GEAR: Gaze-enabled augmented reality for human activity recognition

Kenan Bektaş
kenan.bektas@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Jannis Strecker
jannisrene.strecker@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Simon Mayer
simon.mayer@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Kimberly Garcia
kimberly.garcia@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Jonas Hermann
jonas.hermann@student.unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Kay Erik Jenss
kayerik.jenss@student.unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Yasmine Sheila Antille
yasminesheila.antille@student.unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Marc Elias Solèr
marcelias.soler@student.unisg.ch
University of St. Gallen
St.Gallen, Switzerland

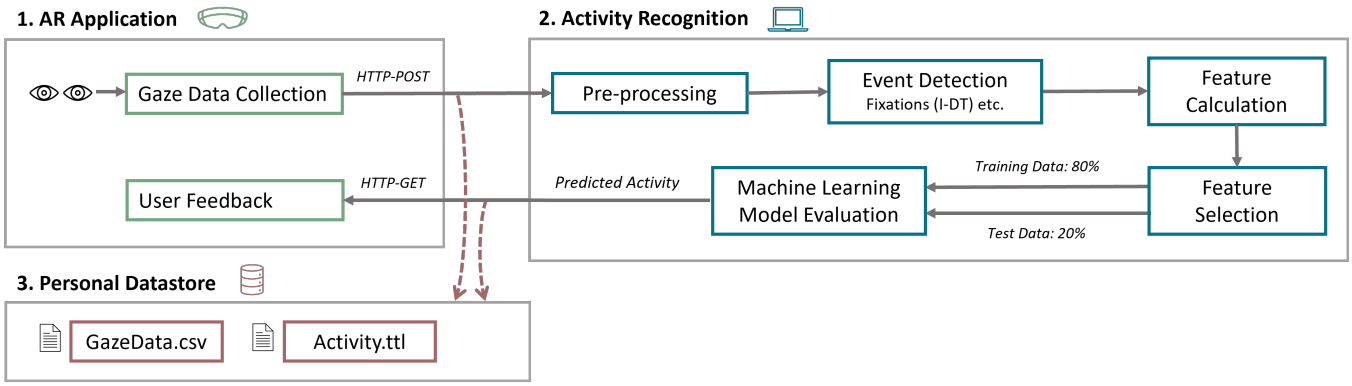

Figure 1: The components of the GEAR. The collected gaze data is sent from the AR Application (1) to the Activity Recognition component (2). The recognized activity is returned to the AR application which displays appropriate feedback. Both, the collected gaze data and the recognized activity, can be stored in a privacy-friendly Personal Datastore (3).

ABSTRACT

Head-mounted Augmented Reality (AR) displays overlay digital information on physical objects. Through eye tracking, they allow novel interaction methods and provide insights into user attention, intentions, and activities. However, only few studies have used gaze-enabled AR displays for human activity recognition (HAR). In an experimental study, we collected gaze data from 10 users on a HoloLens 2 (HL2) while they performed three activities (i.e., read, inspect, search). We trained machine learning models (SVM, Random Forest, Extremely Randomized Trees) with extracted features and achieved an up to 98.7% activity-recognition accuracy. On the HL2, we provided users with an AR feedback that is relevant to their current activity. We present the components of our system (GEAR) including a novel solution to enable the controlled sharing of collected data. We provide the scripts and anonymized datasets which can be used as teaching material in graduate courses or for reproducing our findings.

ETRA '23, May 30–June 02, 2023, Tubingen, Germany

ACM ISBN 979-8-4007-0150-4/23/05…$15.00
https://doi.org/10.1145/3588015.3588402

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Mixed / augmented reality**; • **Computing methodologies** → **Perception**; **Supervised learning**.

KEYWORDS

pervasive eye tracking, augmented reality, human activity recognition, attention, context-awareness

ACM Reference Format:
Kenan Bektaş, Jannis Strecker, Simon Mayer, Kimberly Garcia, Jonas Hermann, Kay Erik Jenss, Yasmine Sheila Antille, and Marc Elias Solèr. 2023. GEAR: Gaze-enabled augmented reality for human activity recognition. In *2023 Symposium on Eye Tracking Research and Applications (ETRA '23), May 30–June 02, 2023, Tubingen, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3588015.3588402

1 INTRODUCTION

Optical see-through head-mounted displays (HMDs) can augment a user's visual field by overlaying virtual information on their physical environment. Early versions of HMDs were complex, bulky, and expensive [Azuma et al. 2001; Billinghurst et al. 2015] (e.g., Sutherland's well-known *the Sword of Damocles* [Sutherland 1968]). Today, the form factor and usability of Augmented Reality (AR) HMDs (e.g., Microsoft Hololens 2, Varjo XR-3, Magic Leap 1 and 2, HTC Vive ProEye) have significantly improved, and they hold the potential to become more than a visual interface. AR HMDs can track the instant 3D position and movements of the user (including their head, hands, and eyes), can detect objects in the environment that appear in their camera feed through computer vision methods, and allow novel ways of interaction with connected devices that are close-by or remote [Strecker et al. 2022], bringing us closer to Weiser's vision [Weiser 1999] of pervasive computing. Across various indoor and outdoor activities, these may provide users with access to relevant information and services [Billinghurst et al. 2015; Grubert et al. 2017; Orlosky et al. 2021; Plopski et al. 2022].

The eyes provide essential visual input to the brain, thus studying eye movements provides insights into various cognitive processes and activities of humans (and animals). Researchers are interested in exploiting the potential of eye tracking for the creation of novel opportunities in human-computer interaction [Jacob and Stellmach 2016] and attention-aware computing [Vertegaal 2003], for selection [Zhai et al. 1999], foveated rendering [Bektas et al. 2015], activity recognition [Bulling et al. 2011], or in retrospective analysis [Salvucci and Goldberg 2000].

Eye tracking sensors can be readily integrated in HMDs for maintaining explicit, implicit, and collaborative interactions in Mixed Reality (MR) applications [Plopski et al. 2022] that can continuously sense and adapt to the requirements and constraints of users' context and activities [Bektaş 2020; Grubert et al. 2017; Orlosky et al. 2021]. Milgram and Kishino presented a well accepted continuum of MR [Milgram and Kishino 1994], focusing mainly on visual experiences in real, augmented, and virtual environments. At the virtual reality (VR) end of this continuum, users are exposed to computer generated visual stimuli (among other stimuli). In VR environments, eye tracking can provide valuable insights about users action planning and execution strategies (e.g., in [Keshava et al. 2021]). However, in VR experiences, users' perception of the virtual content is not necessarily strongly tied with the real environment that they inhabit [Bektaş et al. 2021]. On the other hand, AR HMDs provide users with a hybrid experience that is a synthesis of some virtual content (i.e., often and desirably related) with the natural or real scene. In natural settings (e.g., in daily personal or professional activities), gaze enabled AR HMDs would allow both a better understanding of the cognitive processes of humans and provide them with some assistance (e.g., visual, vocal, or haptic feedback) [Bektaş 2020; Bektaş et al. 2022; Grubert et al. 2017; Orlosky et al. 2021].

There is a growing interest in studying human activity recognition (HAR), where researchers make use of various (wearable) sensors that generate streams of data to train and test machine learning models [Bulling et al. 2014]. In recent years, mobile video-based eye trackers are also being used in HAR-research [Alinaghi et al. 2021; Braunagel et al. 2015; Kiefer et al. 2013]. However, in physical environments, the implications of using gaze-enabled AR HMDs for HAR have not been explored in detail.

In Section 2, we present a systematic review of HAR from gaze with mobile eye trackers and gaze-enabled HMDs. In Section 3, we introduce our main contributions: We present the components of our gaze-enabled AR (GEAR) system and detail an HAR experiment that we conducted with GEAR. We present a benchmark of models that we trained with the collected gaze data and introduce a solution that allows users to retain fine-grained control over sharing of their data and predicted activities. Section 4 re-casts the presented research as an input to graduate teaching on gaze-enabled AR and concludes the paper with a summary of findings and limitations, and an outlook for future directions.

2 HUMAN ACTIVITY RECOGNITION (HAR) FROM GAZE

Research on HAR is relevant for many applications in human-computer interaction and ubiquitous computing [Bulling et al. 2014] that focus on a seamless interaction between human users and interconnected systems. In context-aware computing, the behavior of a system can be adapted to environmental factors (e.g., location) and other factors such as users' expectations, psychophysiology, and activities [Bulling and Zander 2014; Vertegaal 2003]. These factors can be measured with various sensors (see [Bulling et al. 2014; Cornacchia et al. 2017] for reviews) which can be integrated into the environment and objects or can be worn by users. For example, head-mounted (or mobile) eye trackers pave the way towards a pervasive assessment of users' attention, intention, and activities [Bulling and Gellersen 2010]. Since the 1960s (e.g., the seminal work of Yarbus [Yarbus 1967]), eye trackers are used in studying task-dependent cognitive processes, and many studies have shown that it is possible to decode human activities from their eye movements [Borji and Itti 2014].

2.1 Activity Recognition with Mobile Eye Tracking

In mobile eye tracking, one of the most influential works on HAR was presented by Bulling and colleagues [Bulling et al. 2011] who followed a five-step procedure, which was also used by others. In an office environment, they collected raw data (Step 1) with a 128 Hz electrooculography (EOG) system from N=8 participants for recognition of six activities (copying, reading, writing, watching a video, browsing, and resting). After the drift and noise removal (Step 2), they computed a list of eye movement events (Step 3) such as fixations, saccades, and blinks. In feature extraction (Step 4), they calculated 62 features comprising descriptive statistics (e.g., mean, variance, and maximum) of these eye movement events. Lastly, in model training (Step 5), their Support Vector Machine (SVM)

model was able to classify six activities with an average precision of 76.1% and recall of 70.5%. Kiefer and colleagues used in their setup (N=17), a 30 Hz mobile and video-based eye tracker (SMI Eye Tracking Glasses) to recognize six activities on cartographic maps (free exploration, global search, route planning, focused search, line following, and polygon comparison) [Kiefer et al. 2013]. The authors reported a 78% accuracy with an SVM model that was trained with 229 blink-, fixation-, and saccade-based features. Kunze and colleagues used the same eye tracker to detect reading activities of N=8 participants on five different media with variable amount and orientation of text: a comic book with images, a text book, fashion magazine, a novel, and a newspaper [Kunze et al. 2013]. With saccade- and fixation-based features, their decision tree classifier achieved a 74% accuracy in recognizing the type of document. In an outdoor wayfinding scenario, Alinaghi and colleagues studied the recognition of turning activities (left-, right-, or no-turn) with N=52 participants who had variable familiarity with the test routes of 0.9 km and 1.3 km [Alinaghi et al. 2021]. The data was collected with a 200 Hz PupilLabs Invisible eye tracker. They used feature importance ranking (on saccade- and fixation-based features) and tested several models, including SVM and Random Forest, and reached a 91% overall accuracy with Gradient Boosted Decision Trees.

2.2 Activity Recognition with Mobile Eye Tracking and AR

Toyama and colleagues presented a gaze-enabled (with SMI Eye Tracking Glasses) AR prototype [Toyama et al. 2015]. This prototype calculates whether the user's eyes converge on a foreground virtual screen or on the real scene (i.e., the background). While the point of convergence dynamically changes, the system analyzes the user's level of engagement in reading a text on the *virtual screen*. The proposed system provides proactive assistance such as highlighting, scrolling, and reminding the user about the last word read. Eight out of 12 participants rated the system as beneficial, however the system was tested only in a reading activity. Rook and colleagues studied intent prediction in an immersive environment with N=30 participants [Rook et al. 2019]. The 2 Hz data stream included users' head orientation (from Microsoft HoloLens 1) as an approximation to their eye-gaze and auxiliary data from objects of interest, and was used to train a hidden Markov Model (HMM) that yielded an average of 42% precision and 55% recall on three activities (cooking, microwaving, exploring). With Microsoft HoloLens 2 (HL2), Seelinger and colleagues developed a solution to enable safer navigation in a physical environment by presenting users with context-adaptive visual cues [Seeliger et al. 2022]. They trained a deep neural network (DNN) with features such as the angular change of gaze direction and the fixated areas of interest (AOIs), and also included task-specific features. The solution was not directly addressing the question of HAR with a gaze-enabled AR HMD, but their research provides evidence that a gaze-enabled AR display can promote users' autonomy and safety without compromising their performance. In a virtual reality setup (i.e., no interaction with physical objects as in AR), David-John and colleagues used an HTC Vive Pro Eye (with 60 to 120 Hz gaze sampling rate) to predict intentions of N=15 users regarding the selection of items for a given recipe (i.e., onset of interaction) [David-John et al. 2021]. Their logistic regression model was trained with 61 saccade- and fixation-based features as well as the K-coefficient (see [Krejtz et al. 2016]) and showed an above-chance prediction of the onset of interaction. Recently, Lan and colleagues addressed the creation of synthetic gaze data [Lan et al. 2022]. Their solution, EyeSyn, synthesizes realistic eye movement data for four activities (read, communicate, browse a static scene, watch a dynamic scene) using generative models and a range of image and video datasets. In an experimental study, the researchers compared the similarity and activity-recognition performance of EyeSyn-synthesized and actual gaze data collected from N=8 participants. Four participants used a Magic Leap One (30 Hz) and the others used Pupil Labs eye tracker (30 Hz). In all activities, a comparison of the scatterplots showed that the actual and synthetic data had similar spatial characteristics (e.g., reading activity involves horizontal shifts of the gaze). A convolutional neural network was trained with the synthetic data and achieved a 90% accuracy in the classification of the activities. Later, these authors also demonstrated that their solution can provide some AR feedback in two activities but with completely virtual stimuli [Scargill et al. 2022]. The developers of EyeSyn claim that it is a viable solution that can address practical constraints of collecting eye movement data and privacy-related concerns [Lan et al. 2022]. In Figure 2 we present a comparison of the selected previous work on HAR with mobile eye tracking devices and AR displays.

2.3 Data Privacy in Mobile Eye Tracking

Eye tracking data streams are invaluable sources of information, as they can reveal sensitive attributes of individuals (e.g., gender, age, ethnicity, personality traits, health, sexual preference, affect, task focus) [Kröger et al. 2020; Liebling and Preibusch 2014]. Thus, misusage of data from gaze-enabled devices can interfere with the acceptability of eye tracking by the general public [Bozkir et al. 2020; Orlosky et al. 2021] and, most importantly, infringe the privacy of individuals. Now that eye tracking is becoming pervasive, it should be added as a prominent privacy concern in ubiquitous computing technologies [Gressel et al. 2023; Katsini et al. 2020; Langheinrich 2001; Liebling and Preibusch 2014]. In recent years, interest in addressing privacy-related issues in eye tracking research (e.g., in the ETRA, UbiComp, and CHI communities) is growing [Katsini et al. 2020]. Privacy-preserving eye tracking can be maintained by physically obscuring the recordings [Steil et al. 2019b], introducing randomized encodings [Bozkir et al. 2020] or noise [Steil et al. 2019a] to the data (without compromising their utility), or by several other approaches and regulations [Liebling and Preibusch 2014].

Today we have access to mobile eye trackers, AR HMDs, and machine learning models that can be used in individual steps of HAR starting from data collection to activity recognition and providing feedback or assistance to users. To the best of our knowledge, no previous work provides a HAR solution in gaze-enabled AR HMDs where users perform different activities with physical objects. In Section 3, we introduce GEAR, that combines a (5-step) gaze-enabled activity recognition pipeline with an AR app for activity-based feedback. GEAR additionally integrates the Solid specification [Sambra et al. 2016] for storing and sharing data. This empowers users to control who accesses and manipulates their data. Hence, our work

Paper	Activities	Eye tracking (sampling rate in Hz)	N=	Stimuli / Display	AR?	HAR Model	Features	Performance	Feedback to User?
Bulling et al. 2011	**6** (copying, reading, writing, watching a video, browsing, and resting)	128 Hz EOG	8	Digital / Desktop	NO	SVM	**62** (fixations, saccades, blinks, wordbook analysis)	76.1% precision and 70.5% accuracy	NO
Kiefer et al. 2013	**6** (free exploration, global search, route planning, focused search, line following, and polygon comparison)	30 Hz SMI	17	Digital / Desktop	NO	SVM	**229** (fixations, saccades, and blinks)	78% accuracy	NO
Kunze et al. 2013	**1** reading different documents (comic book, text book, magazine, novel, and newspaper)	30 Hz SMI	8	Physical	NO	Decision Tree	saccade and fixation	74% accuracy	NO
Alinaghi et al. 2021	**1** wayfinding (turn left right, and no turn)	200 Hz Pupil Invisible	52	Physical	NO	Gradient Boosted Decision Trees (also SVM & Random Forest)	saccade and fixation (feature importance ranking)	91%	NO
Toyama et al. 2015	**1** reading	30 Hz SMI	12	Physical & Digital / HMD	YES	NO machine learning model	Convergence of the eyes on foregorund (AR overlay) or background (real scene)	8 of 12 liked	YES
Rook et al. 2019	**1** intent prediction in smart environments (cooking, microwaving, exploring)	2 Hz Head-gaze from HoloLens 1	30	Physical & Digital / HMD	YES	Hidden Markov Model	only point of interest	42% precision and 55% recall	NO
Seelinger et al. 2022	**1** navigation in a physical environment	30 Hz HoloLens 2	15	Physical & Digital / HMD	YES	Deep Neural Network	angular change of gaze direction and the fixated areas of interest	participants prefer the solution	YES
David-John et al. 2021	**1** prediction of the onset of item during selection of items for a given recipe	60 - 120 Hz HTC Vive Pro Eye	15	Digital / HMD	NO	Logistic Regression	**61** (saccade, fixation and the K-coefficient)	above chance prediction	NO
Lan and Scargill et al. 2022 Scargill and Lan et al. 2022	**4** (read, communicate, browse a static scene, watch a dynamic scene)	30 Hz Magic Leap One 30 Hz Pupil Labs	8	Digital / HMD	YES	Convolutional Neural Network	fixation and saccade based features	90% accuracy	YES
Bektaş et al. 2023	**3 (read, inspect, search)**	**30 Hz HoloLens 2 200 Hz PupilCore**	**10**	**Physical & Digital / HMD**	**YES**	**SVM, Random Forest, Extremely Randomized Trees**	**19** (fixation and saccade based features)	**98.7% accuracy**	**YES**

Figure 2: A comparison of the selected previous work on HAR with mobile eye tracking devices and AR displays.

provides a blueprint for a more privacy-friendly approach to the storing, processing, and sharing of gaze data.

3 GEAR: A GAZE-ENABLED AR SYSTEM FOR HUMAN ACTIVITY RECOGNITION

GEAR has three main components. The first one is an AR application that collects raw gaze data in real time from an AR HMD and renders activity-based feedback on top of a user's visual field (Figure 1-1.). The second component – Activity Recognition (Figure 1-2.) – implements a procedure for the real-time recognition of three activities (*Reading* a text, the *Inspection* of an object, and the *Search* for an object) from collected gaze data. The last component – Personal Datastore (Figure 1-3.) – implements a solution for the privacy-friendly sharing of such collected data.

3.1 AR Application

We developed an AR application for the HL2 with the Unity Game Engine[1] using building blocks from the Mixed Reality Toolkit (MRTK)[2]. The application has two main functions. In *Gaze Data Collection*, it fetches and sends the gaze data to the GEAR *Activity Recognition* component. The *User Feedback* prompts a visual feedback that is relevant to the recognized activity (Figure 3).

The HL2's eye tracker has a sampling rate of 30 Hz with an accuracy of approximately 1.5°[3]. In Unity, gaze samples can be accessed using the MRTK or the underlying API for the Universal Windows Platform (UWP)[4]. In GEAR, we use the open-source Augmented Reality Eye Tracking Toolkit (ARETT) [5] [Kapp et al. 2021]. ARETT operates on top of the UWP API, reliably delivers gaze samples at a fixed sampling rate (30 Hz) and can be readily included in Unity projects. It also provides a Web interface for storing gaze data in CSV files. The data stream provided by ARETT includes a list of time, gaze, and AOI data, and some auxiliary information. In GEAR's HAR, we make use of the following ARETT data: *eyeDataTimestamp*, *isCalibrationValid*, *gazeHasValue*, *gazeOrigin_(x/y/z)*, *gazeDirection_(x/y/z)*, *gazePoint_(x/y/z)*. The last three vectors are defined in Unity's global coordinate system. When we plotted these different values, we saw that *gazeDirection* might be the most suitable candidate to calculate gaze events and features, respectively. In Figure 3-(1-a), the gazeDirection data for Reading clearly shows the individual lines of the underlying text.

3.1.1 Gaze Data Collection. In a controlled study, we collected gaze data with the AR application (Subsection 3.1) for training and testing HAR models (Subsection 3.2). Our study is informed by the suggestions in [Holmqvist et al. 2022].

Participants: We recruited N=10 participants (3 female) from our lab, with an average age of 30.1 years. 4 participants reported wearing prescription glasses often or all of the time; 40% indicated being extremely familiar with AR headsets and 10% with VR headsets, while 40%/30% reported not being familiar with AR glasses/VR headsets. Most participants (80%, including all participants with moderate or extreme familiarity) reported that they could imagine wearing an AR headset for up to 2 hours in their daily lives.

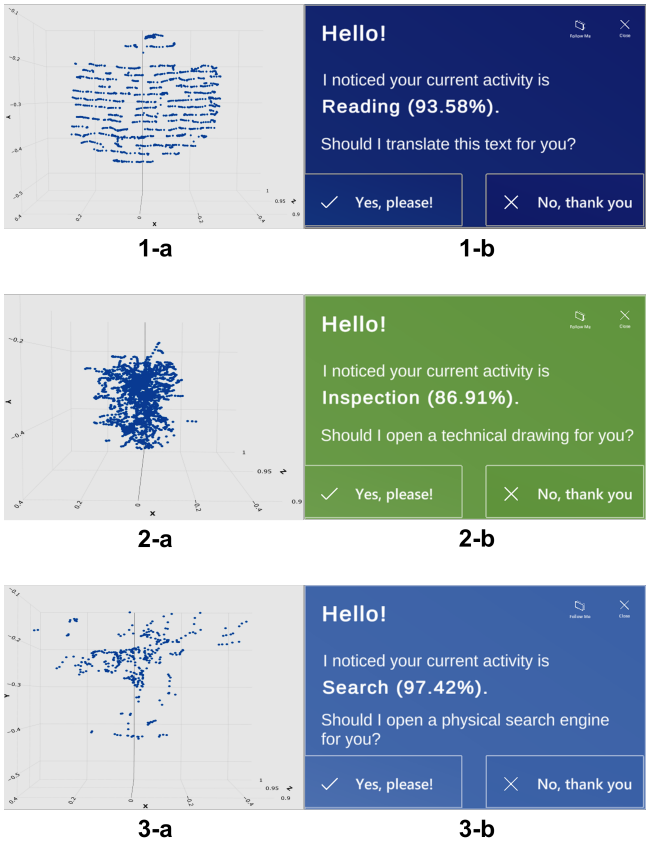

Figure 3: Example 3D-Plots of the normalized gazeDirection_(x,y,z) data points which are collected from one participant in the Reading (1-a), Inspection (2-a), and Search (3-a) activities. GEAR can display an AR feedback that is relevant to the recognized activity (1-b, 2-b , and 3-b).

Apparatus and Material: The gaze data was collected with the AR application and an HL2 as described in Subsection 3.1. Furthermore, in the same setup we collected gaze data with a Pupil Core tracker (200Hz). The Pupil Core data was not used in this study, however, we make it available as supplementary material for further analysis. Our study includes three different physical materials for each to-be-recognized activity. First, a text in English on an A4 paper vertically positioned at a distance of 70cm covering 40° of participants' visual field. Second, we used a toy robot that was positioned at a distance of approximately 40cm covering 20° of participants' visual field. Third, we used a small red pin (about the size of a die) and a workpiece-cabinet (1x1m) with three shelves.

Procedure: During the experiment, we simulated three realistic tasks that comprise three main activities: *reading* instructions, *inspecting* a device, and *searching* for a missing piece of the device. The procedure started with an introduction of the HL2, comfortably adjusting it on the head of the participant, and running the

[1] http://unity.com
[2] https://github.com/microsoft/MixedRealityToolkit-Unity/
[3] https://learn.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking
[4] https://learn.microsoft.com/de-de/uwp/api/windows.perception.people.eyespose
[5] https://github.com/AR-Eye-Tracking-Toolkit/ARETT

default 9-point eye calibration. Then, each participant was asked to read some text, inspect a robot, and search for a missing pin in a cabinet. The reading and inspection activities were performed in a sedentary position. When searching, participants were standing in front of a cabinet. Each activity included a gaze recording of about one minute and participants took short breaks of few seconds between the activities. Finally, we asked each participant to complete a short demographic questionnaire. According to the guidelines of the University of St. Gallen Ethics Committee, all participants gave written consent stating that the collected data can be anonymously used.

3.1.2 User Feedback. The AR application collects gaze data in chunks of ten seconds (i.e., consecutive and non-overlapping time windows) and sends them to the *Activity Recognition* component (details in Subsection 3.2). Each chunk includes the columns of *eyeDataTimestamp*, *isCalibrationValid*, *gazeHasValue*, *gazeOrigin_(x/y/z)*, *gazeDirection_(x/y/z)*, *gazePointHit*, and *gazePoint_(x/y/z)*. The data is sent via HTTP to a computer executing the *Activity Recognition* component. The response of the *Activity Recognition* component (Figure 1), including the predicted activity and its probability, is then sent back to the AR application. Thus, to close the activity-recognition and user-feedback loop, a user had to perform any of the three activities for at least ten seconds. Then, on the HL2, a panel displays the current activity (see Figure 3) along with contextually relevant suggestions. For example, during *Reading* activities, the application suggests a translation of the read text [Strecker et al. 2022]. During *Inspection* activities, the application suggests displaying a technical drawing of the inspected device. To help users find the missing item in *Search* activities, the application suggests whether it should open a semantic hypermedia search engine [Ciortea et al. 2020].

3.2 Activity Recognition

The activity recognition component of GEAR implements a procedure that is similar to those used in previous HAR research [Alinaghi et al. 2021; Bulling et al. 2011; Kiefer et al. 2013]. The procedure starts with the collection of raw gaze data as described in Section 3.1.1. The remaining steps include preprocessing of the raw data, detecting eye movement events, feature calculation, feature selection, and finally training and evaluation of selected machine learning model(s).

3.2.1 Pre-processing and Event Detection. We calculated eye movement events (fixations and blinks) from raw spatio-temporal gaze data (x,y,z,t). We did not compute saccades because of the limited sampling rate of the HL2. In a pre-processing step, before the fixation calculation, we excluded the data where *gazeOrigin* and *gazeDirection* were empty. With the remaining valid data, we calculated the fixations using the I-DT algorithm [Salvucci and Goldberg 2000] with a dispersion threshold of 1.6° and a minimum duration of 100 ms. For the dispersion, we used *gazeDirection* as the spatial input, which describes the normal of the gaze, i.e., the gaze direction in the global coordinate system. Our implementation of the fixation detection is based on a tutorial by Pupil Labs [6], which we adapted to our HL2 setup. Thus, the scripts that we provide in the supplementary material can be used to analyze data collected with the PupilCore [Kassner et al. 2014] or the HL2. To accurately detect blinks (see Chapter 5.7 in [Holmqvist et al. 2011]), the HL2's eye tracking sampling rate is low. Thus, we took the highly simplifying assumption that all missing gaze data were due to closed eyes/blinks.

3.2.2 Feature Calculation. We calculated 19 features from the descriptive statistics (minimum, maximum, mean, variance, and standard deviation) of the fixation duration (5 features) and of the fixation dispersion (5 features), the fixation frequency per second, and the fixation density. Additionally, we calculated the direction of successive fixations for x- and y-directions (2 features). Furthermore, we calculated the following blink-related features: number of blinks, mean, maximum, and minimum blink duration, and the blink rate per second.

In the *reading* activity, direction of successive fixations is decisive because the horizontal eye movements show a pattern that goes from left to right but then exhibits a larger jump from right to left when the participant finishes reading one line and proceeds to the next (similar to carriage-return and line-feed). However, in *inspection* and *search* activities, the eye movements do not necessarily follow a regular pattern, because typically there is no specific scene layout. In scene viewing (e.g., inspection or search activities), people may scrutinize different parts of the stimuli in variable duration [Holmqvist et al. 2011, 2022], thus the visual and spatial properties of targets and distractors may affect features extracted from fixations.

3.2.3 Feature Selection. Before selecting a subset of features for the classification, we formulated several assumptions for each activity. In the *reading* activity, we hypothesized that the successive fixations of the participants should be aligned with the lines of text they were reading. Furthermore, we expected the fixations to be more scattered and of shorter duration in the *search* activity, as participants probably looked quickly at many different places. In the *inspection* activity, we expected fewer fixations but with a longer duration, which are less scattered than those in the search. Based on these assumptions we trained our initial classifier (i.e., a support vector machine as described in 3.2.4) with the following six features: mean fixation duration, maximum fixation duration, variance of the fixation duration, x- and y-fixation direction and the fixation density per area. We then normalized these features and trained three different classifiers to predict the activities.

3.2.4 Model Evaluation. We trained three different machine learning models with the selected features. The performance benchmark of the *Support Vector Machine*, *Random Forest Classifier*, and *Extremely Randomized Trees Classifier* is presented in Table 1.

Support Vector Machine (SVM) Classifier: First, we applied an SVM classifier to the selected features, because SVMs were used in most of the related work on HAR that we documented in Section 2. We implemented the SVM using the `sklearn.svm.SVC` function in the Python package *scikit-learn*[7] and with ***L***inear, ***P***olynomial, Gaussian ***R***adial Basis Function, and ***S***igmoid kernels. We split the data in 80% training and 20% testing data. As the recorded data

[6] https://github.com/pupil-labs/pupil-tutorials

[7] https://scikit-learn.org

Table 1: Performance benchmark of the three models that are trained with 19 features. The number in each cell presents the accuracy (%) of the model (the first column of the table) for a 5, 10, 15, 20 seconds time window.

Model	5sec	10sec	15sec	20sec
Support Vector Machine	78.7	93.3	85.0	**100**
Random Forest	**93.2**	96.6	94.4	94.9
Extremely Randomized Trees	89.3	**98.7**	**96.2**	96.6

for each participant and activity was approximately one minute long, we trained the model with different time windows. With a ten-seconds window for the feature calculation we found the first three kernels to all predict with an accuracy of 93.3% and the Sigmoid kernel to achieve an accuracy of 30%. All kernels achieved lower prediction accuracies when using window sizes of five (L: 65.8%, P: 78.7%, R: 72.1%, S: 26.2%) and 15 seconds (L: 85%, P: 85%, R: 85%, S: 20%). A window size of 20 seconds, however, resulted in an accuracy of up to 100% (L: 86.6%, P: 93.3%, R: 100%, S: 33.3%) which might indicate overfitting to the small sample size.

Random Forest (RF) Classifier: Second, we developed a model with the *sklearn.ensemble.RandomForest* function from the above-mentioned *scikit-learn* library. To select the best subset of the pre-computed features for this classifier, forward and backward feature selection was performed. The results of our experiments show that the Random Forest classifier outperforms all SVM kernels by at least 4 percentage-points regarding accuracy. The best result (96.6% accuracy) was achieved using all of the 19 possible features. To be comparable to the SVM approach, a window of 10 seconds was used for the feature calculation. Other window sizes did not further improve the result (93.2% for 5 seconds, 94.4% for 15 seconds, 94.9% for 20 seconds).

Extremely Randomized Trees (ET) Classifier: Finally, we applied *sklearn.ensemble.ExtraTreesClassifier* while using all 19 features and a window size of 10 seconds. This classifier achieved an accuracy of 98.7% on our test data. As with RF, other window sizes did not improve accuracy (89.26% for 5 seconds, 96.25% for 15 seconds and 96.6% for 20 seconds). The ET classifier improves on the accuracy of the RF classifier while also being significantly faster, requiring 0.639s ±0.035s versus 0.844s ± 0.020s to classify 149 samples, i.e., 24% less time per sample. The difference between the ET and RF classifiers can be characterized as follows: While RF computes the most discriminative decision boundary for each feature, ET chooses the most discriminative boundary among several random boundaries and with different features subsets [Pedregosa et al. 2011]. As a consequence, variance is reduced, mitigating overfitting of the classifiers. Furthermore, by choosing the decision boundary randomly, the ET is computationally less expensive, leading to faster execution times.

3.3 Privacy-Friendly Personal Datastore

To handle the storage of gaze data and of recognized activities in a more privacy-friendly manner (compared to traditional means), we integrated GEAR with Solid. Solid is a specification for the creation of a decentralized data platform for Web applications [Sambra et al. 2016]– its main objective is to decouple applications from the data they use. To achieve this, a user is provided with a (personal) *Pod*, which is a repository that can contain her personal and non-personal data. A Pod is implemented as a Web server with standardized authentication, authorization, and sharing procedures. Through Access Control List (ACL) resources[8], a user can grant (and revoke) read, write, and append rights to Solid applications (or to people) to her full Pod or specific resources (i.e., individual files); access grantees and Pod owners are recognized through a unique identifier (i.e., a WebID). In the Solid system, applications that access user data do not keep a copy of this data, but rather access it (and possibly modify it) transiently, where access rights are checked on every access. Hence, given the sensitivity of users' gaze data, and of the predictions that our approach is capable of making on the current activity that a user might be performing, we set up an instance of the Solid community server, which is an open-source implementation of the Solid specification, developed and maintained by the research community[9]. Moreover, we added to our AR application the capability of writing the collected gaze data directly into a user's Pod in the form of a CSV file (see Figure 1 - GazeData.csv). Likewise, the Activity Recognition component stores the recognized activity in the user's Pod as an RDF file (see Figure 1 - Activity.ttl), which is expressed using well-known schemas (e.g., schema.org[10] and the PROV-O ontology[11]). The Personal Datastore component allows users to decide who to share their gaze and activity data with, which can be also done through the AR application.

4 CONCLUSIONS AND OUTLOOK

In this article, we presented GEAR, a system that allows human activity recognition in a gaze-enabled AR HMD. Using an Extremely Randomized Trees model, GEAR achieved an accuracy of 98.7% when recognizing three different activities in real time, using a window size of 10s of gaze data. The current implementation of GEAR updates user feedback at regular time intervals which is suitable for constrained experimental setups where users may perform the pre-trained activities in a particular order (i.e., read the instructions, inspect the robot, search for the missing piece). However, users have to deliberately continue performing an activity (i.e., without any interruption) to be able to receive some feedback that is relevant to that activity. In more realistic settings, various factors may affect users' mental state that can change from losing attention (e.g., a state of *mind-wandering* [Christoff et al. 2016]) to fully engaging with the current activity (i.e., the state of *flow* [Csikszentmihalyi 2014]). Thus, in a next step, GEAR can be extended with models (e.g., [Huang et al. 2019]) to better estimate users' mental state.

The work presented in this paper was conducted in the context of a graduate course on Ubiquitous Computing, where one of three assignments focused on Gaze-enabled AR. In addition to the code and data that is required to reproduce the results of this paper, we furthermore provide all teaching materials and their sources

[8] https://solid.github.io/web-access-control-spec/
[9] https://github.com/CommunitySolidServer/CommunitySolidServer
[10] https://schema.org/
[11] https://www.w3.org/TR/prov-o/

for reuse by others.[12] In the assignment, students gained experience working with gaze-enabled AR and different machine learning models by building on top of GEAR components. In Task 1 of this assignment, students worked on offline HAR using a Jupyter notebook that they were required to extend and improve to analyze our gaze dataset; in Task 2, they were required to extend a provided software framework for the HL2 to enable the close-to-real-time classification of user activities with the help of the model from Task 1, as described in this paper. Finally, Task 3 focused on the provisioning of AR feedback to the user, where we required students to provide contextual suggestions to users using *any* feedback modality (simple audio, spatial audio, visual feedback, etc.). In a subsequent assignment, the students were then required to enable the more privacy-friendly processing of gaze data through Solid by extending their applications and making use of our setup. As done in this paper, we motivated that assignment with a discussion on the high sensitivity of this data from a privacy perspective and that it may be used to not only estimate the activity that a user currently performs, but can also be used to estimate other users' personal information, from drug consumption to cultural background [Kröger et al. 2020]. Students were then required to implement a gaze-enabled activity recognition pipeline that made use of their Web-accessible personal Pods with the Solid specification, where access rights were granted based on WebID (concretely: Solid OpenID Connect) together with Solid Access Control Lists.

As a next step, we will implement the activity recognition classifier in C# instead of Python so that it can run directly on the HL2. The SharpLearning[13] library for C# provides implementations for RandomForest and Extremely Randomized Trees classifiers, while the event detection algorithms can be implemented analogously to the Python implementation. Using C#'s concurrency support, the activity classifier can be run concurrently with the data collection and feature extraction. We will also extend the user feedback part with other modalities (e.g., audio cues, or speech interfaces), relevant Web-based services (e.g., Optical Character Recognition [Strecker et al. 2022], Object Detection [Spirig et al. 2021]) and by defining dynamic AOIs to make it more useful for users. Additionally, we will explore how we can include the Extended Eye Tracking API[14] for collecting gaze data on the HL2 with a higher frame rate. Lastly, we will test an extended list of features with our system in the recognition of other activities.

ACKNOWLEDGMENTS

This work was funded by the Swiss Innovation Agency Innosuisse (#48342.1 IP-ICT) and the Basic Research Fund of the University of St.Gallen.

REFERENCES

Negar Alinaghi, Markus Kattenbeck, Antonia Golab, and Ioannis Giannopoulos. 2021. Will You Take This Turn? Gaze-Based Turning Activity Recognition During Navigation. *11th International Conference on Geographic Information Science (GIScience 2021) - Part II* (2021), 5:1–5:16. https://doi.org/10.4230/LIPIcs.GIScience.2021.II.5

Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47. https://doi.org/10.1109/38.963459

Kenan Bektaş. 2020. Toward A Pervasive Gaze-Contingent Assistance System: Attention and Context-Awareness in Augmented Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Adjunct)*. ACM, New York, NY, USA, Article 36, 3 pages. https://doi.org/10.1145/3379157.3391657

Kenan Bektas, Arzu Cöltekin, Jens Krüger, and Andrew T. Duchowski. 2015. A Testbed Combining Visual Perception Models for Geographic Gaze Contingent Displays. *Eurographics Conference on Visualization (EuroVis) - Short Papers* (2015). https://doi.org/10.2312/EUROVISSHORT.20151127

Kenan Bektaş, Jannis Rene Strecker, Simon Mayer, and Markus Stolze. 2022. EToS-1: Eye Tracking on Shopfloors for User Engagement with Automation. In *AutomationXP22: Engaging with Automation, CHI'22*. CEUR Workshop Proceedings. http://www.alexandria.unisg.ch/266339/

Kenan Bektaş, Tyler Thrash, Mark A van Raai, Patrik Künzler, and Richard Hahnloser. 2021. The systematic evaluation of an embodied control interface for virtual reality. *Plos one* 16, 12 (2021). https://doi.org/10.1371/journal.pone.0259977

Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272. https://doi.org/10.1561/1100000049

A. Borji and L. Itti. 2014. Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision* 14, 3 (March 2014), 29–29. https://doi.org/10.1167/14.3.29

Efe Bozkir, Ali Burak Ünal, Mete Akgün, Enkelejda Kasneci, and Nico Pfeifer. 2020. Privacy Preserving Gaze Estimation using Synthetic Images via a Randomized Encoding Based Framework. In *ACM Symposium on Eye Tracking Research and Applications*. ACM, Stuttgart Germany, 1–5. https://doi.org/10.1145/3379156.3391364

Christian Braunagel, Enkelejda Kasneci, Wolfgang Stolzmann, and Wolfgang Rosenstiel. 2015. Driver-Activity Recognition in the Context of Conditionally Autonomous Driving. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, Gran Canaria, Spain, 1652–1657. https://doi.org/10.1109/ITSC.2015.268

Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *Comput. Surveys* 46, 3 (Jan. 2014), 1–33. https://doi.org/10.1145/2499621

Andreas Bulling and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12. https://doi.org/10.1109/MPRV.2010.86

Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Tröster. 2011. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 4 (April 2011), 741–753. https://doi.org/10.1109/TPAMI.2010.86

Andreas Bulling and Thorsten O. Zander. 2014. Cognition-Aware Computing. *IEEE Pervasive Computing* 13, 3 (July 2014), 80–83. https://doi.org/10.1109/MPRV.2014.42

Kalina Christoff, Zachary C Irving, Kieran CR Fox, R Nathan Spreng, and Jessica R Andrews-Hanna. 2016. Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience* 17, 11 (2016), 718–731. https://doi.org/10.1038/nrn.2016.113

Andrei Ciortea, Simon Mayer, Simon Bienz, Fabien Gandon, and Olivier Corby. 2020. Autonomous search in a social and ubiquitous Web. *Personal and Ubiquitous Computing* (June 2020). https://doi.org/10.1007/s00779-020-01415-1

Maria Cornacchia, Koray Ozcan, Yu Zheng, and Senem Velipasalar. 2017. A Survey on Activity Detection and Classification Using Wearable Sensors. *IEEE Sensors Journal* 17, 2 (Jan. 2017), 386–403. https://doi.org/10.1109/JSEN.2016.2628346

Mihaly Csikszentmihalyi. 2014. Toward a psychology of optimal experience. In *Flow and the foundations of positive psychology*. Springer, 209–226.

Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*. ACM, Virtual Event Germany, 1–7. https://doi.org/10.1145/3448018.3458008

Céline Gressel, Rebekah Overdorf, Inken Hagenstedt, Murat Karaboga, Helmut Lurtz, Michael Raschke, and Andreas Bulling. 2023. Privacy-Aware Eye Tracking: Challenges and Future Directions. *IEEE Pervasive Computing* 22, 1 (2023), 95–102. https://doi.org/10.1109/MPRV.2022.3228660

Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (June 2017), 1706–1724. https://doi.org/10.1109/TVCG.2016.2543720

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Kenneth Holmqvist, Saga Lee Örbom, Ignace T. C. Hooge, Diederick C. Niehorster, Robert G. Alexander, Richard Andersson, Jeroen S. Benjamins, Pieter Blignaut, Anne-Marie Brouwer, Lewis L. Chuang, Kirsten A. Dalrymple, Denis Drieghe, Matt J. Dunn, Ulrich Ettinger, Susann Fiedler, Tom Foulsham, Jos N. van der Geest, Dan Witzner Hansen, Samuel B. Hutton, Enkelejda Kasneci, Alan Kingstone, Paul C. Knox, Ellen M. Kok, Helena Lee, Joy Yeonjoo Lee, Jukka M. Leppänen, Stephen

[12] https://github.com/Interactions-HSG/GEAR/tree/main/GEAR-4-HAR
[13] https://github.com/mdabros/SharpLearning
[14] The API reportedly allows up to 90Hz sampling (https://learn.microsoft.com/en-us/windows/mixed-reality/develop/unity/extended-eye-tracking-unity).

Macknik, Päivi Majaranta, Susana Martinez-Conde, Antje Nuthmann, Marcus Nyström, Jacob L. Orquin, Jorge Otero-Millan, Soon Young Park, Stanislav Popelka, Frank Proudlock, Frank Renkewitz, Austin Roorda, Michael Schulte-Mecklenbeck, Bonita Sharif, Frederick Shic, Mark Shovman, Mervyn G. Thomas, Ward Venrooij, Raimondas Zemblys, and Roy S. Hessels. 2022. Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior Research Methods* (April 2022). https://doi.org/10.3758/s13428-021-01762-8

Michael Xuelin Huang, Jiajia Li, Grace Ngai, Hong Va Leong, and Andreas Bulling. 2019. Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2254–2262. https://doi.org/10.1145/3343031.3350573

Rob Jacob and Sophie Stellmach. 2016. What you look at is what you get: gaze-based user interfaces. *Interactions* 23, 5 (Aug. 2016), 62–65. https://doi.org/10.1145/2978577

Sebastian Kapp, Michael Barz, Sergey Mukhametov, Daniel Sonntag, and Jochen Kuhn. 2021. ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays. *Sensors* 21, 6 (March 2021), 2234. https://doi.org/10.3390/s21062234

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-Based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

Christina Katsini, Yasmeen Abdrabou, George E. Raptis, Mohamed Khamis, and Florian Alt. 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–21. https://doi.org/10.1145/3313831.3376840

Ashima Keshava, Farbod Nosrat Nezami, Henri Neumann, Krzysztof Izdebski, Thomas Schüler, and Peter König. 2021. Just-in-time: gaze guidance behavior while action planning and execution in VR. *bioRxiv* (2021).

Peter Kiefer, Ioannis Giannopoulos, and Martin Raubal. 2013. Using eye movements to recognize activities on cartographic maps. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, New York, NY, USA, 488–491. https://doi.org/10.1145/2525314.2525467

Krzysztof Krejtz, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska, and Agata Kopacz. 2016. Discerning Ambient/Focal Attention with Coefficient *K*. *ACM Transactions on Applied Perception* 13, 3 (May 2016), 1–20. https://doi.org/10.1145/2896452

Jacob Leon Kröger, Otto Hans-Martin Lutz, and Florian Müller. 2020. What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking. In *Privacy and Identity Management. Data for Better Living: AI and Privacy*, Michael Friedewald, Melek Önen, Eva Lievens, Stephan Krenn, and Samuel Fricker (Eds.). Vol. 576. Springer International Publishing, Cham, 226–241. https://doi.org/10.1007/978-3-030-42504-3_15 Series Title: IFIP Advances in Information and Communication Technology.

Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. 2013. I know what you are reading: recognition of document types using mobile eye tracking. In *Proceedings of the 2013 International Symposium on Wearable Computers*. ACM, New York, NY, USA, 113–116. https://doi.org/10.1145/2493988.2494354

Guohao Lan, Tim Scargill, and Maria Gorlatova. 2022. EyeSyn: Psychology-inspired Eye Movement Synthesis for Gaze-based Activity Recognition. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, Milano, Italy, 233–246. https://doi.org/10.1109/IPSN54338.2022.00026

Marc Langheinrich. 2001. Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems. In *Ubicomp 2001: Ubiquitous Computing*, Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, Gregory D. Abowd, Barry Brumitt, and Steven Shafer (Eds.). Vol. 2201. Springer Berlin Heidelberg, Berlin, Heidelberg, 273–291. https://doi.org/10.1007/3-540-45427-6_23

Daniel J. Liebling and Sören Preibusch. 2014. Privacy considerations for a pervasive eye tracking world. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1169–1177. https://doi.org/10.1145/2638728.2641688

Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

Jason Orlosky, Misha Sra, Kenan Bektaş, Huaishu Peng, Jeeeun Kim, Nataliya Kos'myna, Tobias Höllerer, Anthony Steed, Kiyoshi Kiyokawa, and Kaan Akşit. 2021. Telelife: The Future of Remote Living. *Frontiers in Virtual Reality* 2 (Nov. 2021), 763340. https://doi.org/10.3389/frvir.2021.763340

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *Comput. Surveys* 55, 3 (March 2022), 1–39. https://doi.org/10.1145/3491207

Kelsey Rook, Brendan Witt, Reynold Bailey, Joe Geigel, Peizhao Hu, and Ammina Kothari. 2019. A Study of User Intent in Immersive Smart Spaces. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, Kyoto, Japan, 227–232. https://doi.org/10.1109/PERCOMW.2019.8730692

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. ACM, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

Andrei Vlad Sambra, Essam Mansour, Sandro Hawke, Maged Zereba, Nicola Greco, Abdurrahman Ghanem, Dmitri Zagidulin, Ashraf Aboulnaga, and Tim Berners-Lee. 2016. Solid: a platform for decentralized social applications based on linked data. *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep.* (2016).

Tim Scargill, Guohao Lan, and Maria Gorlatova. 2022. Demo Abstract: Catch My Eye: Gaze-Based Activity Recognition in an Augmented Reality Art Gallery. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, Milano, Italy, 503–504. https://doi.org/10.1109/IPSN54338.2022.00052

Arne Seeliger, Raphael P. Weibel, and Stefan Feuerriegel. 2022. Context-Adaptive Visual Cues for Safe Navigation in Augmented Reality Using Machine Learning. *International Journal of Human–Computer Interaction* (Sept. 2022), 1–21. https://doi.org/10.1080/10447318.2022.2122114

Janick Spirig, Kimberly Garcia, and Simon Mayer. 2021. An Expert Digital Companion for Working Environments. In *Proceedings of the 11th International Conference on the Internet of Things* (St.Gallen, Switzerland) *(IoT '21)*. ACM, New York, NY, USA, 25–32. https://doi.org/10.1145/3494322.3494326

Julian Steil, Inken Hagestedt, Michael Xuelin Huang, and Andreas Bulling. 2019a. Privacy-aware eye tracking using differential privacy. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/3314111.3319915

Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. 2019b. PrivacEye: privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/3314111.3319913

Jannis Strecker, Kimberly García, Kenan Bektaş, Simon Mayer, and Ganesh Ramanathan. 2022. SOCRAR: Semantic OCR through Augmented Reality. In *Proceedings of the 12th International Conference on the Internet of Things*. ACM, Delft Netherlands, 25–32. https://doi.org/10.1145/3567445.3567453

Ivan E. Sutherland. 1968. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*. ACM Press, San Francisco, California, 757–764. https://doi.org/10.1145/1476589.1476686

Takumi Toyama, Daniel Sonntag, Jason Orlosky, and Kiyoshi Kiyokawa. 2015. Attention Engagement and Cognitive State Analysis for Augmented Reality Text Display Functions. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, Atlanta Georgia USA, 322–332. https://doi.org/10.1145/2678025.2701384

Roel Vertegaal (Ed.). 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (March 2003), 3263733. https://doi.org/10.1145/3263733

Mark Weiser. 1999. The Computer for the 21st Century. *SIGMOBILE Mob. Comput. Commun. Rev.* 3, 3 (jul 1999), 3–11. https://doi.org/10.1145/329124.329126

Alfred L. Yarbus. 1967. *Eye Movements and Vision*. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4899-5379-7

Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Pittsburgh, Pennsylvania, United States, 246–253. https://doi.org/10.1145/302979.303053