

# Visual Gaze Labeling for Augmented Reality Studies

S. Öney<sup>†</sup>, N. Pathmanathan, M. Becher, M. Sedlmair, D. Weiskopf, and K. Kurzhals

VISUS, University of Stuttgart, Germany



**Figure 1:** Overview of the presented visualization approach for the annotation of areas of interest (AOIs) in augmented reality. (a) A 3D gaze player and (b) a linked timeline view show fixations of participants on AOIs. The overview can be (f) zoomed and filtered for (c) participants and (d) AOI labels. (e) Different view options allow encoding annotated fixations differently to examine various aspects.

## Abstract

Augmented Reality (AR) provides new ways for situated visualization and human-computer interaction in physical environments. Current evaluation procedures for AR applications rely primarily on questionnaires and interviews, providing qualitative means to assess usability and task solution strategies. Eye tracking extends these existing evaluation methodologies by providing indicators for visual attention to virtual and real elements in the environment. However, the analysis of viewing behavior, especially the comparison of multiple participants, is difficult to achieve in AR. Specifically, the definition of areas of interest (AOIs), which is often a prerequisite for such analysis, is cumbersome and tedious with existing approaches. To address this issue, we present a new visualization approach to define AOIs, label fixations, and investigate the resulting annotated scanpaths. Our approach utilizes automatic annotation of gaze on virtual objects and an image-based approach that also considers spatial context for the manual annotation of objects in the real world. Our results show, that with our approach, eye tracking data from AR scenes can be annotated and analyzed flexibly with respect to data aspects and annotation strategies.

## CCS Concepts

- Human-centered computing → Visualization;

## 1. Introduction

Augmented reality (AR) has seen a revival in recent years, mainly through improved hardware and easy-to-use software that allows

<sup>†</sup> e-mail:Seyda.Oney@visus.uni-stuttgart.de

© 2023 The Authors. Computer Graphics Forum published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

researchers to design augmented content without having to deal with the underlying computer vision technology. This technology provides new ways to display information in the spatial context of the real world. Examples comprise training and support in medical procedures [VRZ<sup>\*</sup>17], industrial manufacturing [FLFCBNVM18], as well as data visualization for situated analysis [BHM<sup>\*</sup>22]. Evaluation of AR applications is mainly conducted by qualitative methods such as interviews and observations, as well as quantitative measurements [DB11, Liv05]. Quantitative methods are often restricted to performance measures (*How long did it take to solve the task? How correct was the result?*), often leaving the question open: *Why were some participants performing better than others?* Investigating visual task solution strategies can provide answers to this question, as it might reveal where people spent attention and important parts of a task that caused problems.

Eye tracking is a means to evaluate visual stimuli (e.g., visualization techniques [KFBW16]) quantitatively and qualitatively. The measured gaze distribution is an indicator of visual attention and can be investigated sequentially to derive a detailed temporal analysis of individual steps while performing a task [AABW12]. Modern head-mounted displays (HMDs) include eye tracking as means for gaze-based interaction which can also be recorded for post-experimental analysis to understand perceptual and cognitive aspects of the task at hand [KKBW22]. In traditional eye tracking analysis, mapping of gaze to semantic objects or areas of interest (AOIs) is a common way to enrich fixations with world knowledge that also improves the comparability of data from different people [BKR<sup>\*</sup>17]. AR scenarios pose a challenge for traditional analysis approaches for multiple reasons:

- Movement and gaze data of multiple participants is often not recorded in a common world-coordinate system, but in individual coordinates per participant. This is comparable to issues with spatial comparability of data recorded with eye tracking glasses [KHSW16].
- AOIs in AR scenarios can be virtual or real. Virtual content can be identified automatically, whereas real AOIs pose a classification problem. For scenarios with predefined AOIs, this classification can potentially be trained algorithmically [WHB<sup>\*</sup>18]. In scenarios where AOIs are not known in advance (e.g., uncontrolled environments), manual annotation is often necessary.
- Video-based manual annotation of AOIs is time-consuming and neglects the 3D spatial context of the data. Multiple approaches for the annotation of such mobile gaze data were proposed in the past [BKR<sup>\*</sup>17], mainly focusing on 2D images recorded by a world-view camera of wearable eye tracking devices.

Hence, although an AOI-based analysis of gaze data from AR scenarios is achievable with established methods such as video-based manual annotation, the data processing results in annotations of real-world AOIs without spatial information in the context of the environment. Especially in AR, this spatial context is essential to understand how people interacted with the virtual and the real surroundings. For example, investigating navigation support in orientation tasks where the location of AOIs is important to investigate spatial cognition processes. Furthermore, visualization techniques depicting movement trajectories and scanpaths of people's gaze from AR focus on single participant analyses. However, to derive generalizable findings about behavior patterns, higher

participant numbers have to be investigated. Current techniques mainly show single trajectories [MT21] or aggregated gaze distributions [LSO20]. They do not support this type of multi-user analysis and annotation is still necessary to investigate data with AOI-based methods.

We contribute a new visualization-based approach to annotate and interpret gaze data of multiple participants simultaneously without neglecting the spatial context of the data (Figure 1). Our focus lies on AR scenarios with a combination of virtual and real AOIs. We display the data by extending the gaze stripes technique [KHH<sup>\*</sup>15], a temporal overview of investigated content based on thumbnail images. In a second linked view, we provide a detailed replay of gaze and movement in the 3D spatial context of the scene. Fixations on virtual content are labeled and visualized automatically. For the remaining unlabeled fixations, we compare different annotation techniques and their applicability for AR scenes: (1) Direct fixation labeling based on the point of regard, (2) image-based labeling with thumbnails and the definition of bounding boxes in 3D space. We showcase our approach with an experiment where people investigated a collection of artwork enriched by interactive virtual content and domain experts annotated this data.

Our results show similar annotation times for both techniques. Labeling in 3D space was preferred by most because the spatial context allows for more intuitive labeling and provides a deeper understanding of the space. The annotated data as well as the source code of our implementation are openly available [OPB<sup>\*</sup>23].

## 2. Related Work

We focus our discussion of related work on eye-tracking-based evaluation in general and how AOI-based analysis is currently applied. Further, we provide an overview of current applications of eye tracking in the context of AR scenarios.

### 2.1. Eye Tracking for User-based Evaluation

Eye tracking for evaluation purposes has a long tradition in research fields such as psychology, cognitive science [Duc17], human-computer interaction [PB06], and visualization [KFBW16]. However, evaluation scenarios consider mostly desktop applications (e.g., examining reading behavior [BBHD10]) and mobile eye tracking with wearable devices (e.g., how people perform everyday tasks [HB05]). Experiments in VR also benefit from eye tracking for the interpretation of viewing behavior [CKK19, MPPO19]. In contrast, AR technology just recently became feasible for the application of eye tracking and therefore opens a new field of research [KBPR22] to adjust established techniques and develop new methods to gain insights into how people use AR applications.

### 2.2. AOI-based Gaze Analysis

As we will further outline in Section 4, there are different approaches to derive AOIs or labels for fixations on AOIs, respectively. Boundary shapes of relevant objects (e.g., [BCNS15]) provide the geometry for testing if the gaze lies inside the respective region. Alternatively, gaze samples or fixations can be labeled directly. This is usually performed by showing individual gaze coordinates on the visual stimulus and letting the annotator assign

the correct label (e.g., [NBW16]). Image labeling based on thumbnails of stimulus regions around the point of regard has been integrated into visualization techniques to improve the annotation process [Kur21, PKP10]. However, these techniques have only been applied to data from remote and wearable eye tracking devices. Although such techniques could also be applied to AR scenarios, the spatial context would be neglected. As a consequence, later analyses could only interpret gaze as a sequence of visited AOIs without a reference to where these AOIs were located. To close this gap, we propose a technique that takes advantage of the context recorded by the device to support an efficient annotation of virtual and real AOIs. As a consequence, we can combine the advantages of spatial annotation with image-based techniques to provide a visualization approach to annotate and interpret gaze patterns from multiple participants recorded in AR.

### 2.3. Evaluation of Augmented Reality

Evaluation of AR applications is often performed with classical performance analysis and qualitative techniques such as observations and interviews [DB11]. Eye tracking provides an additional means to derive insights beyond user performance and can help detect design issues and understand viewing behavior [KFBW16]. For desktop applications, this technique was applied many times [GSL\*02, PHG\*04]. The interest of using eye tracking for AR scenarios increased in recent years, partially due to the availability of the technique in current hardware generations. As an example, there are multiple approaches presenting heat maps and trajectories in a 3D context [SG22]. Although this is valuable information about gaze distributions in general, a comparison between many participants is hard to achieve this way. Consequently, AOIs become necessary to investigate scanpaths semantically. To the best of our knowledge, there is no approach fit for the requirements of AOI-based analysis in AR scenarios. To solve this issue, fixation-based labeling can be applied as a general approach to perform this task sequentially. Further, spatial annotation allows one to annotate data from multiple participants in parallel. A thorough comparison between techniques is outlined in Section 5.

## 3. Data Processing

The data acquisition requires an HMD that supports eye tracking (e.g., Microsoft HoloLens 2). Unity was employed to create the AR scene to capture the viewers' movements and collect 3D gaze data, both of which are needed to provide the spatio-temporal context in the visualization framework later. The Mixed Reality Toolkit (MRTK) [Mic16] was integrated into our scene to make Unity AR-capable and facilitate access to various HoloLens features. The ARETT toolkit [KBM\*21a] was utilized to capture gaze with a stable sampling rate. With this setup, we collected important data that was essential for preparing our visualization framework:

**Hitted object**, which contained the gaze hit with the object. In our case, the virtual objects corresponded to the virtual AOIs.

**Transformed 3D gaze position** in a global coordinate system. From these data, we extracted fixations using the velocity threshold algorithm (I-VT) [SG00] provided as an R package [KBM\*21b].

**Projected gaze point** onto the camera image recorded by the HMD. Thumbnails for image-based annotation were created from the projection of the identified fixations with the video recordings captured during the pilot experiment.

Gaze replay requires a mesh representation of the environment. Therefore, the room was scanned with the HoloLens at the start of the pilot experiment to produce an environment mesh with SLAM-based spatial mapping methods [CCC\*16]. These methods allow localizing the viewers while constructing a mesh from their surrounding. Since the produced mesh was not textured, a second, textured photogrammetry mesh was used and its position, size, and rotation were manually adjusted to match the spatial mesh.

The collected gaze and mesh data of different participants must be in a common coordinate system to provide comparable data of movement and gaze behavior within the replay. For this, we applied World Locking Tools (WLT) [Mic22]. We used 4 QR markers to transform the coordinate space so that the origin was moved from the head position defined at the beginning of the application to a common physical start position. WLT also provides persistence across sessions by storing spatial anchors locally. This technique also avoids inconsistencies in QR marker recognition. The *world-locked* state was then reloaded for different participants.

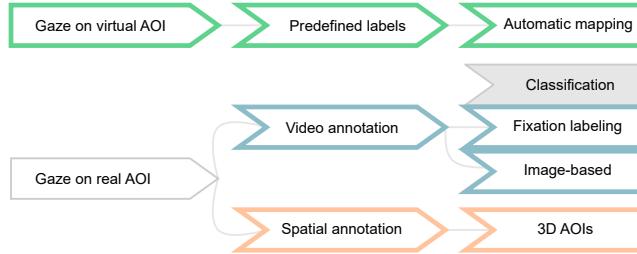
## 4. Technique

Our development was divided into two stages: collecting eye tracking data through a pilot experiment that serves as a case study, and the final design of the visualization framework.

For the design of our approach to annotate gaze data in AR, we first investigated common approaches to solve this task in scenarios of eye tracking with glasses and in VR. We decided on a hybrid approach that handles virtual AOIs automatically and provides an image-based technique in combination with spatial AOIs in 3D. Figure 2 depicts an overview of how AOIs are typically handled. Virtual AOIs can be identified automatically if the virtual scene design is controlled by the people conducting the experiment. Meaningful labels for virtual objects or areas of a virtual mesh can be defined and later be used for hit detection with gaze rays to find out what a person was investigating. Limitations arise if an object consists of multiple zones that need to be separated into different AOIs, or small AOIs in general. To compensate for accuracy issues, it is common practice to add border margins. For both cases, this might lead to intersections between AOIs and ambiguities where the gaze was directed to. Real-world AOIs can be addressed in similar ways as data from mobile eye tracking devices, i.e., by annotations based on an egocentric video. However, the advantage of AR devices is that the multitude of sensors provides rich information about the spatial context of an experiment. Spatial annotation, i.e., defining boundary shapes for AOIs, either in 2D or 3D is the basis for fixation labeling based on hit detection. The recorded spatial context, for instance, represented by a SLAM-based mesh allows performing hit detection in defined areas.

### 4.1. Design Decisions

According to Blascheck et al. [BKR\*17], visualization techniques based on AOIs mainly comprise transition matrices, timelines



**Figure 2:** With predefined labels, gaze on virtual AOIs can be processed automatically (green). For real AOIs, video-based annotation is standard, either by automatic classification, manual fixation labeling, or image-based techniques (blue). Spatial annotation in 3D requires a common coordinate system or semantically matching AOIs between different recordings (orange).

showing AOI visits (e.g., scarf plots [SND10]), as well as transition graphs embedded in the stimulus. Our main focus is on annotation support and an overview of scanpaths. Transition matrices show only pairwise transitions between AOIs and scanpaths embedded in the 3D environment tend to create visual clutter. Hence, we decided to keep the 3D context for details and expand the concept of AOI-based timelines for the interpretation of scanpaths. As indicated in Figure 2, we address the annotation problem for AR scenarios by a combination of three different approaches:

- Augmented, virtual elements are designed with predefined labels, and gaze on elements is processed automatically by mesh-based hit detection.
- Real-world AOIs are annotated with an image-based visualization technique after virtual AOIs are processed.
- The annotation of real-world AOIs is further supported by spatial annotations in a 3D model of the environment.

With this combination of approaches, we support static AOIs that do not have to be trained by an algorithm. Image-based and spatial annotation provides flexible strategies to label gaze: Frequently visited AOIs can be annotated in 3D space, allowing efficient labeling of all fixations in this area. Image-based labeling helps identify where AOIs are, judge ambiguous cases outside of margin areas, and annotate data where it would be more effort to draw AOIs in 3D space. Hence, our resulting visualization framework consists of a 3D scene view and timelines that represent fixations of individual participants by thumbnail images. Both views are linked and complement each other through spatial (3D view) and temporal (timelines) representations of the data.

#### 4.2. Visualization Framework

Our visualization framework consists of two main components (Figure 1) – the gaze replay and a timeline visualization – linked together to enable spatial and image-based annotation. The following paragraphs first describe the components in detail. The supported annotation techniques are explained in Section 4.3.

**Gaze Replay** A textured mesh is integrated into the gaze replay to reconstruct the AR scene from an experiment and simulate the par-

ticipants' movements and gaze behavior in a 3D view. The participants are represented by spheres with their names visible as floating labels. A ray is emitted from the sphere to represent the gaze direction. Movement within the 3D view is enabled to examine the AR scene from different perspectives. The individuals' points of regard are shown during replay and linked for analyses using the timeline. The gaze points up to the current point in time can be visualized by little spheres to highlight the regions already viewed. The spatial annotation is done in the gaze replay and provides a spatial context for the fixations, which is explained in Section 4.3.

**Timeline Visualization** The gaze data and video recordings collected during the experiment are used to create the timeline visualization. Here, the fixations extracted from the gaze data are positioned on the timeline. Each fixation is represented by a thumbnail image obtained from the video recording. The height and width of the image represent the distance of the fixation from the viewer, and a bar around the thumbnail indicates the length of the fixation (Figure 4a). The distance to the investigated AOI is encoded by the height of the displayed bar and decreases when the AOI is investigated from farther away. The consecutive fixations of each participant are positioned horizontally in separate timelines for each participant. The sliders in the timeline and the gaze replay are linked to show the data synchronized from different perspectives. Moving the slider at any point in time shows the position of the participants and their point of regard in the gaze replay. The timeline visualization provides a horizontal and vertical zoom slider to switch between a detailed view and an overview of all participants and all fixations (Figure 1(f)). The timeline visualization supports three different view modes (Figure 3) to provide more information about the labeled fixations. Enabling one of these views (Figure 1(e)) provides more information about the labeled fixations:

**Annotated fixations** are grayed out and do not allow interaction. Only the unlabeled fixations can be selected and labeled for image-based annotation (Section 4.3). In the overview mode, the approximate number and position of the remaining unlabeled fixations can be determined (Figure 3a).

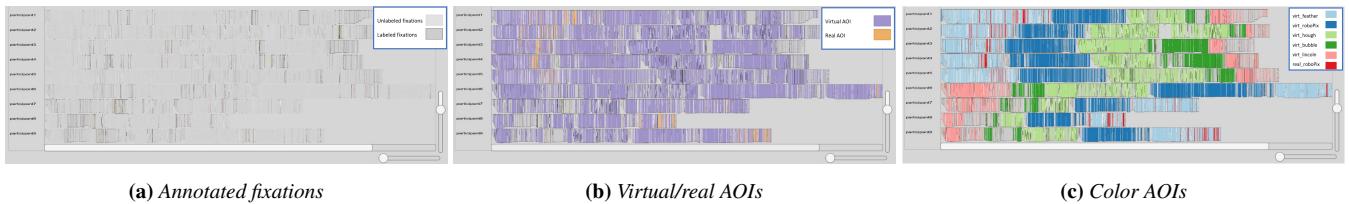
**Virtual/real AOIs** display all labeled fixations in the respective category, according to the AOI legend. In the overview, the ratio of virtual and real AOIs can be detected and AOI patterns in the fixation sequences can be observed (Figure 3b).

**Color AOIs** present a view in which all labeled fixations are colored according to their AOIs. After the labeling process is complete, the AOI sequences and patterns can be identified in the overview mode (Figure 3c).

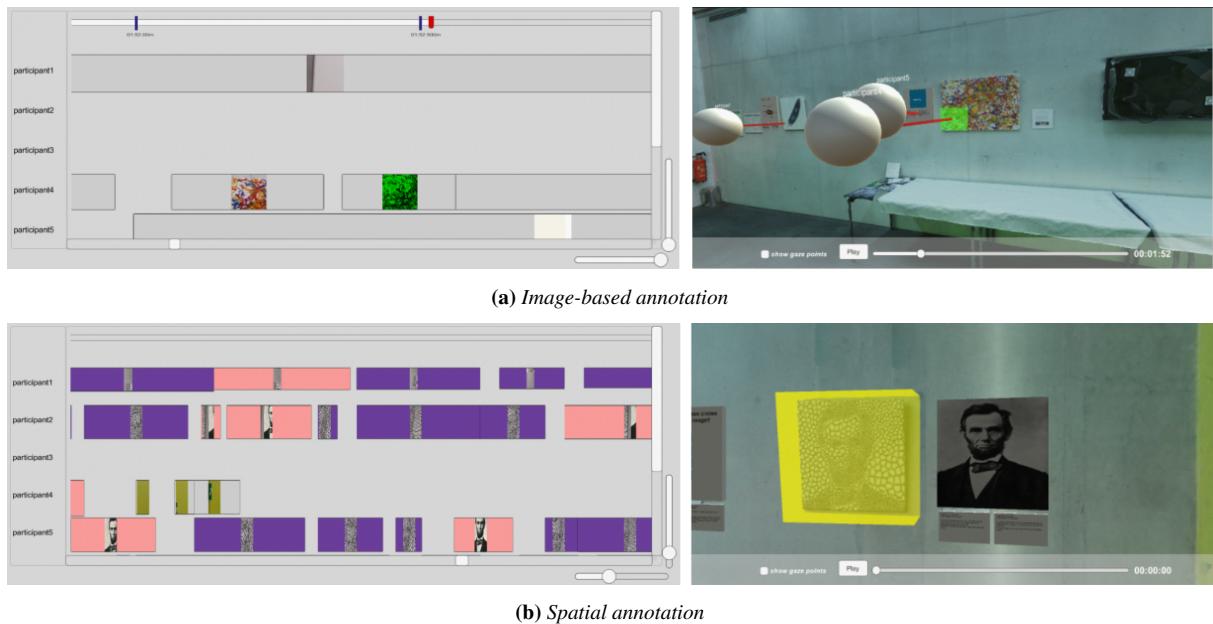
#### 4.3. Annotation Support

We consider three techniques and possible filtering operations to constrain the data to be labeled. Here, image-based and spatial annotation are part of our visualization framework.

**Fixation-based Annotation** We see fixation-based labeling as the baseline approach we want to compare against because it can be performed in most software suites of eye tracking vendors and is independent of dynamically moving 2D boundary shapes for AOI annotation. We implemented this approach within the same framework as our approach to provide a comparable interface for the



**Figure 3:** Different views of the timeline visualization showing annotated fixations and different AOI categories.



**Figure 4:** Annotation approaches supported in our framework. (a) Selecting a fixation in the timeline highlights the fixated region in the gaze replay with a green box. (b) Defining AOI area within the gaze replay labels all fixations that lie within this region.

evaluation. Fixation-based annotation is done by assigning AOI labels based on cross-checking with video replay (Figure 6). Here, individual fixations must be selected to display the fixated region within the video replay and then perform the labeling. In contrast, image-based annotation includes a thumbnail image of the fixated region for each fixation. The main difference with image-based annotation is the information contained in the fixation.

**Image-based Annotation** The individual thumbnails in the timeline visualization are labeled with the corresponding AOI. Each thumbnail represents a fixation and contains the corresponding 3D gaze position to provide a spatial context in the gaze replay. For annotation of fixations, one or more thumbnails are first clicked, which are then highlighted in green, and at the same time this focused region is mapped in the gaze replay using the 3D gaze information provided (Figure 4a). Each time a fixation is selected, the camera moves in gaze replay to show the fixation region directly. For the labeling, an existing AOI can be selected via the AOI legend or a new AOI can be defined. There are two possible options for multi-selection of fixations with our image-based and fixation-based annotation method. The first option is to click on each thumbnail. An alternative is to select a thumbnail and successive fixations

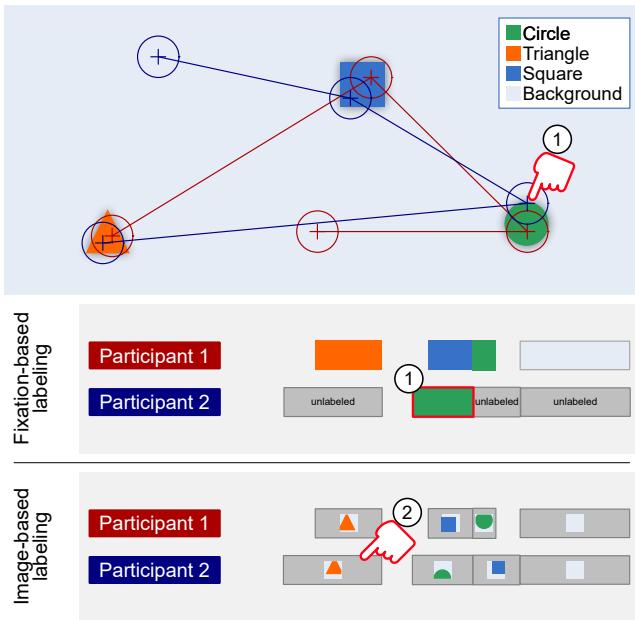
via hotkeys. Additionally, the *remaining fixations* function can be enabled to select all remaining unlabeled fixations.

**Spatial Annotation** The spatial annotation is performed within the gaze replay by placing a cube on the AOI region in the 3D model. This is achieved by clicking on an area in the mesh to align a cube to it. In edit mode, the cube can be moved and resized in all three axes (Figure 5). It can also be deleted as long as it has not yet been saved. Adjusting the rotation of the cube is not supported. Before creating the AOI cube, an existing AOI is selected from the AOI legend, or a new AOI is defined. All fixations in the timeline visualization whose 3D gaze positions are in the selected AOI region are labeled whenever an AOI cube is saved (Figure 4b).

**Filter Options** We can load any number of datasets of participants and decide which of them should be considered and labeled. For this purpose, there is the participant list with all loaded datasets that can be deselected (Figure 1 (c)). Deselection results in the removal of the corresponding fixation sequence in the timeline visualization, thus excluding the corresponding fixations from spatial annotation. Another filter operation is the *annotated fixations* function (Figure 3a). It prevents fixations that are already annotated in both components from being annotated again.



**Figure 5:** An AOI region is defined by placing a cube. It can be moved and scaled using the axes to take the position and size of the real AOI.



**Figure 6:** Different types of annotation. (1) Fixation-based labeling by investigating gaze replays and assigning AOI labels to individual fixations. (2) Image-based labeling where thumbnails directly show AOI images in the timeline and multiple fixations on the same AOI can be labeled simultaneously.

**AOI legend** The AOI legend consists of a virtual and a real part. It lists all predefined AOIs as well as the newly added AOIs. The legend can be used to label the fixations and to create the AOI cubes, but also to determine the percentage of fixations that have been labeled with the corresponding AOIs. Reaching 100% in total corresponds to complete labeling of the fixations.

## 5. Comparative Evaluation

We present a showcase scenario to discuss our approach and compare it with fixation-based labeling as a baseline technique.

### 5.1. Scenario: The Gallery

We decided to showcase our technique with data from multiple people investigating a small art gallery at our institute. The gallery consists of pictures that we augmented with virtual content such as additional images, information text, and videos. This enrichment simulates applications as they have been implemented in the context of a museum or comparable educational environments.

**Stimuli** The investigated pictures comprise artwork of different styles (Figure 7).

1. *Software feathers* show an image of a feather built from software structures. Next to the image, a detailed description of the underlying data was presented virtually.
2. *RoboPix* shows a painting by a robot using the drip technique in the style of Jackson Pollock. The virtual content for the picture showed a video that outlines how such a painting is created.
3. *Hough images* applies Hough transformations to simple shapes in order to create new geometrical shapes. A simple Hough transformation task was given on the virtual board.
4. *Bubble hierarchies* use radial structures to depict hierarchical data structures. The virtual content includes a simple question about the picture.
5. *Frayed cell diagrams* applies Voronoi tessellation to an image of Abraham Lincoln. An original picture of it is shown virtually.

**Task and Participants** Overall, we asked 10 participants to walk through the gallery and solve tasks according to the picture and the virtual information. Two visiting orders for the pictures were predefined and participants were split into two groups. This way, we aimed to create two different patterns for the latter annotation task. On average, it took 7 minutes to complete the tasks.

### 5.2. Annotation Task

To annotate the recorded data, we recruited 8 scientific employees (4 female, 4 male) from our institute. Five of them were in an age group between 26 and 30, while three were between 31 and 40 years of age. All of the participants had expertise in visualization research for at least 1 to 3 years. Four of them had more than 3 years of experience. Four of the participants had additional knowledge of eye tracking data analysis. We do not exclusively address our approach to visualization experts, but for the comparison, we wanted to provide optimal circumstances to achieve the best performance with both techniques. We hypothesize that non-experts are also capable of performing this task with a short learning time and a potentially longer annotation time.

Our study consisted of two annotation tasks using the *Visual Gaze Labeling* approach and the *Fixation Labeling* approach. The *Visual Gaze Labeling* approach provides the participants the possibility to annotate AOIs in 3D space as well as labeling fixations directly based on thumbnail images. The *Fixation Labeling* approach only allows labeling fixations by viewing the corresponding video replay. The viewed artworks (Section 5.1) and their virtual contents were considered as virtual and real AOIs. Here, the fixations that would be labeled with a virtual and real *Hough image* AOI were used for the training phase, so that the fixations for the two



**Figure 7:** Gallery with algorithmically created artwork: (1) Software feathers, (2) RoboPix, (3) Hough images, (4) Bubble hierarchies, (5) Frayed cell diagrams. The enumerated images also provided augmented information in form of text, videos, and interactive questions.

tasks could each be labeled with four virtual and four real AOIs. We predefined the AOIs in the AOI legend so that they could be selected directly. The task was to label all fixations. Fixations that did not belong to any of the artworks were labeled with the real AOI category: *unknown*.

The two tasks were performed sequentially, with the order of the two approaches and the two datasets alternating for each participant. First, the use case study was presented, and all AOIs were briefly shown and described. The first task consisted of a training phase and subsequent labeling of the corresponding dataset. This was followed by a questionnaire to determine the usability of the approach used. The same procedure was repeated for task 2.

In a within-subject design, we separated the dataset into two groups consisting of data from 4 participants in order to keep the time for both annotation tasks within a bearable time frame for the participants. We excluded one part of the data to use it in our training phases for both annotation tasks. The first data group consisted of 999 fixations, of which 305 were unlabeled. The second data group consisted of 766 fixations, of which 300 were unlabeled. We logged annotation times for both tasks and took a screen recording during the study to identify annotation strategies and patterns for the two tasks to compare them between participants.

### 5.3. Results

For the evaluation, we considered qualitative and quantitative aspects. The completion time and labeling performance were inspected, as well as applied strategies to solve the task.

**Performance** On average, both annotation tasks were performed in 43 min in total. Individually, it took 20 min ( $SD = 3.3$  min) to annotate the data with *Visual Gaze Labeling* and 23 min ( $SD = 6$  min) with the other approach. Half of the participants performed better with *Visual Gaze Labeling* (mean = 18 min, mean = 26 min with *Fixation Labeling*). The other half of the participants performed better with *Fixation Labeling* (mean = 18 min, mean = 23 min with *Visual Gaze Labeling*), while one of them performed equally with both techniques.

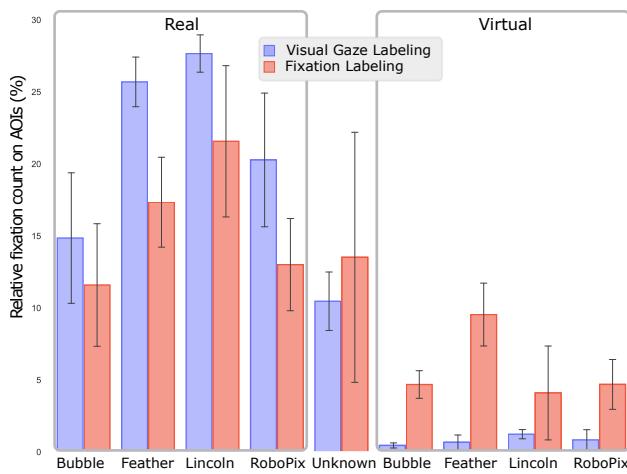
**Fixation Distribution on AOIs** To examine the distribution of gaze on AOIs, we first assigned each fixation to the AOI with the most votes. Then, we examined the relative distribution for both

approaches (Figure 8). The distribution of the data shows that more than 25% of the fixations were labeled with the real *Lincoln image* AOI. Fixations with *Hough image* were removed from the dataset to serve as an example for the training phase. Despite the fact that fixations with virtual AOIs were already labeled automatically, additional fixations in this category were labeled during the task (4.2% for *Visual Gaze Labeling*, 23.5% for *Fixation Labeling*).

**Agreement** We further investigated how consistently participants annotated the data. Deriving a ground truth for the data is hard as point-based eye tracking measures inherently contain uncertainty because not a point but a foveated area is perceived. We used the majority vote to measure agreement. A comparison of the two tasks resulted in an average agreement of 96.06% for *Visual Gaze Labeling* and 88.51% for *Fixation Labeling*. Therefore, we examined the agreement between the individual AOI categories (Figure 9). The distribution of agreement of the AOIs for *Fixation Labeling* is more spread out. The average agreement of virtual AOIs for *Fixation Labeling* is higher compared to the other approach. It should be noted that participants labeled fixations with virtual AOIs much less frequently in *Visual Gaze Labeling* (4.2%) than with *Fixation Labeling* (23.5%). In addition, a lower agreement was obtained for fixations with unknown AOIs. We also measured inter-annotator agreement (IAA) using the Fleiss-Kappa statistic [Fle71] (0.97 for *Visual Gaze Labeling* and 0.92 for *Fixation Labeling*).

**Applied Strategies** There are various approaches to label fixations with our implemented framework. Figure 10 illustrates a workflow for an annotation strategy centered around spatial annotation. Several strategies were used during the study.

The annotation strategy for *Fixation Labeling* did not allow too much variability. The only way was to select fixations in the timeline and then label them. Key shortcuts could be used for more efficient labeling of successive fixations. For example, 7 of the participants enabled the *annotated fixations* feature to skip fixations that were already labeled and used key shortcuts for labeling. Another participant instead enabled the *colored AOIs* feature to decide which AOI the fixations should have based on the neighboring fixations. One participant chose the strategy of selecting multiple images with the same AOI assignment and then annotating them to make fewer annotations. In addition, this participant explicitly did not annotate the unknown fixations in order to select and annotate them at the end using the *remaining fixations* feature.



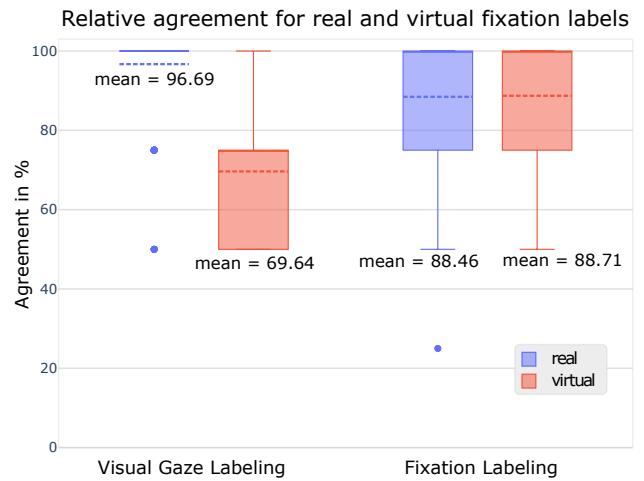
**Figure 8:** Average relative distribution of fixations on real and virtual AOIs after labeling.

Our *Visual Gaze Labeling* approach allowed different strategies to achieve the final annotation. This required first exploring the data and then applying the annotation techniques. Figure 11 shows the workflow of the participants' annotation process, which reveals the annotation strategies used.

**Exploration** The initial strategy of all participants was to apply spatial annotation. To do this, they started with a brief exploration phase to figure out where to place the AOI cube. They examined where all unlabeled fixations were placed. The fixation positions were shown in the gaze replay by enabling the *remaining fixations* option. Dense fixation areas were evident, as shown in the picture (Figure 10 (2)). To locate the different AOIs, they walked around in the 3D view. Another approach was to select some of the unlabeled fixations in the timeline and display their position in the gaze replay. On this basis, spatial annotation was performed. The sequences of **P3** and **P4** (see Figure 11) exhibit this behavior.

**Spatial Annotation** By defining AOI cubes in the reconstructed 3D space, it is possible to label numerous fixations simultaneously by performing hit detections along the gaze ray. **P1** and **P4** mainly used spatial annotation for labeling. After a brief exploration, a bounding box was defined to see if all fixations were detected. If there were still many fixations around the defined AOI cube, a larger box was created. These participants annotated the images according to the hanging order. All participants checked where the remaining fixations were located using the *remaining fixations* option before proceeding with spatial annotation. AOI cubes were also defined for the virtual AOIs if there were multiple unlabeled fixations. Remaining isolated fixations that did not belong to any AOI were finally labeled directly as *unknown*. Three participants recognized that there were several fixations in a region that did not belong to one of the important AOIs. An AOI cube was placed at this place with the *unknown* AOI so that several fixations were labeled.

**Combination with Image-based Annotation** Six participants mainly used spatial annotation. After larger areas were covered by the AOI cubes, approximately the remaining 40 fixations were la-

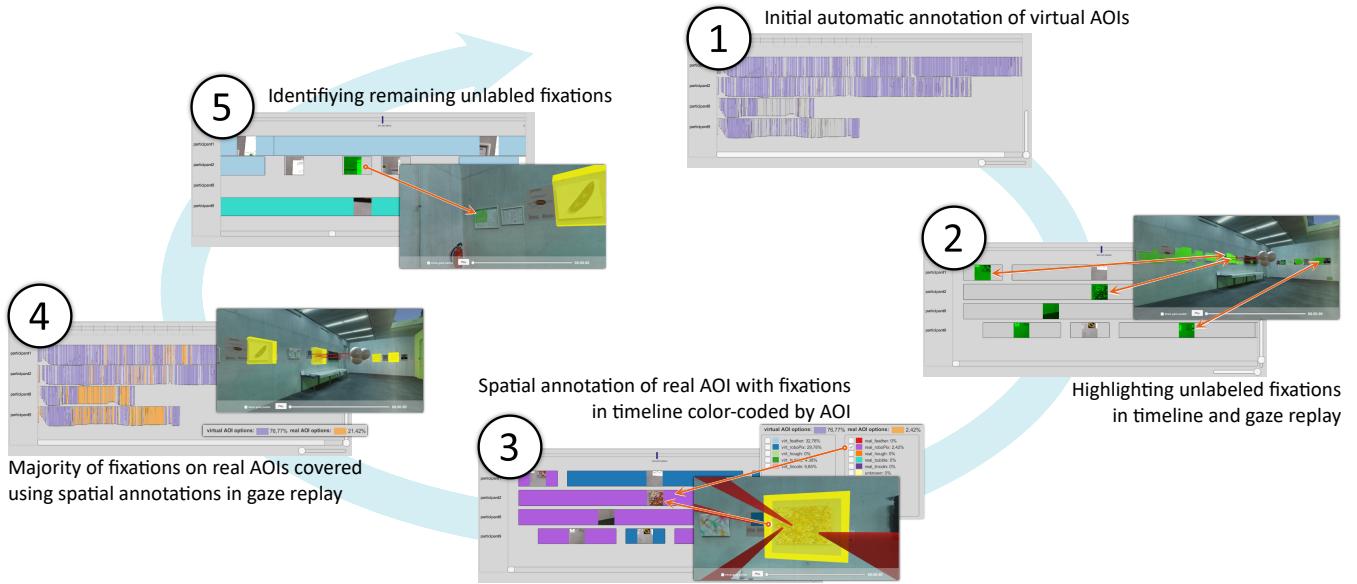


**Figure 9:** Each fixation was assigned a real or virtual AOI. On this basis, the relative agreement was calculated for both approaches.

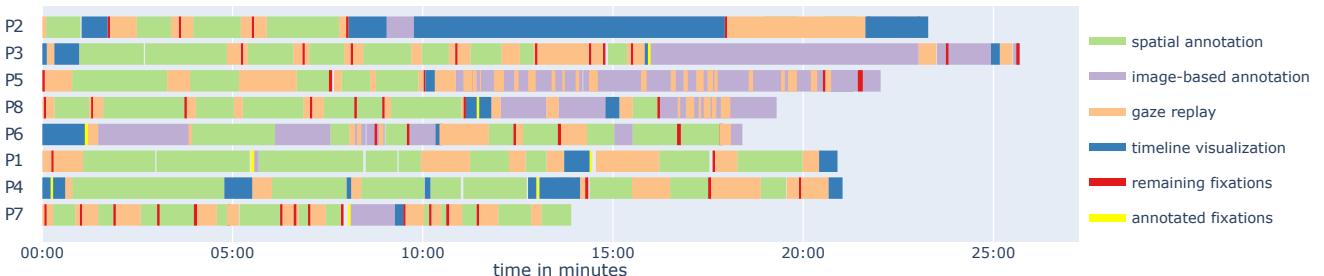
beled by image-based annotation. Alternatively, the remaining fixations could be selected and labeled individually. Two participants used the technique of looking at the remaining fixations first to investigate whether they could all be labeled as unknown. If not, the corresponding fixations were then searched for and labeled, and finally, the unknown fixations were labeled. **P4** regularly switched between the two annotation methods. Some of the participants switched to timeline visualization to detect multiple fixations in the same region and again created an AOI cube.

**Patterns in Annotation Sequences** Figure 11 reveals patterns in the history of sequences of the participants. **P2**, **P3**, **P5** required the longest time for the annotation task. We suspect that this is because all of them spent too much time on the image-based approach. They started with the spatial annotation approach and spent a lot of time on the image-based annotation halfway through. Another pattern we noticed concerns all participants: After creating spatial annotations, they spent time exploring fixations by enabling the *remaining fixations* options or selecting fixations in the timeline visualization. During the study, there was repeated alternating between the two methods, which resulted in a frequent change between 2D and 3D views. Except for **P2** and **P7**, all participants increased their spatial annotation speed over time (indicated by decreasing length of green boxes). This is consistent with usability observations regarding how easy it is to use the technique after a learning period. We assume that with larger datasets, this would have resulted in only minor changes in the sequence history and a slightly higher completion time, depending on the fixation distribution outside the AOI regions. On the other hand, fixation-based annotation would require significantly more time, as the annotation process here is constrained by the manual annotation of all fixations from each dataset.

**Usability** For both tasks, we asked participants how they would use the tool. We asked them to choose the methods they felt were most useful for the *Visual Gaze Labeling* task. **P1**, **P2**, **P3**, **P6** and **P8** preferred the combination of spatial annotation and image-



**Figure 10:** Workflow for a strategy using spatial annotation. (1) In the initial state all virtual AOIs are annotated automatically (shown as purple in timeline). (2) All unlabeled fixations can then be highlighted both in the timeline and gaze replay view to identify regions that require manual annotation. (3) Placing a cube around a real AOI region in the gaze replay annotates all fixations that intersect with that region. In the timeline, fixations can be color-coded by AOI. (4) After covering the most frequently viewed real AOIs using spatial annotation, the number of unlabeled fixations is greatly reduced compared to the initial state (annotated real AOIs now shown in orange in the timeline). (5) Still remaining unlabeled AOIs can be selected in the timeline to highlight their spatial location for further refinement of the spatial annotation.



**Figure 11:** Annotation strategies used by the participants over time. We can observe different patterns in participants' workflows, e.g., **P2**, **P3**, **P5** shift from primarily using spatial annotation to using image-based annotation during the second half.

based annotation. They liked that the context to the real world was provided by a 3D model and that they could see the attention of the AOIs more clearly in this model than in the timeline. **P4**, **P5** and **P7** found only the spatial annotation helpful. **P4** and **P7** also solved the task mainly using this technique. **P4** found it very intuitive and convenient to use. The participant argued that with spatial labeling, human error could be avoided because there was no need to spend effort on individual fixations, as spatial labeling was more automated. We also observed this during the study. When using fixation-based labeling, they sometimes mislabeled fixations without their knowledge. With spatial annotation, this would have been less likely to occur. Additionally, this error would be more likely

to be detected because more data would be affected. The problem could be resolved by adding a new AOI cube.

Positive aspects for *Fixation Labeling* were also mentioned. **P1** and **P6** liked the straightforward and structured linear workflow where fixations could be labeled directly one after the other so that nothing was forgotten to be labeled. In addition, **P1**, who had first performed the *Visual Gaze Labeling* task, found it more convenient that they did not have to switch between the 2D and 3D views, so the workflow was not interrupted, allowing them to achieve good labeling performance. In general, it was also mentioned that usability was easy and efficient for time spans with fixations that could be assigned to the same AOI, as multiple fixations could be quickly se-

lected via hotkeys. Frequent switching between AOIs, on the other hand, required significantly more interaction so that here manual labeling had taken more time and resulted in frustration for the participants. Thus, the efficiency of this approach depended on the labeled dataset.

When asked about the two tasks, positive aspects were mentioned for both approaches. However, when they had to decide which approach they would choose, spatial and image-based annotation was chosen by all of them. They argued that they got a better overview of the scene in the gaze replay, so gaze attention could be explored better. In addition, spatial annotation has more potential for efficient labeling once it is learned because less interaction was needed for labeling here and the performance did not depend on the type of dataset while *Fixation Labeling* suffered from frequent attention shifts. Some participants mentioned that mental effort was significantly lower for spatial labeling. It was also suggested that the defined AOIs in gaze replay could be reused for other datasets in the same environment.

We further asked our participants to fill out standardized questionnaires for each labeling technique, i.e., the NASA task load index (TLX) and the system usability scale (SUS). The SUS resulted in an average score of 77 for *Fixation Labeling* ( $SD = 8.5$ ) and 61 for *Visual Gaze Labeling* ( $SD = 19.5$ ). The high standard deviation results from one participant rating the approach with a score of 25, although this participant had achieved a better annotation time with our method. We interpret the low score with the fact that the user interface was more complex in design with interactions in 3D which not all participants were familiar with.

The NASA TLX scale ranged from 0 (low) to 20 (high). The evaluation showed that the participants perceived a higher performance when solving the task with the *Visual Gaze Labeling* approach (mean = 15.5,  $SD = 4.3$ ) than with the *Fixation Labeling* approach (mean = 13.5,  $SD = 5$ ). Additionally, they perceived less physical demand with *Visual Gaze Labeling* (mean = 3.6,  $SD = 2.6$ ) compared to *Fixation Labeling* (mean = 5.6,  $SD = 3.6$ ). The average scores for the other workload demands were slightly higher for the *Visual Gaze Labeling* approach. The participants perceived the approach as more mentally and temporally demanding and required more effort. This is due to the fact that creating the AOI cubes in 3D was more challenging than manual labeling.

## 6. Discussion

We base our discussion on the results of our evaluation and observations we made during the development and application of the presented approach.

**Scalability** We tested the annotation performance on a rather small subset of our recorded data. This was mainly due to the limited time frame we had for testing per participant. As a consequence, fixation-based labeling and the mixed approach resulted in comparable annotation times. We hypothesize that *Fixation Labeling* scales linearly with an increasing number of fixations to label while spatial annotation in particular will create simultaneous annotations for all fixations within an AOI. Therefore, there exists a threshold where *Fixation Labeling* becomes the less efficient approach

which we will have to determine in future experiments. Regarding the scalability with respect to the number of AOIs, the color-based scatterplot visualization is limited to the number of perceivable colors [HB03]. For the annotation phase, the addition of more labels would be less problematic, but for the overview of the data, alternative visualizations, for instance, based on hierarchical aggregation [BKR<sup>\*</sup>16] might support the managing of many AOIs.

**Dynamic AOIs** Dynamic AOIs (e.g., moving people) are currently covered by fixation and image-based labeling. Spatial annotation is not possible at the moment. Tracking spatially annotated AOIs is already challenging in 2D and would in many cases require additional external tracking devices in AR experiments [KBPR22]. For individual recordings, keyframe-based adjustment of boundary shapes might be an option. However, movement in different recordings still has to be handled individually.

**Spatial Context** Even if the annotation strategies did not result in substantial performance differences, we noticed from the comments of the participants that spatial context played an important role in solving the task, and spatial annotation was preferred in general. We see this as an indicator that AR analysis scenarios should consider including spatial information for context.

## 7. Conclusion

We presented a visualization approach to analyze gaze in AR scenarios. We address the common issue of annotation of AOIs in this context. While we take advantage of labeling gaze on virtual elements automatically, real-world content is included by spatial and image-based annotation. Compared to fixation-based labeling, our new approach provides efficient means to annotate and interpret gaze data from multiple participants simultaneously.

For future improvements, we plan to facilitate the interaction in gaze replay with additional capabilities to improve usability in 3D space. We also want to extend our support for dynamic objects. Although it is possible to identify moving content in the thumbnails of the image-based timeline visualization, spatial annotation of movement trajectories might further improve annotation. However, this probably has to be applied to individual recordings, as objects will in most cases not move identically for each participant. Overall, we see this work as an important step to help understand how people perceive augmented environments and how cognitive processes in this context work to solve specific tasks and perform interactions with novel visual interfaces.

## Acknowledgements

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618 and SFB 1244 – Project ID 279064222. Open Access funding enabled and organized by Projekt DEAL.

## References

- [AABW12] ANDRIENKO G., ANDRIENKO N., BURCH M., WEISKOPF D.: Visual analytics methodology for eye movement studies. *IEEE*

- Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2889–2898. 2
- [BBHD10] BUSCHER G., BIEDERT R., HEINESCH D., DENGEL A.: Eye tracking analysis of preferred reading regions on the screen. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (2010), p. 3307–3312. 2
- [BCNS15] BIANCO S., CIOCCA G., NAPOLETANO P., SCHETTINI R.: An interactive tool for manual, semi-automatic and automatic video annotation. *Computer Vision and Image Understanding* 131, 1 (2015), 88–99. 2
- [BHM\*22] BECHER M., HERR D., MÜLLER C., KURZHALS K., REINA G., WAGNER L., ERTL T., WEISKOPF D.: Situated visual analysis and live monitoring for manufacturing. *IEEE Computer Graphics and Applications* 42, 2 (2022), 33–44. 2
- [BKR\*16] BLASCHECK T., KURZHALS K., RASCHKE M., STROHMAIER S., WEISKOPF D., ERTL T.: AOI hierarchies for visual exploration of fixation sequences. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications* (2016), pp. 111–118. 10
- [BKR\*17] BLASCHECK T., KURZHALS K., RASCHKE M., BURCH M., WEISKOPF D., ERTL T.: Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum* 36, 8 (2017), 260–284. 2, 3
- [CCC\*16] CADENA C., CARLONE L., CARRILLO H., LATIF Y., SCARAMUZZA D., NEIRA J., REID I., LEONARD J. J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* 32, 6 (2016), 1309–1332. 3
- [CKK19] CLAY V., KÖNIG P., KOENIG S.: Eye tracking in virtual reality. *Journal of Eye Movement Research* 12, 1 (2019), 3:1–18. 2
- [DB11] DÜNSER A., BILLINGHURST M.: Evaluating augmented reality systems. In *Handbook of Augmented Reality*, Furht B., (Ed.). Springer, NY, 2011, pp. 289–307. 2, 3
- [Duc17] DUCHOWSKI A. T.: *Eye tracking methodology: Theory and practice*. Springer, Cham, 2017. 2
- [Fle71] FLEISS J. L.: Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (1971), 378–382. 7
- [FLFCBNVM18] FRAGA-LAMAS P., FERNÁNDEZ-CARAMÉS T. M., BLANCO-NOVOA O., VILAR-MONTESINOS M. A.: A review on industrial augmented reality systems for the industry 4.0 shipyard. *IEEE Access* 6, 1 (2018), 13358–13375. 2
- [GSL\*02] GOLDBERG J. H., STIMSON M. J., LEWENSTEIN M., SCOTT N., WICHANSKY A. M.: Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 symposium on Eye tracking research & applications* (2002), pp. 51–58. 3
- [HB03] HARROWER M., BREWER C. A.: ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37. 10
- [HB05] HAYHOE M., BALLARD D.: Eye movements in natural behavior. *Trends in Cognitive Sciences* 9, 4 (2005), 188–194. 2
- [KBM\*21a] KAPP S., BARZ M., MUKHAMETOV S., SONNTAG D., KUHN J.: ARETT: Augmented reality eye tracking toolkit for head mounted displays. *Sensors* 21, 6 (2021), 2234:1–18. 3
- [KBM\*21b] KAPP S., BARZ M., MUKHAMETOV S., SONNTAG D., KUHN J.: ARETT R Package: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays, 2021. Last accessed 13.03.2023. URL: <https://github.com/AR-Eye-Tracking-Toolkit/ARETT-R-Package>. 3
- [KBPR22] KURZHALS K., BECHER M., PATHMANATHAN N., REINA G.: Evaluating situated visualization in AR with eye tracking. In *Proceedings of the Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (2022), pp. 1–8. 2, 10
- [KFBW16] KURZHALS K., FISHER B., BURCH M., WEISKOPF D.: Eye tracking evaluation of visual analytics. *Information Visualization* 15, 4 (2016), 340–358. 2, 3
- [KHH\*15] KURZHALS K., HLAWATSCH M., HEIMERL F., BURCH M., ERTL T., WEISKOPF D.: Gaze stripes: Image-based visualization of eye tracking data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 1005–1014. 2
- [KHSW16] KURZHALS K., HLAWATSCH M., SEEGER C., WEISKOPF D.: Visual analytics for mobile eye tracking. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 301–310. 2
- [KKBW22] KOCH M., KURZHALS K., BURCH M., WEISKOPF D.: Visualization psychology for eye tracking evaluation. In *Proceedings of the Workshop on Visualization Psychology* (2022), pp. 1–18. 2
- [Kur21] KURZHALS K.: Image-based projection labeling for mobile eye tracking. In *Symposium on Eye Tracking Research and Applications* (2021), pp. 1–12. 3
- [Liv05] LIVINGSTON M. A.: Evaluating human factors in augmented reality systems. *IEEE Computer Graphics and Applications* 25, 6 (2005), 6–9. 2
- [LSO20] LI T.-H., SUZUKI H., OHTAKE Y.: Visualization of user's attention on objects in 3D environment using only eye tracking glasses. *Journal of Computational Design and Engineering* 7, 2 (2020), 228–237. 2
- [Mic16] MICROSOFT: MixedRealityToolkit-Unity, 2016. Last accessed 17.03.2023. URL: <https://github.com/microsoft/MixedRealityToolkit-Unity>. 3
- [Mic22] MICROSOFT: MixedRealityToolkit-Unity, 2022. Last accessed 17.03.2023. URL: <https://learn.microsoft.com/en-us/mixed-reality/world-locking-tools/>. 3
- [MPPO19] MEISSNER M., PFEIFFER J., PFEIFFER T., OPPEWAL H.: Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *Journal of Business Research* 100, 1 (2019), 445–458. 2
- [MT21] MUCHEN Y., TAMKE M.: Augmented reality for experience-centered spatial design: A quantitative assessment method for architectural space. In *Towards a new, configurable architecture: Proceedings of the eCAAdE Conference-Volume 1* (2021), pp. 173–180. 2
- [NBW16] NETZEL R., BURCH M., WEISKOPF D.: Interactive scanpath-oriented annotation of fixations. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications* (2016), pp. 183–187. 3
- [OPB\*23] ÖNEY S., PATHMANATHAN N., BECHER M., SEDLMAIR M., WEISKOPF D., KURZHALS K.: Visual Gaze Labeling for Augmented Reality Studies, 2023. doi:10.18419/darus-3384. 2
- [PB06] POOLE A., BALL L. J.: Eye tracking in hci and usability research. In *Encyclopedia of Human Computer Interaction*, Ghaoui C., (Ed.). IGI Global, Hershey, PA, 2006, pp. 211–219. 2
- [PHG\*04] PAN B., HEMBROOK H. A., GAY G. K., GRANKA L. A., FEUSNER M. K., NEWMAN J. K.: The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research & applications* (2004), pp. 147–154. 3
- [PKP10] PONTILLO D. F., KINSMAN T. B., PELZ J. B.: Semanticode: Using content similarity and database-driven matching to code wearable eyetracker gaze data. In *Proceedings of the ACM Symposium on Eye-Tracking Research and Applications* (2010), pp. 267–270. 3
- [SG00] SALVUCCI D. D., GOLDBERG J. H.: Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research & Applications* (2000), pp. 71–78. 3
- [SG22] SUNDSTEDT V., GARRO V.: A systematic review of visualization techniques and analysis tools for eye-tracking in 3d environments. *Frontiers in Neuroergonomics* 3, 1 (2022), 910019:1–15. 3
- [SND10] STELLMACH S., NACKE L., DACHSELT R.: Advanced gaze visualizations for three-dimensional virtual environments. In *Proceedings of the Symposium on Eye-Tracking Research & Applications* (2010), pp. 109–112. 4

[VRZ\*17] VÁVRA P., ROMAN J., ZONČA P., IHNÁT P., NĚMEC M., KUMAR J., HABIB N., EL-GENDI A.: Recent development of augmented reality in surgery: A review. *Journal of Healthcare Engineering* 2017, 1 (2017), 4574172:1–10. [2](#)

[WHB\*18] WOLF J., HESS S., BACHMANN D., LOHMEYER Q., MEBOLDT M.: Automating areas of interest analysis in mobile eye tracking experiments based on machine learning. *Journal of Eye Movement Research* 11, 6 (2018), 6:1–11. [2](#)