Fig. 2c) and thus prevents a clean calibration.

Automatically switching off the light before calibration prevents reflections but introduces another problem. As the room is dark and the screens are bright, the automatic camera exposure will be high and lead to overexposed pixels (see Fig. 2d). To account for that, we reduce the screen brightness to 15 % during calibration. Finally, we observed some desaturation effects when having a camera at a steep angle relative the the calibration screens. As Fig. 2e shows, this desaturation is not homogeneous. The calibration pattern is green on the left side (far

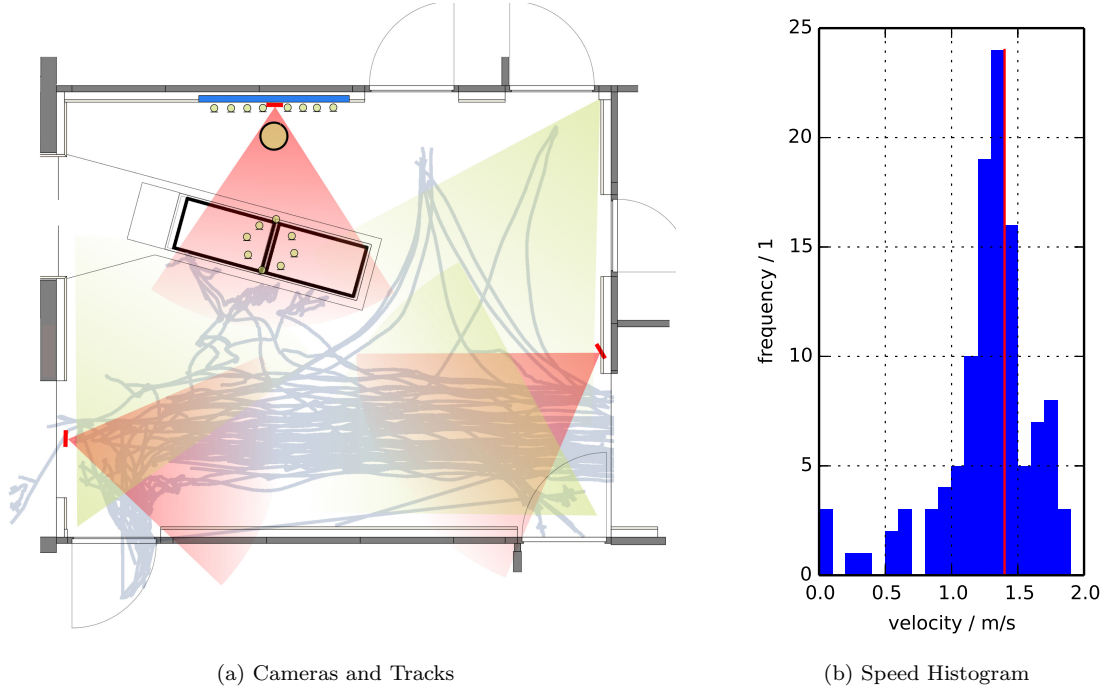


Figure 3: (a) Layout of the lobby area with Kinects (reddish cones) and RGB-cameras (greenish cones). The large orange dot refers to the PTZ-camera. The small green dots refer to the microphones. (b) Walking speed histogram

side) of the display but is almost white on the right (close side). We counteract this mainly by reducing the pattern extent on the ultra-wide screens.

The actual calibration procedure is done with OpenCV [6] (see Fig. 2f). Fig. 2g shows the extrinsic parameter measurement of the center Kinect over the course of 296 days. In this plot a notable drift in orientation is observable. Furthermore, the plot shows some events of strong perturbations. These were caused by maintenance work on the pan-tilt-zoom unit which is located on the ceiling above the center Kinect.

In order to measure the accuracy of the automatic calibration we generated a manual Ground Truth calibration by manually annotating the corners of the screen calibration patterns in the camera images. We found the automatic calibration error to be between 0.36° and 0.64° for camera orientations and between 0.1 m and 0.4 m for camera positions. The eligibility of the manual Ground Truth was judged with a simple experiment: We assumed that humans are able to localize corner points with an accuracy of at least 1 pix. By jittering the human screen corner point annotations by ± 1 pix we found that the average accuracy of the manual calibration is about 0.01 m in position and 0.13° in orientation. Since the average error caused by 1 pix jitter is much smaller than the difference between automatic and manual calibration, we can regard the manual calibration as a reasonable ground truth.

4.2 People Tracking

For tracking people in the SmartLobby, we employ OpenPTrack [26]. One important aspect of OpenPTrack is its support for distributed computing. In our setup, we use one ZBox (Zotac Magnus-1070k, i5-7500T CPU @ 2.70GHz Quadcore) for each Kinect. OpenPTrack allows to run point cloud generation and detection on the ZBox while fusion and tracking is done on one of the central SmartLobby computers. Hence, the computational load is highly distributed and data bandwidth is reduced as the data-heavy point clouds are not sent over the network.

Fig. 3a shows that the arrangement of Kinects allows to track humans in the lobby on all reasonable paths. During one morning about 120 persons pass the SmartLobby. The histogram of average walking speeds corresponds to the typical speed of 1.4 m/s (see red line in Fig. 3b).

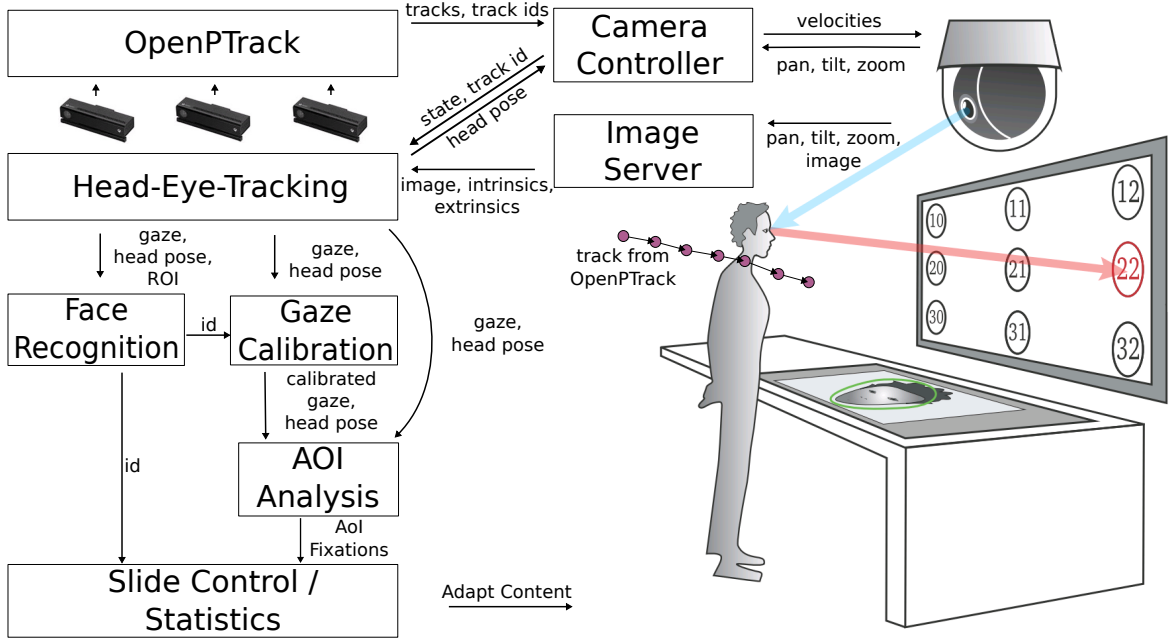


Figure 4: The Head-Eye-Tracking is driven by the PTZ-camera. The outcomes of the people tracking (OpenPTrack) are used to control the PTZ-camera’s viewing direction and zoom. Head orientation and gaze are fed into face recognition and AoI analysis, to find out who has seen which content.

4.3 Head-Eye-Tracking

Head pose and eye gaze are regarded as a major cue to infer the internal human state [2]. In the context of smart environments usually vision based remote eye tracking systems are applied, because of their non-intrusiveness and their possibility to promptly interact with the system. Gaze estimation relies on qualitative images, which results in a confined tracking box. The enhancement of the tracking box is an integral challenge of such systems. There are systems with movable cameras (rays, drone), systems that try to cope with less resolution and pan-tilt-zoom cameras.

We decided to integrate an Axis PTZ-camera, which is placed above the large display as it is the most prominent information hub in the room (see Fig. 4). Thus, it is possible to track all humans that enter the area and capture their glances while watching the screen. Two threaded processes retrieve the images and forward them as ROS messages at 40 Hz and a latency of 75 ms. We developed a simple kinematic model of the PTZ-camera given the assumption that the image plane is moving on a sphere located in camera center. Here, we used a common checkerboard pattern to retrieve the anchor point of the camera, the radius of the sphere, and the pan offset. Given the possibility to change the camera’s focal length from 3.9 mm to 46.8 mm with a resolution of 10000 steps there could be the same amount of intrinsic calibrations. The high quality optics allows us to neglect the distortion and we set the principal point to the center of the image. Furthermore, we estimated the actual focal length (camera scale) for five operation points and generated a look-up-table.

The PTZ-camera provides an extensive REST-API (VAPIX) to monitor, parametrize and control the device. We implemented a closed-loop PI-controller, which controls the camera orientation and its zoom by adjusting the velocities of pan, tilt, and zoom. The controller runs at 20 Hz and its major parameters (K_p and T_n) have been optimized to capture an entering human within maximal 1.5 s. The maximum speed of the PTZ-camera is set to $90^\circ/s$.

The head-eye-tracking-(HET) uses tracking messages from OpenPTrack as an initial guess for a head position in order to follow the people’s faces with the PTZ camera. Head pose and eye gaze are estimated by means of the `dlib`-library [18, 19] and GazeML [28] with an average accuracy of 10° . This accuracy is sufficient to estimate grid-wise areas-of-interest (AoIs), which are used by other applications, such as the personalization (see Fig. 7c).

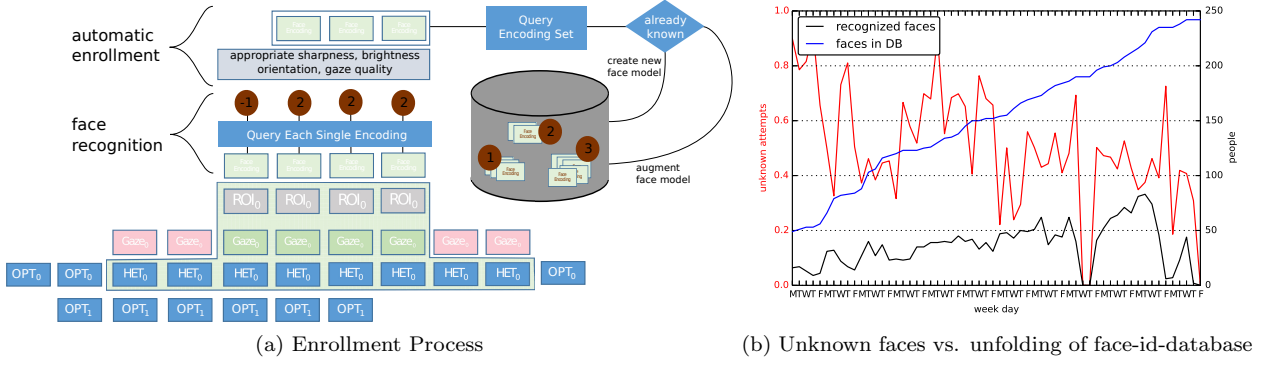


Figure 5: (a) Face recognition mechanism with automatic enrollment. (b) During a period of 12 weeks approx. 250 face models have been enrolled automatically (blue line). At the same time, the rate of recognition queries classified as “unknown” continuously decreases (red line).

4.4 Face Recognition

Recognizing individuals entering the SmartLobby is an inevitable capability to implement personalized behavior. The most common and non-intrusive approach is face recognition. Usually, face recognition is encompassed by an enrollment process wherein a user deliberately provides one or more images of his face together with his name. In spirit of tighter data protection regulations and its uncertain reading, we decided to implement an anonymized face recognition. The system learns new faces in an unsupervised fashion without associating the actual names of the users and still providing unique id numbers.

Our face recognition system is based on the `dlib`-library [19] and the HET-module outcomes. Two processes are running in parallel (see Fig. 5a): First, every time the HET-module gives a head pose, the system tries to recognize a face. To this end it passes an image cutout of the head pose to the face recognition of the `dlib`-library, which transforms these into a 128-dimensional face-feature-vector. This measured feature vector is used to search for the most similar face model in the database. The system computes the L^2 -norm of the measured face feature and the stored ones. A match is found for the pair with the minimal norm that is above a certain threshold ($1 - L^2$ -norm). Second, all face RoIs of a coherent head tracking are filtered according to particular requirements (brightness, orientation, sharpness). The remaining RoIs are assumed to relate to a single individual and the facial-feature-vector is computed for each RoI. By using the aforementioned search method a matching face feature is added to the database (augmentation) or, if not available, a new face model is created (enrollment).

These two processes rely on different confidence thresholds that have been decided on the basis of face recognition performance measures. According to [15] and [35] we identified the *false non-match rates* (FNMR), *false match rates* (FMR), *false acceptance rates* (FAR), and *false rejection rates* (FRR). Fig. 6a shows the comparison of the `dlib`-face recognition performance with the best approach from a recent NIST evaluation. At a FMR of 0.1 % genuine facial features are classified as unknown for 27 % of the queries (FNMR). This is a magnitude higher than state-of-the-art (2.5 % FNMR). It is representative only for our setup and potentially caused by imperfect lighting conditions and unfavorable viewing angles. Fig. 6b is used to decide on the confidence thresholds for the recognition and unsupervised augmentation processes. Usually, a threshold is selected that gives equal FARs and FRRs. The developers of the `dlib`-face recognition software recommend a threshold of 0.4 (light green dots). In our setup, we used for the regular face recognition and the initial enrollment a threshold of 0.475, which yields a FRR and FAR of 9 %. For the unsupervised augmentation, we used a more strict threshold of 0.6. Thus, we can prevent the face model database from adding incorrect facial features and mixing up individuals.

Fig. 5b shows the face database unfolding over a period of 12 weeks. Currently, approximately 100 researchers are with the institute and should regularly pass-by the SmartLobby. The automatic enrollment prefers to enrich the database with persons that are captured by the HET-system, i.e. persons that either are present in the lobby or interact with the screens. Over time, the number of enrolled faces continuously increases, whereas the percentage of unknown faces reduces continuously. It can be observed that the system generates more IDs than the expected 100 researchers. There are two main reasons: First, the system also captures guests and maintenance staff. Second, the system sometimes creates more than one instance for a person, which needs to be improved in the future.

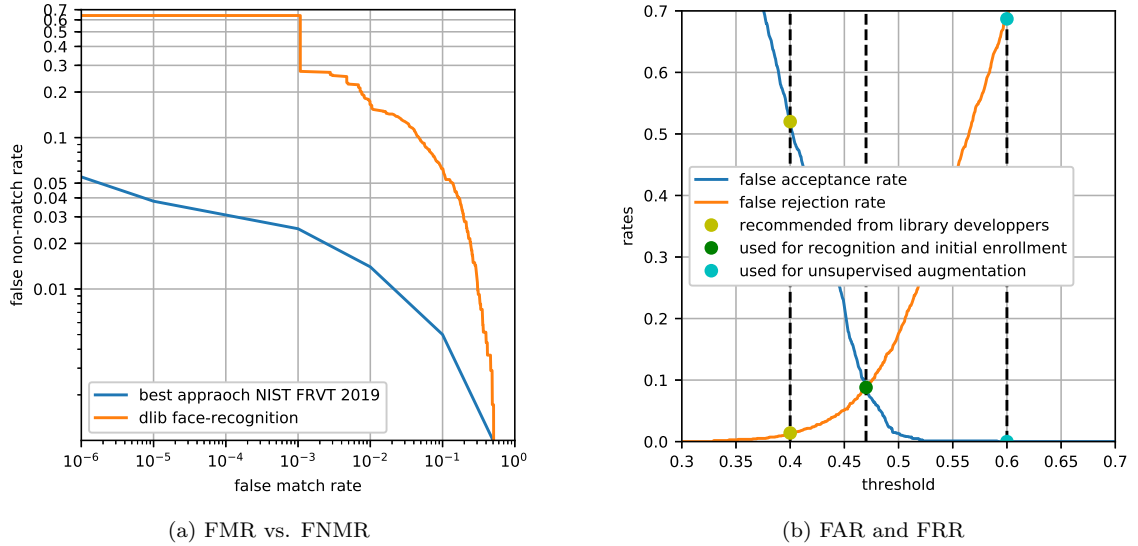


Figure 6: (6a) false non-match rate (FNMR) is plotted over false match rate (FMR) to show detection error tradeoff characteristics of different face recognition approaches. The scales are logarithmic. In our setup `dlib`-face recognition is inferior to the best state-of-the-art approach (*innovatrics* [15]), but it is still sufficient for our application: (6a) false acceptance rate (FAR) and false rejection rate (FRR) are plotted to find the confidence threshold at the equilibrium. With a minimal confidence of 0.475 we can have a FAR and FRR of about 9%.

4.5 Presentation Software

A major task of the SmartLobby is to represent the institute and to visualize ongoing research activities with their achievements. We targeted for a system that can express the creativity and innovative strength of the institute by using state-of-the-art devices and recent technology. As a result, we decided to use multi-touch displays (MTD) as interaction hubs, which not only react on touch gestures but also recognize physical tokens placed on the displays.

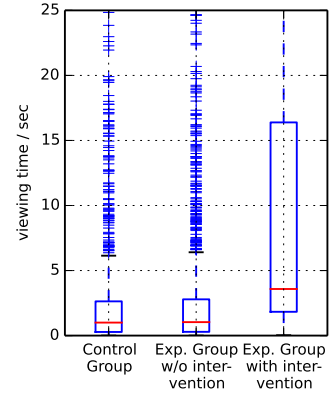
The latter capability is a basic feature of the presentation software *easire* (developed by *interactive scape*), which govern the left MTD and the large display. For example, *easire* can present the structure of the institute, CVs of the people working at the institute or give information about running and finished projects. Fig. 7a depicts a view when standing in front of the left MTD. The large, upright screen can either show a duplicate of the content of the left MTD or additionally display videos, documents and charts. This feature is especially useful when presenting information to a larger audience, because the horizontal touch screen is hardly visibly from farther away due to an unfavorable view perspective. We found that people enjoyed this kind of presentation much more than dusty old slides. Furthermore, the token concept allows them to easily explore the things they are interested in. The right MTD is set aside for RViz, which visualizes ROS topics, such as point clouds (see Fig. 2h), robot trajectories or results of our algorithms.

4.6 Personalization Experiment

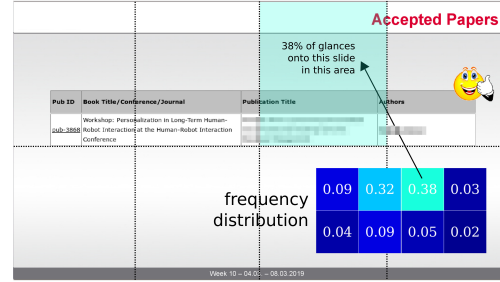
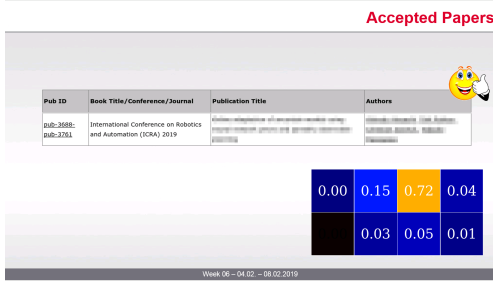
During idle times a slide show welcomes visitors, announces talks and comprises a weekly updated press review. Typically a dozen pages are repeatedly shown, each for 2 min. Given the face recognition and the AoI analysis, the system can memorize the slides a person has already seen. A first multi-week lasting experiment investigates a basic content personalization method targeted at all kinds of visitors to the SmartLobby. We divided the people into two groups, one where the system intervenes in the repetitive slide process and one where it does not. The groups are split randomly by intervening only for people with an even ID. For the intervention group, the system skips the slide show to pages that have not been regarded by the person yet. This means, if the current slide, when entering the room, has not been seen by the person yet, the system will also not switch to another slide for people in the intervention group. Fig. 7b shows the impact of the slide-skipping intervention. The analysis of the intervention group is split into events where no intervention happened (slide was not yet known anyway) and actual skipping events. As one can expect, there is no significant difference between the control group (no intervention) and the events of the intervention group where no intervention happened (slide



(a) Tokens trigger the presentation of various contents



(b) Personalization



(c) Glance statistics of two particular slides

Figure 7: (a) View when standing in front of the multi-touch-table. The tokens give access to information about the institute, its people and projects. When there is no mutual interaction, a weekly update slide show is presented continuously. (c) We are collecting statistics on the interest and acceptance of the individual slides. The glance frequencies are rather similar for slides with the same structure and content. (b) The experimental group is shown slides it has not seen yet.

unknown yet). The median viewing time is similar for both groups with about 1s. This seems rather short, but it includes also people passing by without reading the screen. In contrast, skipping the pages attracts the attention of the visitors and increases the mean viewing time tremendously to 4s.

Furthermore, we investigated the reading behavior for different content categories: tech news and publication announcements. Fig. 7c shows the glance frequency distribution for the publication slides. The gaze transition entropy [21] is a promising criteria to measure interest in content: Though the median entropy is similar for the two categories with 2.8, the entropy variance for the news slides is more than twice as big. This demonstrates, that different news topics cause a larger spread in the gaze transition entropy values of the readers.

5 CONCLUSION AND OUTLOOK

We have presented an intelligent environment integrated into the lobby of a research institute. We described how to equip a specialized public space with ubiquitous technologies, so as to showcase research results as well as to serve as a workbench for collecting data for research in the domain of human machine cooperation. This so-called SmartLobby is available 24/7. The interior design, the hardware and software infrastructure allow for versatility and extensibility. Thus, we can quickly set up experiments and conduct unobtrusive user studies in an every-day environment.

For a start, we developed a head-eye-tracking system, that covers the whole lobby, and an anonymous face recognition to enable long-term user monitoring in the SmartLobby. On this basis, we equipped the lobby with a simple intelligent behavior and personalized the news feed shown in the lobby. Our results show the positive personalization impact by increasing the median viewing time. Furthermore, we investigated gaze transition entropy as a measure for interest in the news feed slides. As a next step, we will enhance the personalization by also considering the personal interest.

ACKNOWLEDGMENT

We would like to thank *studio klv GmbH* for designing and managing the reconstruction of the lobby and *Interactive Scape GmbH* for implementing the presentation software as well as the token and touch user-interface.

References

- [1] http://wiki.ros.org/multimaster_fk1e.
- [2] A. Al-Rahayfeh and M. Faezipour. Eye tracking and head movement detection: A state-of-art survey. *IEEE Journal of Translational Engineering in Health and Medicine*, 1, 2013.
- [3] A. Alberdi, A. Aztiria, A. Basarab, and D. J. Cook. Using smart offices to predict occupational stress. In *International Journal of Industrial Ergonomics*, volume 67, pages 60–66, 2018.
- [4] Y. Andreu, F. Chiarugi, S. Colantonio, G. Giannakaki, D. Giorgi, P. Henriquez, E. Kazantzaki, D. Manousos, K. Maria, B. J. Matuszewska, M. A. Pascali, M. Pediaditis, G. Raccichini, and M. Tsiknakis. Wize Mirror - a smart, multisensory cardio-metabolic risk monitoring system. In *CVIU*, pages 3–22, 2016.
- [5] J. C. Augusto, V. Callaghan, D. Cook, A. Kameas, and I. Satoh. Intelligent Environments: a manifesto. *Human-centric Computing and Information Sciences*, 3:12, 2013.
- [6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [7] D. Calvaresi, D. Cesarini, P. Sernani, M. Marinoni, A. F. Dragoni, and A. Sturm. Exploring the ambient assisted living domain: a systematic review. *Journal of Ambient Intelligence and Humanized Computing*, 8(2):239–257, 2017.
- [8] CherryHome. Cherryhome.ai website, 2019.
- [9] S. Colantonio, G. Coppini, D. Giorgi, M.-A. Morales, and M. A. Pascali. Computer vision for ambient assisted living: Monitoring systems for personalized healthcare and wellness that are robust in the real world and accepted by users, carers, and society. In *Computer Vision for Assistive Healthcare*, pages 319–336, 2018.
- [10] R. Cucchiara, A. Prati, and R. Vezzani. A multi-camera vision system for fall detection and alarm generation. *Expert Systems*, 24(5):334–345, 2007.
- [11] M. Danninger and R. Stiefelhagen. A context-aware virtual secretary in a smart office environment. In *Proceedings of the 16th ACM International Conference on Multimedia*, pages 529–538, 2008.
- [12] M. Eriksson, V. Niitamo, S. Kulkki, and K. Hribernik. Living labs as a multi-contextual R&D methodology. *The 12th ITM Conference*, January 2006.
- [13] L. Fischer, S. Hasler, J. Deigmöller, T. Schnürer, M. Redert, U. Pluntke, K. Nagel, C. Senzel, A. Richter, and J. Eggert. Where is the tool? - grounded reasoning in everyday environment with a robot. In *International Cognitive Robotics Workshop (CogRob)*, 2018.
- [14] S. Fuchs, N. Einecke, and F. Eisele. Smart-lobby: Using a 24/7 remote head-eye-tracking for content personalization. In *Proceedings of UbiComp/ISWC*, 2019.
- [15] P. Grother, M. Ngan, and K. Hanaoka. Ongoing face recognition vendor test (frvt). Technical report, NIST National Institute of Standards and Technology, 2019.
- [16] S. Hasler, J. Kregar, and U. Bauer-Wersing. Interactive incremental online learning of objects onboard of a cooperative autonomous mobile robot. In *Neural Information Processing*, pages 279–290, 2018.
- [17] M. B. James Dooley and M. R. Al-Mulla. Beyond four walls: Towards large-scale intelligent environments. pages 287–294, 2012.
- [18] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. June 2014.

- [19] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.
- [20] K. Kiyokawa, M. Hatanaka, K. Hosoda, M. Okada, H. Shigeta, Y. Ishihara, F. Ooshita, H. Kakugawa, S. Kurihara, and K. Moriyam. Owens luis a context-aware multi-modal smart office chair in an ambient environmen. In *IEEE VR Workshops (VRW)*, pages 1–4, 2012.
- [21] K. Krejtz, A. Duchowski, and N. Villalobos. Gaze transition entropy. *ACM Trans. Appl. Percept.*, 13(1), December 2015.
- [22] M. Krüger, C. B. Wiebel, and H. Wersing. From tools towards cooperative assistants. In *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 287–294, 2017.
- [23] D. Lymberopoulos, A. Bamis, and A. Savvides. Extracting spatiotemporal human activity patterns in assisted living using a home sensor network. *Universal Access in the Information Society*, 10(2):125–13, 2011.
- [24] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. Affectaura: An intelligent system for emotional memor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 849–858, 2012.
- [25] H. A. Müller. The rise of intelligent cyber-physical systems. *Computer*, 50(12):7–9, 2017.
- [26] M. Munaro and E. Menegatti. Fast RGB-D people tracking for service robots. *Journal on Autonomous Robots*, 37(3):227–242, 2014.
- [27] S. Pandya, H. Ghayva, K. Kotecha, M. Awais, S. Akbarzadeh, P. Gope, S. C. Mukhopadhyay, and W. Chen. Smart Home Anti-Theft System: A Novel Approach for Near Real-Time Monitoring and Smart Home Security for Wellness Protocol. In *Applied System Innovations*, volume 1, 2018.
- [28] S. Park, A. Spurr, and O. Hilliges. Deep pictorial gaze estimation. In *European Conference on Computer Vision (ECCV)*, ECCV ’18, 2018.
- [29] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. ROS: An open-source Robot Operating System. *ICRA Workshop on Open Source SW*, 3:1–6, 2009.
- [30] M. K. Rainer Planinc, Alexandros Andre Chaaraoui and F. Florez-Revuelta. Computer vision for active and assisted living. In *Active and Assisted Living: Technologies and Applications*, pages 57–79, 2016.
- [31] A. M. Rajjeet Phull, RamiroLiscano. Comparative Analysis of Prominent Middleware Platforms in the Domain of Ambient Assisted Living (AAL) for an Older Adults with Dementia (OAwd) Scenario. In *Procedia Computer Science*, volume 83, pages 26–33, 2016.
- [32] P. Rashidi and A. Mihailidis. A survey on ambient-assisted living tools for older adults. *IEEE Journal of Biomedical and Health Informatics*, 17(3):579–590, 2013.
- [33] M. Sacristan, A. Brunete, M. Hernando, and E. Gambao. Robohealth: Smart room interface for hospital patients. In *Robots for Assisted Living - IROS Workshop*, 2018.
- [34] A. Saffiotti and M. Broxvall. Peis ecologies: Ambient intelligence meets autonomous robotics. In *Smart Objects and Ambient Intelligence Conference*, 2005.
- [35] S. K. Sahoo, T. Choubisa, and S. R. M. Prasanna. Multimodal biometric person authentication : A review. *IETE Technical Review*, 29(1):54–75, 2012.
- [36] H. Sun, V. D. Florio, N. Gui, and C. Blondia. Promises and challenges of ambient assisted living systems. In *Sixth International Conference on IT: New Generations*, pages 1201–1207, 2009.
- [37] M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):66–75, January 1991.