Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

SmartLobby: Using a 24/7 Remote Head-Eye-Tracking for Content Personalization

Nils Einecke, Stefan Fuchs, Fabian Eisele

2019

Preprint:

Copyright ACM 2019. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in UbiComp/ISWC '19 Adjunct, https://doi.org/10.1145/3341162.3343837.

Poster: SmartLobby: Using a 24/7 Remote Head-Eye-Tracking for Content Personalization

Stefan Fuchs
stefan.fuchs@honda-ri.de
Honda Research Institute Europe

Nils Einecke
nils.einecke@honda-ri.de
Honda Research Institute Europe

Fabian Eisele
Technical University of Darmstadt

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 with head-eye-tracking as key component. 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 head-eye-tracking as a key component to continuously observe persons using the system and customize content shown to them. The results of a multi-week lasting experiment demonstrate the effectiveness of the system.

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing systems and tools.*

KEYWORDS

smart environment, personalization, HCI, head-eye-tracking

ACM Reference Format:
Stefan Fuchs, Nils Einecke, and Fabian Eisele. 2019. Poster: SmartLobby: Using a 24/7 Remote Head-Eye-Tracking for Content Personalization. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341162.3343837

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6869-8/19/09.
https://doi.org/10.1145/3341162.3343837



Figure 1: View into the SmartLobby area.

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 [4], 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 interaction mechanisms as well as on innovative interactions. 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. Head pose and gaze are regarded as one major cue to infer the internal human state. Thus, we developed an unobtrusive remote head-eye-tracking system that covers the whole lobby area. As a first step, we employed this system to learn the faces of passers-by and continuously track their perception of information shown on a display in the lobby. The information gathered is used to personalize the content and investigate criteria to measure interest in content.

2 SMARTLOBBY SYSTEM

General Requirements

Our goal was to design a room that has both a representative character and an intelligent behavior according to the

concept of cooperative intelligence [4]. 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 and should welcome visitors or announce talks embedded into a weekly press review. One major target is the 24/7 operation of the system's basic functionalities, i.e. the presentation software, 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.

Structural Concept

The SmartLobby is located at an essential junction of the institute's traveling paths and connects two building parts with the employee offices and laboratories. With adjacent administration offices, a coffee kitchen and two meeting rooms, the SmartLobby is frequently visited by employees and visitors alike. Fig. 1 gives an image of the lobby. The room has a size of 6 m × 5.5 m. A large bar table extends from the coffee kitchen's service hatch into the room and features embedded touch screens. 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 hidden extension slots that can house additional sensors, effectors or computation hardware.

Sensors and Effectors

The SmartLobby 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. The room is also equipped with two microphone arrays 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×2160 pix) at 55 inch screen diagonal. The left touch display 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×0.35 m at a resolution of 3840 pix×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 as well as performing speech-based interaction.

The lamellas in the walls are partially replaced by multicolor LED stripes. These LED stripes are grouped into eight segments (two for each wall), which can be controlled independently. The light (including the ceiling lights) is driven by DMX protocol and a LANBox-CLX device.

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 middle-ware that allows for distributed processing as well as fast integration of new sensors and algorithms. We decided to use ROS [6], which comes with a lot of available sensor packages (e.g. `iai_kinect2`, `ueyecamera`) and deployable applications (e.g. `OpenPTrack`). Also ROS provides means for web-based usage that 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.

3 MODULES FOR CONTENT PERSONALIZATION

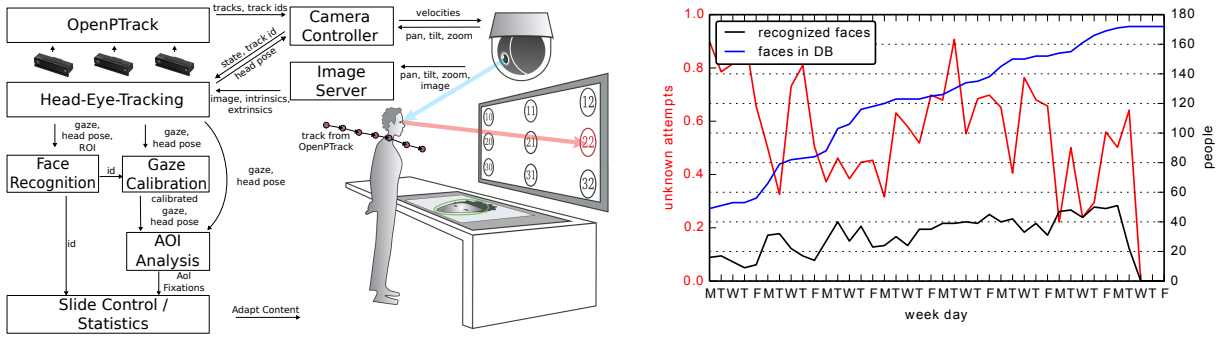
The SmartLobby features a large variety of modules. The following sections describe those modules that are important for content personalization.

People Tracking

For tracking people in the SmartLobby, we employ OpenPTrack [5]. One important aspect of OpenPTrack is its support for distributed computing. In our setup, we use one ZBox (Zotac Magnus-1079k, 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. Three Kinects are arranged on opposite sites of the room in order to track humans in the lobby on all reasonable paths.

Head-Eye-Tracking

Head pose and eye gaze are regarded as a major cue to infer the internal human state [1]. 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 high-quality images, which results in a confined



(a) Head-Eye-Tracking System **(b) Unknown faces vs. unfolding of face-id-database**

Figure 2: Head-Eye-Tracking system and unfolding of face database.

tracking box. The enhancement of the tracking box is an integral challenge of such systems.

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. 2a), in particular weekly news slides are presented here. With the PTZ it is possible to track all humans that enter the area and capture their glances while watching the screen. 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 model parameters. 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. Hence, 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, parameterize 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 at a maximum PTZ-camera speed of 90 °/s.

The head-eye-tracking-(HET)-system 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 gaze estimation consist of facial landmark detection and eye shape registration with an average gaze accuracy of 10 ° (see Fig. 1). Given this accuracy, we estimate grid-wise areas-of-interest (AoIs), which are suitable for other applications. Two examples can be seen in Fig 3b.

Face Recognition

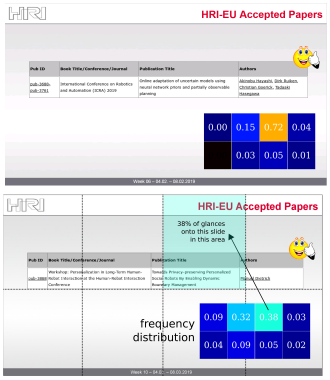
Recognizing individuals entering the SmartLobby is an inevitable capability to implement personalized behavior. In spirit of tighter data protection regulations and its uncertain reading, we decided to implement an anonymized face recognition. The system learns new faces in an unsupervised fashion without associating the actual names of the users and still providing unique ID numbers.

Our face recognition system is based on the `dlib` library [2] and the HET-module outcomes. Two processes are running in parallel: 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. Then the measured feature vector is used to search for the most similar face model in the database and its unique ID. 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 an available face model either is augmented or a new face model is created.

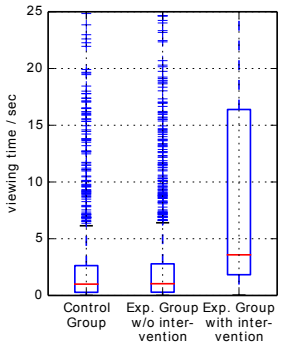
Fig. 2b shows the face database unfolding over a period of several weeks. Currently, approximately 100 researchers are with the institute and should regularly pass-by the SmartLobby. The automatic enrollment prefers to enrich the database with persons that are captured by the HET-system, i.e. persons 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



(a) Tokens trigger various contents



(b) Glance statistics of 2 slides



(c) Personalization

Figure 3: (a) View when standing in front of the multi-touch-table. When there is no mutual interaction with the tokens, a weekly update slide show is presented continuously. (b) 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. (c) The experimental group is shown slides it has not seen yet.

creates more than one instance for a person, which needs to be improved in the future.

Content Personalization

A major task of the SmartLobby is to represent the institute and to visualize ongoing research activities with their achievements. For this purpose, we decided to use multi-touch displays as interaction hubs, which not only react on touch gestures but also recognize physical tokens placed on the displays (see Fig. 3a).

During idle times a news 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 area of interest analysis, the system can memorize the slides a passerby has already seen. We conducted a simple experiment and divided the passersby into two groups. As soon as an experimental group participant enters the lobby, the slide show skips to pages that have not been seen yet. Fig. 3c shows the impact of this intervention. The median viewing time is similar for both groups with about 1 s. This seems rather short, but it includes also people passing by without reading the screen. Apparently, skipping the pages attracts the attention of the passersby and increases the median viewing time to 4 s.

We investigated the reading behavior for different content categories: tech news and publication announcements. Fig. 3b shows the glance frequency distribution for the publication slides. The gaze transition entropy [3] 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 entropies of the readers.

4 CONCLUSION AND OUTLOOK

We have presented an intelligent environment integrated into the lobby of a research institute. The so-called SmartLobby is running 24/7 and can be used for presenting our institute as well as conducting research in the domain of human machine cooperation. With a head-eye-tracking system, that covers the whole lobby, and an anonymous face recognition we personalize a news feed shown in the lobby. Our results show the positive personalization impact by increasing the median viewing time. 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.

REFERENCES

[1] A. Al-Rahayfeh and M. Faezipour. 2013. Eye Tracking and Head Movement Detection: A State-of-Art Survey. *IEEE Journal of Translational Engineering in Health and Medicine* 1 (2013).

[2] Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.

[3] Krzysztof Krejtz, Andrew Duchowski, and Natalia Villalobos. 2015. Gaze Transition Entropy. *ACM Trans. Appl. Percept.* 13, 1 (Dec. 2015).

[4] Matti Krüger, Christiane B. Wiebel, and Heiko Wersing. 2017. From Tools Towards Cooperative Assistants. In *Proceedings of the 5th International Conference on Human Agent Interaction*. 287–294.

[5] M. Munaro and E. Menegatti. 2014. Fast RGB-D people tracking for service robots. *Journal on Autonomous Robots* 37, 3 (2014), 227–242.

[6] Morgan Quigley, Brian Gerkey, Ken Conley, Josh Faust, Tully Foote, Jeremy Leibs, Eric Berger, Rob Wheeler, and Andrew Ng. 2009. ROS: An open-source Robot Operating System. *ICRA Workshop on Open Source SW* 3 (2009), 1–6.