A Spiral into the Mind: Gaze Spiral Visualization for Mobile Eye Tracking

MAURICE KOCH*, University of Stuttgart, Germany
DANIEL WEISKOPF, University of Stuttgart, Germany
KUNO KURZHALS, University of Stuttgart, Germany

Fig. 1. Video and gaze data of a person looking at different magazines (0:40 min). For each frame, a vertical scanline is extracted to create the visualization. The location and size of the scanline is either constant (orange) or influenced by the current point of regard (blue, green). The video starts in the center of the spiral and time continues clockwise away from the center. The spiral pattern allows one to represent long time spans and supports a good overview, even on multiple recordings. The visualization can be created directly for the egocentric video and with gaze data in the global and local context of the video. Black scanlines are inserted for frames with missing gaze data.

Comparing mobile eye tracking data from multiple participants without information about areas of interest (AOIs) is challenging because of individual timing and coordinate systems. We present a technique, the gaze spiral, that visualizes individual recordings based on image content of the stimulus. The spiral layout of the slitscan visualization is used to create a compact representation of scanpaths. The visualization provides an overview of multiple recordings even for long time spans and helps identify and annotate recurring patterns within recordings. The gaze spirals can also serve as glyphs that can be projected to 2D space based on established scanpath metrics in order to interpret the metrics and identify groups of similar viewing behavior. We present examples based on two egocentric datasets to demonstrate the effectiveness of our approach for annotation and comparison tasks. Our examples show that the technique has the potential to let users compare even long-term recordings of pervasive scenarios without manual annotation.

Authors’ addresses: Maurice Koch, maurice.koch@visus.uni-stuttgart.de, University of Stuttgart, Stuttgart, Germany; Daniel Weiskopf, University of Stuttgart, Stuttgart, Germany, daniel.weiskopf@visus.uni-stuttgart.de; Kuno Kurzhals, kuno.kurzhals@visus.uni-stuttgart.de, University of Stuttgart, Stuttgart, Germany.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2577-6193/2022/6-ART20 $15.00
https://doi.org/10.1145/3530795

Proc. ACM Comput. Graph. Interact. Tech., Vol. 5, No. 2, Article 20. Publication date: June 2022.
INTRODUCTION

To capture gaze data in realistic, everyday situations, mobile eye tracking is deployed in many areas, such as behavioral research [Land and Hayhoe 2001], pervasive computing [Bulling et al. 2011], and augmented reality [Naspetti et al. 2016]. It combines gaze data and egocentric video, two data sources already difficult to analyze on their own. Each recorded video consists of individual coordinate systems, movement, and timing. Hence, an analysis beyond descriptive statistics of gaze characteristics (e.g., average fixation duration) requires contextual information from all recordings. Especially for the analysis of scanpath sequences, relevant areas and objects have to be identified in each video, so this point often manually, rendering the definition of AOIs as one of the most time-consuming steps in the analysis process.

Interactive data visualization provides means to investigate recorded videos and their respective scanpaths under numerous aspects [Blascheck et al. 2017]. Most common is the depiction of measures over time (e.g., pupil dilation, coordinates) that provide more details than aggregated descriptive statistics. In the stimulus, gaze replays overlaid on the egocentric video can be investigated with most software packages shipped with the device. However, a comparison between multiple participants is difficult to achieve with gaze replays alone. Hence, most techniques to visualize mobile eye tracking data in a static overview require the definition of AOIs as a preprocessing step. With the information in which order people looked at specific objects or areas, sequential representations (e.g., scarf plots [Richardson and Dale 2005]) help interpret data and compare between multiple participants [Kurzhals et al. 2014]. Overview and comparative visualizations that work without AOIs are sparse. Although progress in machine learning techniques allows one to identify specific AOIs automatically [Wolf et al. 2018], this typically requires pretrained categories for specific AOIs. An application to arbitrary eye tracking experiments is still limited to rather general object categories. For example, it is possible to detect when a person looked at different cars, but to identify when it was their own car requires external knowledge that has to be fed into the algorithm. Visualization can help communicate such human interpretations to the computer, making the data more accessible for analysts, and it helps interpret results from automatic processing steps.

We introduce a visualization approach for the overview, annotation, and comparison of numerous recordings from mobile eye tracking. This is achieved by creating a compact, static representation of video content incorporating gaze positions: the gaze spiral. In this way, visual patterns such as dwells on an area, as well as changes between different areas of the stimulus are preserved. For example, Figure 1 shows three types of spirals created under different conditions. The gaze spirals convey frequently occurring patterns of gaze behavior that can be visually identified. The compact representation of the gaze spiral allows to visualize even hour-long recordings and highlight results from queries without the need for horizontal or vertical scrolling. Furthermore, multiple recordings can be displayed simultaneously for the interpretation of automatic comparison methods. Overall, we contribute a new technique to (1) visualize scanpaths from mobile eye tracking in a compact representation, (2) investigate visual behavior patterns by similarity queries and annotation, (3) and facilitate projection-based comparison of recordings. Our approach does not require AOIs and
can be applied directly to recorded gaze data. Hence, we see this as a starting point for the analysis of data from mobile eye tracking experiments, to assess the quality of recordings, identify relevant scanpath patterns, and derive AOIs for established analysis techniques.

2 RELATED WORK

Relevant work with respect to this technique comprises comparable approaches that focus on the analysis of data from mobile eye trackers and more general visualization techniques for temporal data, especially video data.

2.1 Mobile Eye Tracking Analysis

Eye tracking data from mobile glasses can be analyzed without contextual information from an egocentric camera, for example, the detection of vigilance [Ji and Yang 2002] or cognitive load [Rafiqi et al. 2015]. We focus on techniques that include the video stimulus because they generally comprise more scenarios to investigate behavior such as spatial orientation [Kiefer et al. 2017] and activities in everyday situations [Land and Hayhoe 2001].

As mentioned, with AOIs available, established means of statistical analysis can be applied and used for visualization. Blascheck et al. [2017] provide an overview of visualization methods based on AOIs. Techniques that were applied to mobile data are hierarchical techniques such as transition matrices and graph visualization [Blascheck et al. 2016], as well as timeline-based techniques such as scarf plots [Richardson and Dale 2005].

Since the definition of AOIs is a complex problem by itself, some techniques have been proposed that visualize image content of the video directly while incorporating information about gaze positions. Techniques such as fixation-image charts [Kurzhals and Weiskopf 2016] visualized scanpaths from videos by thumbnail image sequences, extracted at the point of regard. Scanpath comparison techniques were applied based on image features [Castner et al. 2020; Kurzhals et al. 2015], providing results comparable to AOI-based methods. This image-based approach was also proposed for the annotation of AOIs, supported by automatic processing of similar images [Kurzhals et al. 2016a]. However, these methods did not focus on representing the context of the stimulus in its temporal order. Instead, by annotating AOIs in the visualization, the sequential scanpath was partially displayed by color-coded scarf plots. In contrast, we provide a visualization that supports comparisons between recordings without the need for annotation.

Multi-layered timelines with spectrograms for different feature intensities were proposed for long-term eye tracking recordings by Kurzhals et al. [2020]. By querying selected time spans, similar sequences could be searched and were depicted by short thumbnail sequences of the recorded field of view with a gaze overlay. Although the authors investigated hour-long videos, they focused on the analysis of single videos. A comparison between recordings was not possible. With our approach, we aim to fulfill both requirements, by providing a compact representation of long video sequences and methods to compare between participants.

2.2 Visualization of Temporal Data

The visualization of video data can be separated into applications for analytical purposes and video-based graphics [Borgo et al. 2012]. We see our approach mainly on the analytical side, however, it is inspired by techniques for video artwork based on the slitscan technique, for instance, the Last Clock application1, which presented the images resulting from video processing with a clock metaphor on iPads.

1https://gnomalab.es/last-clock-ipad/. Last checked on April 14, 2022.
Gaze spirals serve as compact representation of visual context and gaze properties of a scanpath. When investigating multiple recordings, they can be applied as glyphs in a 2D embedding, based on scanpath comparison metrics. The presented approach aims to support four common tasks for eye tracking analysis: the visual assessment of data quality, identification of patterns, annotation of AOIs and behavior, and the comparison of scanpaths from different recordings.

In the context of video analysis, this technique has been applied to create visual fingerprints that summarize long videos [Bezerra and Lima 2006; Seitz and Kim 2003]. Kurzhals and Weiskopf [2016] presented the idea to include gaze data to create slitscans for videos shown with a static remote eye tracking system. However, they did not apply the technique to mobile eye tracking data. Kurzhals [2021] used slitscans as a visual aid to determine thresholds for the segmentation of video data. The visualization was represented as horizontal timelines. This layout impairs the comparison for long recordings because horizontal scrolling for individual timelines becomes necessary once the number of timesteps exceeds the horizontal screen resolution. Burch [2018] suggested applying dimensionality reduction on subsequences of scanpaths from static images, but the visualization of the results was restricted to scatterplots, impairing a direct interpretation of the results. Weber et al. [2001] argued that spiral visualization of time-series data scales well with time and helps identify periodic and recurring patterns. A linear representation either needs line wraps, excessive zooming, or scroll interactions to be displayed on the screen. Compared to that, our approach produces a continuous scanpath representation without line wraps. Zooming still might be required to fit the entire sequence on the screen, but to lesser degree than linear representations.

3 TECHNIQUE

We designed a visualization technique for data from mobile eye trackers in unrestricted scenarios. Based on the literature [Kurzhals et al. 2017] and our own experience with eye tracking experiments, we identified four tasks (Figure 2) we address in this context:

Data Quality Assessment. Data quality has numerous aspects [Holmqvist et al. 2012]; we help assess quality in terms of data loss because it is a prerequisite for most analyses. Since eye tracking is prone to issues with the detection of gaze samples, data will be missing for a multitude of reasons. Important is that experimenters are aware of missing data and can decide whether a recording is sufficient to be included in the analysis. This is often determined by defining a threshold for the acceptable percentage of valid samples. Investigating the temporal distribution of missing data is helpful [Schulz et al. 2017], as short periods of missing samples might be less problematic than long periods without gaze detection during important events.
Fig. 3. Comparison of linear and spiral gaze-guided slitscans for three recordings. (1) At the center of the spiral, the compression of visual content is the strongest, which decreases with later timesteps. (2) The linear representation is displayed with a reading direction from left to right, a clockwise spiral inverts the respective patterns horizontally. (3) If the rotation of the scanlines is continuous, the patterns in the upper part of the spiral are vertically inverted. In general, the spiral uses display space more efficiently than the linear slitscan.

Pattern Identification. Mobile eye tracking data consists of multiple data sources, i.e., recorded eye movements and a stimulus video that provides context. In this work, we assume that patterns are manually identified by the user and that the visualization aims to support this process. Patterns can be identified in both, e.g., horizontal scanpath patterns during reading [Kunze et al. 2013], or a specific sequence of how AOIs are investigated. Our main focus is on experiments in unrestricted environments where the context of the stimulus is necessary to identify behavioral patterns and visual strategies people apply to solve a task.

Annotation. AOIs are the prerequisite for the majority of statistical and visual analysis methods in this mobile context [Blascheck et al. 2017]. If gaze samples or fixations can be assigned to a specific AOI, distributions can be calculated to identify frequency and duration of gaze on each AOI. Similarly, behavior related to eye movement (e.g., visual search) often needs to be identified and annotated. In many scenarios, especially in in-the-wild studies, a fully automated solution is not possible and manual annotation is necessary.

Scanpath Comparison. Scanpath comparison [Anderson et al. 2015] is applied for the identification of common visual strategies (e.g., how people read metro maps [Netzel et al. 2017]) or the grouping of levels of expertise (e.g., for chess players [Reingold and Sheridan 2011]). For mobile eye tracking data, this is typically achieved with string-comparison methods of annotated data. These methods provide pairwise similarity values that can be clustered for a larger number of recordings, resulting in groups of similar scanpaths.

To address all these tasks together, existing techniques are not sufficient, either because they require annotated data (e.g., scarf plots, comparison metrics), do not preserve the sequential order of a scanpath (e.g., gaze thumbnails [Kurzhals et al. 2016b]), or are just not well suited for long sequences. Hence, we designed a compact representation of slitscans that conveys important scanpath information, particularly, what visual context was investigated over time by combining information about the point of regard and the stimulus. This single representation of a scanpath—the gaze spiral (Section 3.1)—can serve as a compact glyph that conveys important scanpath aspects and is further applied to compare numerous recordings. We discuss the parameter space for gaze spirals and how individual parameters influence the rendered results (Section 3.2). Important patterns indicating gaze on AOIs can be queried and annotated directly in the visualization (Section 3.3). For comparison, we create a 2D embedding of all scanpaths based on established comparison metrics (Section 3.4). The result shows gaze spirals in a spatial layout where similar scanpaths are in spatial
proximity and visual patterns, mainly referring to gaze on AOIs, can be investigated, queried, and annotated in the visualization.

### 3.1 Gaze Spiral

A slitscan is generated by placing a scanline in a video, extract pixels along the scanline over time, and render the scanlines into a static overview. The technique is not limited in terms of location and orientation of the scanline [Tang et al. 2008]. For gaze-adaptive slitscans [Koch et al. 2018; Kurzhals and Weiskopf 2016], the location of the scanline is adjusted, according to the current point of regard. For the representation, horizontal timelines have the advantage that they represent linear time in the most common way [Aigner et al. 2011]. Considering the scalability of temporal extent, individual timesteps are bound to the screen resolution, which means that for long sequences, the visualization either has to be scrolled, making it hard to compare numerous timelines, or zoomed out, which results in a loss of visual information. Hence, we suggest applying a spiral pattern to represent slitscans (Figure 3). Different patterns, such as a space-filling curve [Dafner et al. 2000], would show the data with a comparable density.

We decided for the spiral because it represents time with a clock metaphor, which is familiar to most people. Spiral data visualization can further help detect recurring events, which is often the case in eye tracking data, e.g., when people are repeatedly switching attention between different areas of interest. Further, the radial representation is often perceived as aesthetic, which renders the visualization as a good method for the illustration and communication of the data. We argue that the slitscan representation comes with a good trade-off between compactness and interpretability. In principle, other types of representations could be wrapped into a spiral form, but they have their own set of problems. For example, gaze stripes [Kurzhals et al. 2015] are based on image thumbnails, thus providing more context to the underlying stimulus than scanlines. However, this property makes gaze stripes also impractical for longer recordings. In contrast, scarf plots are compact but assume the existence of AOI definitions, which are difficult to provide in free-viewing contexts.

For data visualization, the Archimedes’ spiral is the most commonly applied spiral form [Weber et al. 2001]. It reads in polar coordinates:

\[
\phi(t) = t^k \\
\tau(t) = a \cdot \phi(t)/2\pi
\]

Here, time \( t \) is a unitless quantity and increases with a fixed step size. It can be derived from physical time (recording time or frame number) by applying appropriate scale and bias operations. We assume that the data starts at \( t = 0 \).

The Cartesian coordinates \( x(t) = r(t) \cdot \cos(\phi(t)) \) and \( y(t) = r(t) \cdot \sin(\phi(t)) \) define the position of the spiral baseline. Finally, the gaze spiral is generated by placing the scanlines orthogonal onto the spiral’s baseline, pointing outward from the perspective of the baseline. The parameter \( a \) is a constant for the distance between spiral arms and the rotation angle between timesteps can be influenced by the exponent \( k \).

### 3.2 Parameter Space

For the design of gaze spirals, a set of parameters can be adjusted to influence the outcome of the rendering process.

**Level of Detail.** First, the data can be represented with different levels of detail, mainly by focusing either on the egocentric video, or the underlying gaze data (Figure 1). Focusing on the egocentric video, a vertical scanline in the center of the video can be applied. Such data is often used as an approximation for visual attention, given the assumption that people mainly focus on the center of
their field of view. However, as Figure 1 shows, this is not always the case and including eye tracking data provides further details about the viewing behavior. With global gaze-adaptive slitscans, the $x$-position of the point of regard is applied to position the scanline. The $y$-position can be indicated with a visual marker, in the example rendered as white dots with a red border. Due to the rapid nature of eye movements, the smoothness of patterns is interrupted, similar to slitscans resulting from edited video material. Local gaze-adaptive slitscans focus on a small area around the point of regard. While this provides the most details on a fixation level, the resulting slitscan is visually the least smooth, as it is disrupted by horizontal and vertical eye movements.

**Distance.** By increasing $a$, additional blank space between the arms of the spiral can be created (Figure 4a). While this impairs the compactness of the spiral, it also provides new space to render additional information into the visualization. For instance, annotations or related eye tracking metrics could be integrated by color codings [Koch et al. 2018]. We propose setting $a = 1.2$ (where the slitscan height is assumed to be 1) and placing a fixation color encoding into the empty space. For that, prior to rendering the gaze spiral, we extract fixations from the gaze and video data using the method described by Steil et al. [2018]. For the purpose of applying the spirals as small multiples, we suggest favoring the compact representation without labels or color coding.

**Angle.** The angular sampling of the visualization determines how dense consecutive scanlines will be placed in the visualization. The angle also influences the total size of the spiral. We control the angle mapping via the parameter $k$. The pattern at $k = 1$ corresponds to equidistant samples in time (Figure 4b). As one can see, the distance between sample points on the spiral increases quickly. In general, setting $k < 1$ compresses the spiral, whereas setting $k > 1$ spreads the spiral.
We typically use $k = 1$ (for equidistant sampling) or values slightly less than one for moderate reduction of the angular speed toward the outward parts of the spiral (e.g., $k = 0.9$).

**Scanline Orientation.** For the described technique, we applied a vertical scanline. However, as Tang et al. [2008] have shown, it is possible to take arbitrary orientations and lengths for the scanlines to focus on specific motion and locations. We focus on the current point of regard, which determines the location of the scanline. Considering the orientation, we applied vertical scanlines for consistency between the different renderings. However, with respect to gaze behavior, it might be beneficial to orient the scanline according to the angle between consecutive fixations.

**Sampling.** The sampling of the data can be reduced to decrease the size of the resulting spirals. With respect to eye tracking, the sampling of this data source is adjusted to the sampling rate of the video. Hence, a video with 25 fps results in 40 ms duration per frame. With an average fixation duration of 200–300 ms [Rayner 1998] for information processing, we do not recommend reducing the sampling of the data below one fourth of the original samples, since it could result in missing static patterns typical for fixations. Figure 4c displays an example of how the visualization changes when the samples are reduced uniformly. Future work could consider adaptive sampling techniques to represent only important data and to reduce redundant information.

### 3.3 Query and Annotation

So far, the gaze spiral is static but user interaction is necessary for performing annotation tasks. The annotation system is based on four aspects: inspecting fixations; selection of fixations; annotation of selections; and accepting recommendations. Thumbnail images associated to each fixation in the user selection support the visual inspection of fixations. This selection is adjustable, thus the user can skim through the whole sequence fixation-by-fixation. This selection can also be expanded or shortened upon user request. The actual annotation is, when initiated by the user, always performed on the current selection of fixations. Annotation also automatically triggers a query mechanism to find candidate fixations that are similar in terms of image content. These candidate fixations are placed as thumbnails along the spiral. The user now can continue the annotation process from those candidate fixations by a mouse click on the corresponding thumbnail image, which anchors the user selection on that specific fixation.

### 3.4 2D Embedding

After discussing how a single spiral is created, this section considers how to compare different spirals and respective recordings. Scanpaths can be compared with string-based sequence alignment algorithms, like the Levenshtein distance [Levenshtein et al. 1966] or the Smith-Waterman algorithm [Smith and Waterman 1981]. String-based representations are usually obtained from scanpaths by mapping each fixation to a symbol that corresponds to a specific region in the stimulus. Instead of relying on AOIs, we use a feature-based representation of fixations, which has been proposed before by Kurzhals et al. [2015] and Castner et al. [2020] to perform sequence alignment between scanpaths. For the Levenshtein and Smith-Waterman algorithms, the cost is defined by the cosine distance. The gap cost for both algorithms needs to be defined prior to alignment. In our following discussions, we define the gap cost for both algorithms to be 0.5 for feature-based sequences and to be one for string-based sequences. Additionally, we use Dynamic Time Warping [Berndt and Clifford 1994] to perform alignment without the need of specifying gap costs.

Hereafter, we present a visual approach to support the interpretation of metric outputs. First, the distances between all pairs of sequences is computed with the target metric. The resulting matrix of pairwise distances is used to calculate a 2D embedding of scanpaths using dimensionality reduction. Notice that the previously described preprocessing is identical to that performed before clustering.
Fig. 5. 2D embeddings of UMAP projection algorithm with three different metrics applied to 15 scanpaths. Each scanpath is represented by a gaze spiral, and its screen position is defined by the 2D embedding.

Each scanpath is visually represented by its corresponding gaze spiral, and its center position on the screen is defined by the 2D embedding. Figure 5 shows an embedding of 15 scanpaths (see Section 4.1). The 2D embedding is obtained with Uniform Manifold Approximation and Projection (UMAP) [McInnes et al. 2018], which is a state-of-the-art technique for dimensionality reduction. As shown in the Figure 5, depending on the applied metric, the embedding results in different local clusters. Consequently, it is necessary to investigate individual clusters and interpret why they were rated as similar by the respective algorithm. This task can be performed directly in the visualization with the gaze spirals. An example is discussed in Section 4.

4 COMPARISON WITH ESTABLISHED METHODS

The purpose of the following analysis is to evaluate how a 2D embedding of gaze spirals can be interpreted and how the approach compares to established methods that rely on annotated data, i.e., Hierarchical Cluster Analysis (HCA) on string-based AOI sequences.

4.1 Dataset and Annotation Scheme

To showcase our approach, we recorded a dataset that emulates a scenario for the application of mobile eye tracking, the visit of an art gallery. The scenario is mainly defined by a set of AOIs, i.e., different paintings and the description text next to each piece. Visitors walk through the gallery, look at the paintings, read descriptions, and switch their visual attention in different order and duration between AOIs. Hence, it could help investigate how people view the paintings, for example, to rearrange some parts of the gallery. Further, there might be some groups in the data with different viewing behavior, for example, children or elder people. Note that the data was recorded

Fig. 6. The art collection scene consists of six different paintings of algorithmic artwork. Each participant was asked to attend the paintings in two specific orders as indicated by the blue and yellow arrows.
(a) Levenshtein feature-based alignment. (b) Levenshtein AOI-based alignment.

Fig. 7. Comparison between 2D embedding and Hierarchical Cluster Analysis (HCA) based on two different types of sequence alignments (a) and (b). Each alignment between the 14 sequences results in a pairwise distance matrix from which the clustering and 2D embedding are generated. Both outputs show that clustering and 2D embedding lead to comparable results.

for illustration purposes and does not claim any scientific significance about the interpretation of the presented stimuli. In the following, we will show how our approach could be applied to answer such questions and compare how this would be achieved with established methods.

We asked 8 participants to walk through an art gallery at our institute, investigate the artwork, and read the text. This task was performed twice, once in order from left to right (R) and once from right to left (L), simulating a first and a second impression of the work. The starting order was altered between participants. Data was recorded with Pupil Invisible glasses. An offset correction was performed at reading distance to text descriptions. This resulted in 16 scanpaths, one was removed due to recording issues. The scanpaths contain two induced overall patterns (L and R), and individual differences between participants looking at the AOIs.

The gallery consists of six paintings of algorithmic art (Figure 6), each defined as an individual AOI. Additionally, we define one AOI that encompasses the description texts of all paintings and one for all remaining areas. Fixations on AOIs were annotated with our approach and serve as the baseline we compare against.

4.2 Comparison between Hierarchical Cluster Analysis and 2D Embeddings

First, we investigate the pairwise distances between all sequences in the dataset under consideration of the different alignment methods (see Section 3.4). Taking into account the correlation between AOI-based (A) and feature-based sequence (F) alignment, we calculated the Pearson correlation between distance matrices. We get the correlations $\kappa(\text{Levenshtein-A, Levenshtein-F}) = 0.939$, $\kappa(\text{Smith-A, Smith-F}) = 0.901$, and $\kappa(\text{DTW-A, DTW-F}) = 0.904$. The highest correlation between feature and AOI alignment can be observed with the Levenshtein distance. This shows that, independent of the applied alignment method, the calculated similarities based on image features, which can be performed directly on the recorded data, highly correlate with annotated results.
Second, we want to compare how clusters derived from annotated data compare to cluster results in the 2D embedding. Figure 7 illustrates the visual comparison of these results. There is strong correspondence between the HCA result represented by the dendrogram and the 2D embedding, in particular on the lowest dendrogram level (colored groupings in red, yellow, purple, green). The induced pattern between different viewing directions (L and R) is easy to identify at the highest level of the cluster hierarchies. Going up the dendrogram hierarchy, we observe some larger deviations to the 2D embedding. For example, in (b) the red and yellow groups belong to the same high-level cluster according to the dendrogram, but are rather far apart in the 2D embedding. It should be noted that preserving distances in the 2D embedding cannot be guaranteed in UMAP at any granularity and depends strongly on the number of neighbors factor. In general, a higher neighbors factor better preserves the global structure of the distances at the cost of fewer local clusters. A neighbors factor = 3 and higher results in a clear separation between L01–L07 and R01–R07, but misses the local clusters that can be observed when neighbors factor = 2. Next, we want to investigate how clusters differ.

4.3 Detailed Cluster Analysis

Quantitative analysis of scanpaths based on Hierarchical Cluster Analysis fails to provide sufficient details on why a particular grouping emerged. The 2D embedding of scanpaths, visually represented with gaze spirals, aids such interpretation on a qualitative level. Figure 8 provides a more detailed view into the 2D embedding obtained from Levenshtein distance. If we focus on individual clusters between scanpaths, we can investigate, for example, why the sequences L04, L06 and L05, L07 emerged as distinct groups in the embedding. L05 and L07 show a distinctive pattern that is less prominent in the scanpaths L04 and L06 (highlighted in blue). If AOs are available, we can visualize the scanpath with scarf plots. The scarf plots confirm that the dark blue AOI is more prominent in the upper cluster. Hence, scanpaths L05 and L07 attended the AOI Painting with Flow longer than...
Fig. 9. Example of gaze spirals for five sequences of the Extended GTEA Gaze+ dataset. In all sequences, the participants were asked to prepare a hamburger dish. Left: Annotated spiral with occurring events during the recording. Right: Results of an example query to find out when people opened the fridge. Time spans similar to the yellow framed slitscan pattern in the left spiral are visually highlighted in the other four spirals.

L04 and L06. Similar investigations could be performed to describe the characteristics of individual scanpath groups, for example, to classify behavior.

5 EXAMPLES

The presented examples (Figures 1, 3, and 4) comprise three videos of a dataset presented by Kurzhals et al. [2016a]. All videos show a person looking at four magazines, picking one up, skimming through it, and putting it back on the table. All videos are below one minute recording time. For Figures 4a and 10, we used a video recorded of a car ride [Kurzhals 2021].

To further demonstrate our approach, we investigated recordings from the Extended GTEA Gaze+ dataset for activity recognition in egocentric video [Li et al. 2018]. This dataset contains 28 hours of content from 86 unique sessions of 32 participants performing several cooking activities. Figure 9 shows the gaze spirals of five different participants preparing a hamburger dish. The duration of the shown sequences is between 8 and 19 minutes. We used $k = 0.8$ for rendering and retained all samples to generate the spirals with high level of detail. The left side of Figure 9 depicts events during the preparation of a hamburger dish, e.g., cutting a tomato, looking at an information screen, or frying the burger. Reoccurring events such as opening the refrigerator can be identified in the spiral as blueish patterns (highlighted by yellow borders 1–4). Considering sequences from multiple subjects, one could be interested in finding similar events to compare how participants organized the preparation of the ingredients. Based on the previously described query mechanism (Section 3.3), we selected the slitscan pattern in the left spiral of Figure 9 (yellow framed segment) to query events in sequences from four different participants. The right side of Figure 9 shows the result of this query. Investigating the thumbnail images confirms that there is indeed a close similarity of all found time spans to the event of interest. Most importantly, we can identify the results in the overview visualization immediately and the slitscan further helps interpret if a highlighted segment is visually similar to the selection. This query mechanism facilitates detecting
Fig. 10. A linear slitscan and gaze spiral showing a video of a car ride (29:08 min). To display the 367,122 timesteps, we had to rescale the linear slitscan to 5% of its original width. Even with this drastic reduction, only a short time span can be investigated when scaled to a height comparable to the gaze spiral.

similar regions automatically, which is important for long sequences where a manual identification of similar time spans becomes more difficult. Identified time spans can be investigated, confirmed, and annotated with the respective label.

6 DISCUSSION

Our examples showed that it is possible to achieve similar interpretations with the image-based 2D embedding of gaze spirals, as with clustering based on annotated data. This is an advantage for early analysis stages where no AOIs are available, because it helps identify groups of similar viewing behavior, which could be missed by statistical methods on the data as a whole. Furthermore, the reasons why scanpaths were grouped can be investigated directly with the gaze spirals, reducing the frequency of time-consuming investigations in the recorded videos. However, we identified some points that have to be considered for the application of gaze spirals in the future.

**Scalability:** Considering the temporal scalability of our approach, we tested gaze spirals on data with up to one hour of recording time. While this comprises a large number of possible experiments, it does not cover pervasive recordings over multiple hours. In such cases, we recommend reducing the sampling rate, as a trade-off between overview and representation of details. Individual fixations might become unrecognizable, but in the context of long-term recording, it is typically more important to identify patterns of specific activities, which would still be possible, even with a reduced sampling rate. If a frame-by-frame representation is depicted, gaze spirals scale better than their linear counterpart. Figure 10 shows a comparison between a linear slitscan and the corresponding gaze spiral. It should be noted that the linear slitscan results in an image with a resolution of $367,122 \times 100$ pixels, which is often not even a valid resolution for rendering libraries. We scaled the width of the slitscan down to 5% of the original size before we could compare both
visualizations. Given the same screen width, the gaze spiral is easier to interpret without extensive zooming.

**Interpretability:** Visual patterns in gaze spirals mainly depict time spans when people where fixating a region (constant pattern) and changes of the point of regard (abrupt change in the pattern). The identification of similar patterns over time and between recordings is supported by queries. However, the interpretation and assignment of specific patterns to the respective AOI or activity requires analysts to investigate the underlying video, especially if a new pattern is analyzed. Hence, we implemented preview thumbnails that show on demand the corresponding video frame to help interpret patterns in the spiral.

**Sequence Interpretation:** We see the main purpose of gaze spirals in the compact representation of scanpath properties in the context of the stimulus. This means, we aim to overview the sequential order participants looked at AOIs, starting in the center of the spiral. With time progression, the spiral will grow and new timesteps appear on the outer arm of the spiral. Due to the inherent design of the gaze spiral, these new time spans will have more visual space than the inner parts. With this design, gaze spirals are also useful for live visualization, where current events are more important than older events. However, this design impairs a comparison of sequence durations. We could achieve this by decreasing $k$ depending on $t$. Due to general perceptual issues of people interpreting angular segments in comparison to linear segments, we would rather suggest including an additional view with selected linear segments, if the exact visual comparison of sequence durations is necessary.

**Head and Eye Movement.** A more general discussion relates to the question in which scenarios an egocentric video is sufficient and when eye tracking provides benefits. Our approach supports replacing the gaze position by a static scanline to account for scenarios with much head movement. We noticed that a slitscan from the egocentric video alone results in much smoother patterns while the gaze-based approach results in short patterns with numerous abrupt changes. However, in situations with multiple AOIs in the field of view, the gaze-supported slitscan provides much more details. A hybrid solution could provide smooth results while preserving important details, by combining the egocentric approach in situations with much movement (e.g., walking) with the gaze-based approach when people investigate a situation in detail.

# 7 Conclusion and Future Work

We presented a new approach to visualize scanpaths of multiple recordings from mobile eye tracking data. The resulting gaze spirals show pattern sequences that indicate in which order people looked at different AOIs and performed specific activities. Individual gaze spirals serve as glyphs for each recorded scanpath and are displayed in a 2D embedding to help identify and interpret groups of similar behavior based on image-features, without the necessity to annotate AOIs. Note that similar to cluster analysis, our approach requires the user to be somewhat proficient with sequence alignment and 2D embeddings. One of our future goals is to make this process more accessible to the general, less proficient, user. At this time, we provide open access to our scripts for generating gaze spirals\(^2\). With our technical evaluation, we showed that image-based calculations with established scanpath comparison methods lead to results comparable with annotated data.

For future work, we further plan to conduct a user study on the annotation performance of gaze spirals in comparison other techniques such as the drawing and tracking of polygons. For the annotation of the art gallery dataset, we measured an average annotation time of 5:49 minutes per recording, which is approximately twice the average run time of the videos. We see the proposed

\(^2\)https://github.com/Maurice189/gaze-spirals. Last checked on April 14, 2022.
approach as a starting point for data exploration because it can be linked easily with detail views, e.g., gaze replays, while the overview is maintained. Thus, we think that gaze spirals are particularly relevant for hypothesis-generation purposes. Due to the high scalability with respect to time, we further plan to investigate if the technique could handle long-term pervasive eye tracking data, which will become increasingly important in the years to come.

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project ID 251654672 – TRR 161 (project B01) and under Germany’s Excellence Strategy – EXC 2120/1 – 390831618.

REFERENCES

Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. 2011. Visualization of Time-Oriented Data. Springer-Verlag, London.

Nicola C. Anderson, Fraser Anderson, Alan Kingstone, and Walter F. Bischof. 2015. A comparison of scanpath comparison methods. Behavior Research Methods 47, 4 (2015), 1377–1392.

Donald J. Berndt and James Clifford. 1994. Using Dynamic Time Warping to find patterns in time series. In Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop, Seattle, Washington. Technical Report WS-94-03, Usama M. Fayyad and Ramasamy Uthurusamy (Eds.). 359–370.

F. N. Bezerra and E. Lima. 2006. Low cost soccer video summaries based on visual rhythm. In Proceedings of the ACM International Workshop on Multimedia Information Retrieval. 71–78.

T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. 2017. Visualization of eye tracking data: A taxonomy and survey. Computer Graphics Forum 36, 8 (2017), 260–284.

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Stefan Strohmaier, Daniel Weiskopf, and Thomas Ertl. 2016. AOI hierarchies for visual exploration of fixation sequences. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications. 111–118.

Rita Borgo, Min Chen, Ben Daubney, Edward Grundy, Gunther Heidemann, Benjamin Höferlin, Markus Höferlin, Heike Leitte, Daniel Weiskopf, and Xianghua Xie. 2012. State of the art report on video-based graphics and video visualization. Computer Graphics Forum 31, 8 (2012), 2450–2477.

Andreas Bulling, Andrew T. Duchowski, and Päivi Majaranta. 2011. PETMEI 2011: the 1st international workshop on pervasive eye tracking and mobile eye-based interaction. In Proceedings of the 13th International Conference on Ubiquitous Computing. 627–628.

Michael Burch. 2018. Identifying similar eye movement patterns with t-SNE. In Proceedings of the Conference on Vision, Modeling, and Visualization. 111–118.

Nora Castner, Thomas C. Kuebler, Katharina Scheiter, Juliane Richter, Therese Eder, Fabian Huettig, Constanze Keutel, and Enkelejda Kasneci. 2020. Deep semantic gaze embedding and scanpath comparison for expertise classification during OPT viewing. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications. 18:1–18:10.

Revital Dafner, Daniel Cohen-Or, and Yossi Matias. 2000. Context-based space filling curves. Computer Graphics Forum 19, 3 (2000), 209–218.

Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvev. 2012. Eye tracker data quality: What it is and how to measure it. In Proceedings of the Conference on Vision, Modeling, and Visualization. 111–118.

Qi Jiang and XiaoJie Yang. 2002. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. Real-time Imaging 8, 5 (2002), 357–377.

Peter Kiefert, Ioannis Giannopoulos, Martin Raubal, and Andrew Duchowski. 2017. Eye tracking for spatial research: Cognition, computation, challenges. Spatial Cognition & Computation 17, 1-2 (2017), 1–19.

Maurice Koch, Kuno Kurzhals, and Daniel Weiskopf. 2018. Image-based scanpath comparison with slit-scan visualization. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications. 55:1–55:5.

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. 113–116.

Kuno Kurzhals. 2021. Image-based projection labeling for mobile eye tracking. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications. 1–12.

Kuno Kurzhals, Michael Burch, Tanja Blascheck, Gennady Andrienko, Natalia Andrienko, and Daniel Weiskopf. 2017. A task-based view on the visual analysis of eye-tracking data. In Eye Tracking and Visualization: Foundations, Techniques,
Marc Weber, Marc Alexa, and Wolfgang Müller. 2001. Visualizing time-series on spirals. In Anthony Tang, Saul Greenberg, and Sidney Fels. 2008. Exploring video streams using slit-tear visualizations. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications. 11–18.

Kuno Kurzhals, Marcel Hlawatsch, Florian Heimerl, Michael Burch, Thomas Ertl, and Daniel Weiskopf. 2015. Gaze stripes: Image-based visualization of eye tracking data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 1005–1014.

Kuno Kurzhals, Marcel Hlawatsch, Christof Seeger, and Daniel Weiskopf. 2016b. Visual analytics for mobile eye tracking. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 301–310.

Kuno Kurzhals, Nils Rodrigues, Maurice Koch, Michael Stoll, Andreas Bruhn, Andreas Bulling, and Daniel Weiskopf. 2020. Visual analytics and annotation of pervasive eye tracking video. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications*. 1–9.

Kuno Kurzhals and Daniel Weiskopf. 2016. Visualizing eye tracking data with gaze-guided slit-scans. In *Proceedings of the IEEE Workshop on Eye Tracking and Visualization (ETVIS)*. 45–49.

Michael F. Land and Mary Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision Research* 41, 25-26 (2001), 3559–3565.

Vladimir I. Levenshtein et al. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10. 707–710.

Yin Li, Miaoyan Liu, and James M Rehg. 2018. In the eye of the beholder: Joint learning of gaze and actions in first person video. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 619–635.

Leland McInnes, John Healy, and James Melville. 2018. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

Simona Naspetti, Roberto Pierdicca, Serena Mandolesi, Marina Paolanti, Emanuele Frontoni, and Raffaele Zanoli. 2016. Automatic analysis of eye-tracking data for augmented reality applications: A prospective outlook. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. 217–230.

Rudolf Netzel, Bettina Ohlhausen, Kuno Kurzhals, Robin Woods, Michael Burch, and Daniel Weiskopf. 2017. User performance and reading strategies for metro maps: An eye tracking study. *Spatial Cognition & Computation* 17, 1-2 (2017), 39–64.

Sohail Rafiq, Chatchai Wangwiwattana, Jasmine Kim, Ephrem Fernandez, Suku Nair, and Eric C Larson. 2015. PupilWare: Towards pervasive cognitive load measurement using commodity devices. In *Proceedings of the ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*. 42:1–42:8 pages.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372–422.

Eyal M. Reingold and Heather Sheridan. 2011. Eye movements and visual expertise in chess and medicine. In *The Oxford Handbook of Eye Movements*, Simon P. Liversedge, Iain D. Gilchrist, and Stefan Everling (Eds.). Oxford University Press, Oxford, 523–550.

Daniel C. Richardson and Rick Dale. 2005. Looking to understand: The coupling between speakers’ and listeners’ eye movements and its relationship to discourse comprehension. *Cognitive Science* 29, 6 (2005), 1045–1060.

Christoph Schulz, Michael Burch, Fabian Beck, and Daniel Weiskopf. 2017. Visual data cleansing of low-level eye-tracking data. In *Eye Tracking and Visualization: Foundations, Techniques, and Applications (ETVIS 2015)*, Michael Burch, Lewis Chuang, Brian Fisher, Albrecht Schmidt, and Daniel Weiskopf (Eds.). Springer International Publishing AG, Cham, Switzerland, 199–216.

Steven M. Seitz and Jiwon Kim. 2003. Multiperspective imaging. *IEEE Computer Graphics and Applications* 23, 6 (2003), 16–19.

Temple F. Smith and Michael S. Waterman. 1981. Identification of common molecular subsequences. *Journal of Molecular Biology* 147, 1 (1981), 195–197.

Julian Steil, Michael Xuelin Huang, and Andreas Bulling. 2018. Fixation detection for head-mounted eye tracking based on visual similarity of gaze targets. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications*. 23:1–23:9.

Anthony Tang, Saul Greenberg, and Sidney Fels. 2008. Exploring video streams using slit-tear visualizations. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 191–198.

Marc Weber, Marc Alexa, and Wolfgang Müller. 2001. Visualizing time-series on spirals. In *Proceedings of the IEEE Symposium on Information Visualization*. 7–14.

Julian Wolf, Stephan Hess, David Bachmann, Quentin Lohmeyer, and Mirko Meboldt. 2018. Automating areas of interest analysis in mobile eye tracking experiments based on machine learning. *Journal of Eye Movement Research* 11, 6 (2018), 1–6.

Proc. ACM Comput. Graph. Interact. Tech., Vol. 5, No. 2, Article 20. Publication date: June 2022.