

ISeeCube: Visual Analysis of Gaze Data for Video

Kuno Kurzhals*, Florian Heimerl†, Daniel Weiskopf‡

University of Stuttgart

Abstract

We introduce a new design for the visual analysis of eye tracking data recorded from dynamic stimuli such as video. ISeeCube includes multiple coordinated views to support different aspects of various analysis tasks. It combines methods for the spatiotemporal analysis of gaze data recorded from unlabeled videos as well as the possibility to annotate and investigate dynamic Areas of Interest (AOIs). A static overview of the complete data set is provided by a space-time cube visualization that shows gaze points with density-based color mapping and spatiotemporal clustering of the data. A timeline visualization supports the analysis of dynamic AOIs and the viewers' attention on them. AOI-based scanpaths of different viewers can be clustered by their Levenshtein distance, an attention map, or the transitions between AOIs. With the provided visual analytics techniques, the exploration of eye tracking data recorded from several viewers is supported for a wide range of analysis tasks.

CR Categories: Human-centered computing [Visualization]: Visualization application domains—Visual analytics;

Keywords: Eye tracking, space-time cube, dynamic areas of interest, visual analytics

1 Introduction

How people look at different stimuli is important for research questions about human cognition [Duchowski 2002] and for the optimization of visualization designs [Burch et al. 2011]. The main focus of the past was on the analysis of static stimuli. Over the last years, however, the analysis of dynamic stimuli such as video has been gaining increasing relevance. For example, Tien and Zheng [2012] measured gaze overlaps of a video that showed a surgical task to compare experts' gazes with the gazes of trainees. Goldstein et al. [2007] examined similarities in the viewing behavior of several users to identify centers of interest in movie scenes. Marchant et al. [2009] described an approach to investigate the influence of directorial techniques on film viewers' experience. Our goal is to provide visual analytics techniques that can be used to complement the quantitative means of analysis from those papers, such as timespans of high attentional synchrony.

The analysis of the recorded gaze data can be performed statistically, resulting in quantitative measures that often require advanced expertise for interpretation. Combined with appropriate visualization techniques, the statistical results can be supported visually and an explorative analysis of the data is possible. To this end, ISeeCube provides a visual analytics approach for the qualitative

exploration of the data. Visual analytics can be defined as “the science of analytical reasoning facilitated by visual interactive interfaces” [Thomas and Cook 2005]. It combines automated methods (e.g., statistical analysis, data mining) with visualization techniques and human-computer interaction to extract knowledge from the data.

Visualization methods for eye tracking data mainly concentrate on static stimuli. Bee swarms, scanpaths, and heat maps are the most common methods to visualize the gaze of individual or several viewers. Although the application of these methods to dynamic stimuli is possible, their use is tedious, since the user has to watch the complete video with the visualization overlay to gain insights. With the Space-Time Cube (STC) and a linked timeline visualization, we provide an overview of the video that allows for a non-sequential search for timespans of interest. Density-based color mapping of gaze points and spatiotemporal clustering indicate where the user should look for timespans of potential interest (see Section 4). For a detailed analysis of selected timespans, the bee swarm visualization and heatmaps (animated and static) are still available in ISeeCube.

For a statistical analysis of eye tracking data, the definition of Areas of Interest (AOIs) in the stimulus becomes inevitable. Metrics such as dwell times [Poole and Ball 2006] can be applied to the AOIs to provide a quantitative summarization of the recorded data or for further data visualization. To provide an interface between qualitative and quantitative analysis, ISeeCube includes an editor that can be used to annotate AOIs before the gaze data is analyzed, or in combination with additional information gained from the visualized data. With these annotations, AOI-based scanpath comparisons of selected timespans can be performed interactively.

To demonstrate the effectiveness of ISeeCube, we re-investigate a previously evaluated data set [Kurzhals and Weiskopf 2013] and present examples of how information can be extracted with the new visualization approach (see Section 7).

2 Related Work

Space-Time Cube visualizations provide a static overview of spatiotemporal data like recorded gaze. The data is represented in a 3D volume with one temporal two spatial and axes. Fixations can be visualized as series of static points along the time axis and scanpaths as 3D lines. Li et al. [2010] used the STC to visualize eye trajectories for static stimuli. For the application to dynamic stimuli, Duchowski and McCormick [1998] described a space-time representation of Volumes Of Interest for aggregated eye movement trajectories. Kurzhals and Weiskopf [2013] visualized gaze data from recorded video stimuli in an STC with density-based color mapping of data points and spatiotemporal clustering in order to extract AOIs and find timespans of attentional synchrony. These papers did not include the possibility to analyze scanpaths of viewers, based on dynamic AOIs. We extend the work of Kurzhals and Weiskopf [2013] by new AOI-based visualization techniques and interactive viewer comparison methods.

*e-mail: kuno.kurzhals@visus.uni-stuttgart.de

†e-mail: florian.heimerl@visus.uni-stuttgart.de

‡e-mail: daniel.weiskopf@visus.uni-stuttgart.de

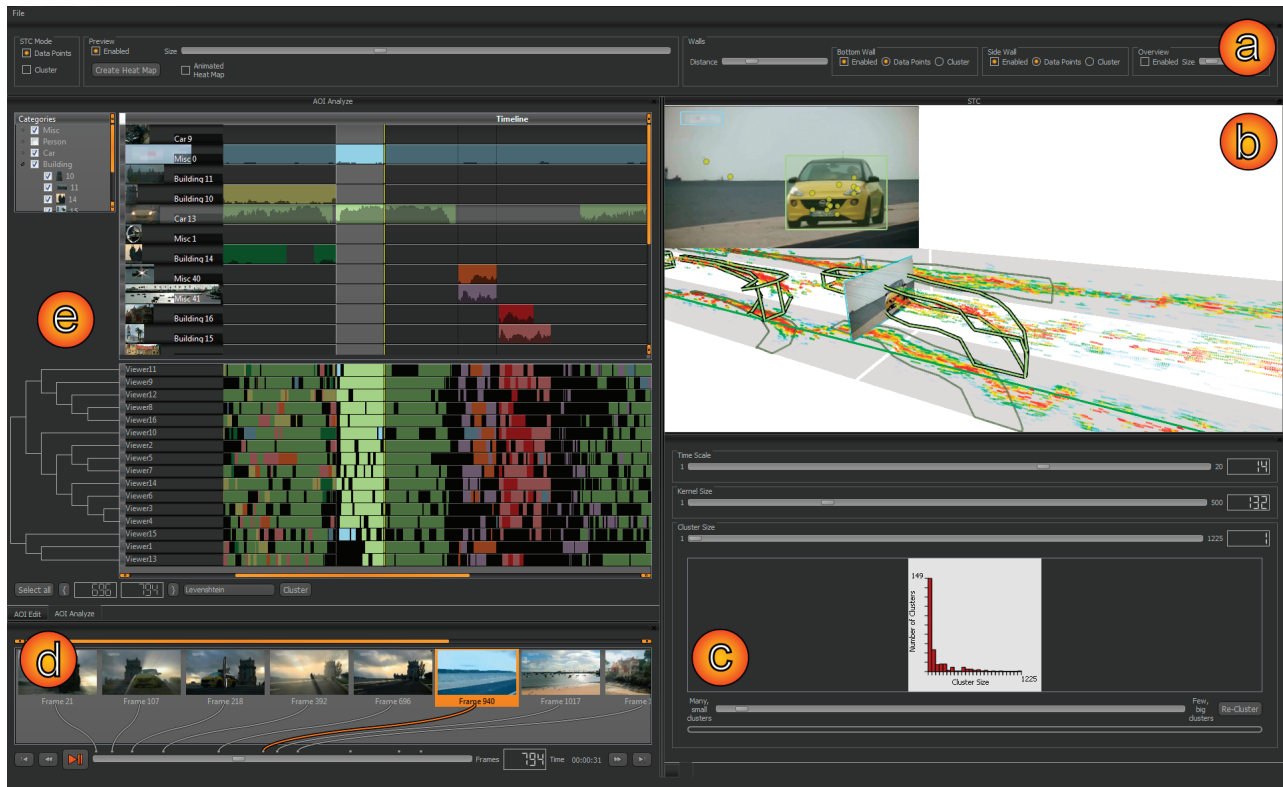


Figure 1: Overview of the main design components: (a) viewer controls; (b) space-time visualization view; (c) parameter controls; (d) video controls; (e) timeline visualization view.

Timeline representations are a common method to visualize the temporal progression of events. André et al. [2007] designed detailed timelines for hierarchies, relationships, and scale. We adapt the principle to provide additional AOI information on demand and adjust the presented information to the special requirements of gaze data recorded from videos. In the field of eye tracking, Andrienko et al. [2012] used horizontal segmented bars in a temporal view to visualize the distance of eye trajectories to selected AOIs of static graphs. Ristovski et al. [2013] used fixation time series similar to scarf plots with a highlighting function for fixations on the same AOI. These papers focused on static stimuli.

For the analysis of dynamic stimuli, Richardson et al. [2005] used a scarf plot to visualize the recurrence of eye movements between two persons. Weibel et al. [2012] integrated mobile eye tracking data in ChronoViz, a tool to visualize multiple streams of time series data simultaneously. They used separate scarf plots for individual AOIs and concentrated their analysis on individual viewers; a similar approach was used by Lessing and Linge [2002]. Stellmach et al. [2010] introduced a models-of-interest timeline that shows a viewer’s gaze distribution between various 3D objects in a virtual environment with individually selectable colors for each object. In their work, the main focus lies on the visualization of a single viewer’s gaze data over time with a constant set of objects. In our approach, we extend these ideas by allowing comparisons between multiple viewers and providing an AOI timeline that visualizes the dynamic changes of AOIs during a video.

For the visualization of scanpath comparisons, West et al. [2006] introduced eyePatterns, a software to identify patterns and similarities across fixation sequences. In their work, AOIs on a static stimulus are labeled by characters, leading to visualization approaches that require the user to remember for which AOI a label stands.

In dynamic stimuli, the number of AOIs can increase significantly compared to static stimuli and with a character-based string, the identification of the corresponding AOIs becomes an additional, inefficient search task. Tsang et al. [2010] used a timeline in combination with a tree visualization to compare duration, frequency, and orderings of fixations, recorded from a mobile eye tracking device. In their approach, only the temporal information from fixation data is used. We include additional information from dynamic AOIs to solve further research questions; e.g. we can analyze how long an AOI was visible before the viewers looked at it.

Our main contribution is a visualization design that combines spatiotemporal analysis methods of the STC with a new AOI-oriented timeline visualization with viewer scarf plots that allows for an interactive scanpath analysis and an efficient interpretation of the results. ISeeCube allows for an explorative analysis of gaze data that is supported by automated techniques. Although it is designed for video stimuli, the application to static stimuli is also possible.

3 Design Overview

To support various analysis tasks that require different visualization approaches, ISeeCube consists of adjustable, multiple coordinated views. Figure 1 shows an overview of the main components that comply with the design of Kurzhals and Weiskopf [2013], except for the timeline visualization view that provides additional new methods for the analysis of the data. For the data representation of each viewer, we use a comma separated format, containing x- and y-coordinates and the frame number. ISeeCube includes no fixation filtering algorithms itself, but can be used to visualize prefiltered or raw gaze data.

The main components of ISeeCube are:

- (a) **Viewer controls:** The viewer controls adjust the space-time visualization view to the needs of the user. Data points and clusters for the STC, the video preview with animated heat maps, and additional wall projections of the data can be enabled as needed.
- (b) **Space-time visualization view:** Within the STC, data points, clusters, and scanpaths can be visualized. Additionally, the spatiotemporal extent of annotated AOIs can be visualized on demand (see Section 5.2).
- (c) **Parameter controls:** By parameter manipulation, the STC visualization can be adjusted interactively. The parameters provide the user with the possibility to control the time scale, kernel size for density-based filtering of data points, and filtering options for the clustering.
- (d) **Video controls:** The video navigation is achieved by standard controls including a time slider and buttons for playback and forward-/backward navigation. We can directly jump to shot boundaries of a video by surrogate images connected to their position on the time slider.
- (e) **Timeline visualization view:** The timeline visualization provides temporal information of annotated AOIs and scarf plots of individual viewers (see Section 6). Selecting an AOI in the timeline shows its representation in the STC and provides additional information of its spatiotemporal extent.

Additionally, ISeeCube provides an editor (see Section 5.1) for the annotation of AOIs that are required for advanced visual analytics techniques, and a control dialog to activate individual 3D scanpaths.

4 Analysis of Unlabeled Stimuli

The STC visualization allows for explorative analysis tasks that require no AOI annotation. The static representation of the spatiotemporal data can be investigated interactively to find interesting timespans directly without a complete sequential analysis of the video. This part of our visual analytics approach uses the techniques described by Kurzahls and Weiskopf [2013]. For completeness, we briefly summarize their techniques in this section and refer to their paper for more details. The data point and the cluster representation of the data can be used to find sequences of high attentional synchrony as well as timespans that contain multiple AOIs.

4.1 Attentional Synchrony

The presence of motion in a video has a significant influence on the viewers' attention [Mital et al. 2011]. As a result, the viewers' gaze can concentrate on a moving object in the same timespan, even if the viewers watched the video independently from each other. This effect is commonly known as attentional synchrony [Smith and Henderson 2008]. To find attentional synchrony in a data set, gaze is represented by data points that can be filtered by their distance to the centroid of all points in a frame. As a result, only dense regions with high attentional synchrony remain and can be detected efficiently in the STC overview. Figure 2 shows an example scene. Although the presented scene contains various AOIs, the viewers attended to the car while it was moving toward the camera.

4.2 Multiple AOIs

During timespans of attentional synchrony, usually one AOI is of high importance. In scenes with several moving objects or in static

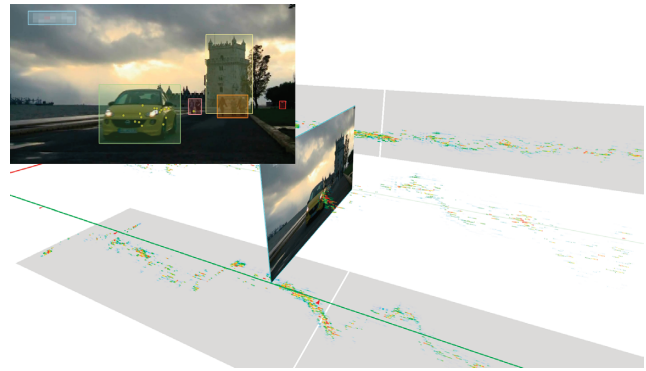


Figure 2: Attentional synchrony: several viewers attended to the same object simultaneously.

scenes without salient motion, the viewers' gaze is usually distributed more asynchronously between different AOIs. A data mining approach to find potential AOIs in a data set applies clustering algorithms to the gaze data. We use the mean shift algorithm for clustering eye tracking data [Santella and DeCarlo 2004] with additional shot boundary information.

With a visualization of the spatiotemporal clustered gaze data, timespans and positions of multiple AOIs in the video can be identified in the STC. The clusters are represented as smoothed hulls around the gaze points they contain. In the video preview, animated bounding boxes show the current spatial extent of a cluster as well as a number that represents the cluster size by the amount of included gaze points. Figure 3 shows an example of a timespan with two AOIs and a blue and a yellow cluster. The corresponding video frame reveals that the viewers' attention was distributed between two buildings. The blue cluster on the lighthouse has a higher number for cluster size, indicating that more attention was on this region.

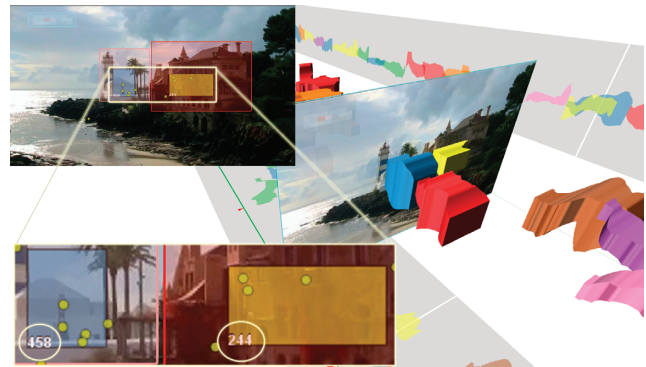


Figure 3: Multiple AOIs: the extracted gaze clusters (blue and yellow) indicate the spatial position and timespan where potential AOIs exist in the video.

As Kurzahls and Weiskopf [2013] pointed out, the STC visualization is not sufficient for all analysis tasks. The analysis of individual viewers and the visualization of similarity measures between viewers can hardly be achieved with the STC alone. Therefore, additional analysis methods are required. In particular, the definition of dynamic AOIs provides new possibilities to solve a wide range of additional research questions. In the following sections, we describe how we extend their work with AOI-based visualization techniques.



Figure 4: The editor helps define dynamic, axis-aligned AOIs.

5 Annotation and Representation of AOIs

To support quantitative research on the data, the definition of AOIs becomes inevitable. The user can either define AOIs in advance to investigate scanpaths or the distribution of attention, or define additional AOIs after an explorative analysis of the data. For this purpose, the extracted clusters (see Section 4.2) can be investigated, as they indicate the position of the AOIs where most of the viewers’ attention was. The definition of dynamic AOIs is supported by an integrated editor that transforms the annotations into representations for the STC and the timeline visualization.

5.1 AOI Editor

For the analysis of dynamic stimuli, the definition of AOIs that adjust to the changes of moving objects becomes an important step in the analytical process. We provide an editor (Figure 4) that allows for the definition of dynamic, axis-aligned bounding boxes to mark areas of interest. In our system, the ViPER-GT [Doermann and Mihalcik 2000] file format can be imported; ViPER-GT is a ground truth editor commonly used for the annotation of video data sets. Different categories can be defined to specify the analysis. With the information provided by the cluster visualization, the user can identify the most important objects and areas in a video and annotate them with the editor. A new AOI is created by drawing key bounding boxes in the video at different positions during the playback. Between the key positions, the bounding box is interpolated linearly. Successive IDs are used for new AOIs, independent from the category.

5.2 AOI Representation in the STC

Similar to the cluster visualization, the defined AOIs can be represented by static objects in the STC. Figure 1(b) shows the AOI that annotates a car that appears in several shots of the video. The spatiotemporal shape of the bounding box summarizes movement and changes in size of the object for the complete video. The AOI representations are also projected onto the walls to overcome problems of depth perception. To solve occlusion problems due to overlaps of different AOIs, we show only the outlines of the AOI representations and provide the user the possibility to show individual AOIs, selected from the AOI timeline (see Section 6.1).

6 Timeline Visualization

The representation of AOIs in the STC provides valuable information about the spatiotemporal extent of individual objects, but due to occasional overlaps, displaying all AOI representations simultaneously leads to visual clutter. In the timeline visualization, this problem is solved by reducing the visualization to the temporal compo-

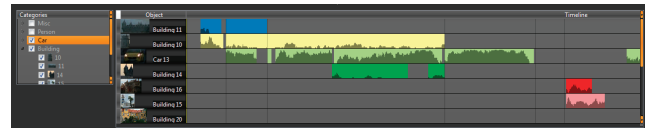


Figure 5: The AOI timeline (right) shows annotated objects ordered by their first appearance. Colored bars with attention histograms indicate when an AOI was visible and how many viewers looked at it over time. The category tree (left) allows the user to hide or unhide AOIs from the timeline.

nent of the data and providing spatial information on demand in the STC. The timeline consists of two synchronized streams that convey the information of the chronological appearance of AOIs and the scanpaths of all viewers. This combination allows for an efficient identification of AOIs existing in a timespan and the order in which viewers attended to them.

6.1 AOI Timeline

Similar to the STC visualization, the AOI timeline provides an overview of the complete data set, but without the spatial information. All AOIs are represented by rows, ordered by their first appearance in the video. The first column shows the name and a representative image of each AOI. The second column shows a colored bar for each timespan in which the AOI exists (Figure 5). The distribution of all viewers’ attention is displayed by an attention histogram in the colored bar. To distinguish between different AOIs, a qualitative color scheme of 11 colors [Harrover and Brewer 2003] was used. A color is locked to an AOI as long as the AOI exists and can be mapped to another one as soon as the respective AOI disappears. This strategy ensures an unambiguous mapping from AOIs to colors, as long as there are fewer than 12 AOIs with overlapping life spans. For additional AOIs, the color scheme is repeated and possible ambiguities have to be solved by looking at the attention histograms and the video preview.

Ordering the AOIs by their first appearance results in a timeline where early appearing objects are placed in upper rows and late appearing objects in lower rows. This leads to problems when objects appear early and reappear several times in the video. In this case, a gap of empty rows occurs between the late appearing objects at the lower rows and the reappearing object in the upper row. Comparing the histograms of the involved AOIs becomes more difficult the farther the rows are apart. To solve this problem, we included a tree view left to the AOI timeline (Figure 5) that shows all objects ordered by their category. Either the complete category or individual objects can be disabled to hide them in the timeline. With this approach, users can exclude all objects that are not present in the currently investigated timespan to concentrate on the relevant information.

To obtain additional AOI information on demand, each row can be selected individually to show an overview (see Figure 6). Each overview is presented in a separate window and can be activated as needed. The information provided by the overview consists of a filmstrip and four histograms. The filmstrip shows representative frames from the timespan of the marked AOI. The frames are chosen by dividing the timespan in four equal parts. If an AOI does not exist in one of the inner frame positions, the parts are further divided until a valid frame is reached. The histograms show the distribution of attention, AOI size, and AOI position:

- **Attention:** The attention histogram is the same as in the timeline. The numbers mark the position of the frames from the filmstrip.

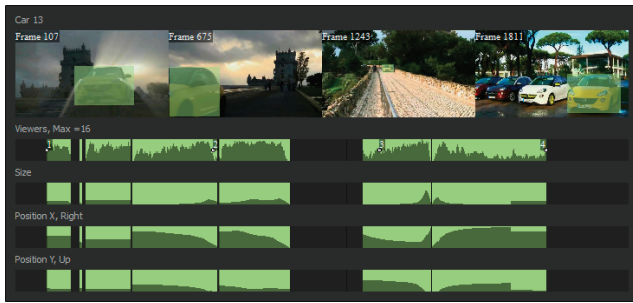


Figure 6: The overview provides a filmstrip of the selected AOI and histograms for attention, size, and x/y position.

- **Size:** The size of an AOI is measured as area size relative to the video resolution. In Figure 6, the size histogram shows several timespans where the size of the AOI increases at the end of a shot. In combination with the information provided by the position histograms, we can interpret that the car was moving close to the camera in these situations.
- **Positions X,Y:** x- and y-coordinates are measured at the bottom-center point of the AOI. In the histograms, high values represent a position in the right part of the scene and in the upper part, respectively.

For future work, the provided information about the temporal development of AOI-related measures can be enriched by additional histograms such as an attention measure, normalized by the size of an object. To analyze the scanpaths of the recorded viewers, the scarf plot visualization can be used.

6.2 Viewer Scarf Plots

Each viewer in the data set is represented by a horizontal scarf plot that shows a frame-wise mapping of the viewer's gaze data to the annotated AOIs. If a gaze point in a frame is considered to be in an AOI, the corresponding color of the AOI is used to mark this frame in the plot. If either no gaze point is available, or cannot be assigned to an AOI, the frame is marked black. A gaze point is assigned to an AOI when it lies inside the bounding box of the AOI in the respective frame. Due to the dynamic content of video stimuli, overlaps between AOIs are often inevitable. Two common methods to handle this problem are either to distribute attention between the overlapping AOIs, or to calculate the distances between the gaze point and the involved AOI centers and assign the point to the AOI with the shortest distance [Holmqvist et al. 2011]. We decided to use the latter, since ambiguities are not supported by the common string comparison algorithms that we used for scanpath analysis.

Since the AOI timeline and the viewer scarf plots are synchronized, the user can see directly which AOIs are involved in the current timespan and what object they represent. As an example, Figure 1(e) shows a selected timespan with two AOIs (Misc 0, Car 13). Corresponding to the attention histograms, the viewer scarf plots show that all viewers attended to Car 13, only Viewer 10, Viewer 11, and Viewer 15 attended to Misc 0 for a short period of time.

The generated scarf plots can be interpreted as strings of IDs that refer to their AOIs. These strings can be used for interactive scanpath analysis, whereas the scarf plot visualization allows for an easy interpretation of the results.

6.3 Scanpath Similarity and Cluster Analysis

To support advanced analysis in the context of visual analytics, ISeeCube integrates automatic processing of eye tracking data to assess the similarity of viewers' scanpaths. We provide three different similarity functions that can be used with hierarchical clustering. Users can thus explore different facets of the dataset and select the distance function that fits their objectives and analytical goals best. The similarity functions available are the Levenshtein distance, a function based on attention distribution, and one based on AOI transitions. They are comparable to those by Ristovski et al. [2013], but have been extended to support video eye tracking data.

The Levenshtein distance [Levenshtein 1966] was used previously as a distance measure for object scanpaths [Holmqvist et al. 2011] by representing scanpaths as strings of AOIs. Levenshtein's algorithm yields a distance for two input strings that represents the number of edit operations to transform one string into the other. As our measure is specifically tailored to videos rather than static images, we incorporate gaze duration for each AOI by constructing a frame-wise string of AOIs, where each character of the scanpath string corresponds to the AOI focused in each frame of the video. The inversion of the Levenshtein distance yields a similarity measure incorporating local and temporal coherence of the scanpath strings and penalizes similar object transitions that have low temporal correlation.

The second similarity measure focuses on the distribution of viewers' attention. It aggregates the overall attention that each of the viewers gave to each of the AOIs by counting the number of video frames during which a viewer was looking at the respective AOI. It then normalizes this value with the maximum attention that a viewer gave to an AOI. To quantify the difference between the resulting attention maps we use the squared difference between each of the components, which is normalized with the overall number of AOIs. A similar measure for the attention map difference for still images is mentioned by Holmqvist et al. [2011].

The third similarity measure focuses on viewers' transitions between AOIs. It is based on a transition matrix for each viewer including an initial and a final state to incorporate the first and last AOI of a scanpath. Each of the values of the transition map is normalized by the maximum number of transitions for a pair of AOIs. Again, similarity is quantified by the sum of squares of pairwise differences in normalized transition frequency, which is normalized with the overall number of pairs of AOIs.

Based on a selected timespan, we provide a hierarchical clustering algorithm [Hastie et al. 2009] that finds groups of similar scanpaths according to the chosen similarity measure. The clustering is visualized as a dendrogram to the left of the scarf plots (Figure 7), and provides users with a global picture of scanpath similarities. Hierarchical clustering starts out with a maximal number of clusters, with each scanpath forming its own cluster. The clusters are then merged consecutively, with the pair of most similar clusters being merged at each iteration of the algorithm. We use average linkage to measure cluster similarity, i.e., the arithmetic mean of all pairs of instances in two different clusters.

Figure 7 illustrates the different effects of the similarity measures on scanpath clustering. The selected scene (see Figure 8) contains three labeled objects: a TV station logo (light blue), a woman (orange), and a car with its engine opened up (purple). Figure 7(a) shows the clustering with the Levenshtein measure. Here, the order of the AOIs and their temporal correlation play a major role for the similarity of two scanpaths. The first cluster on the highest level comprising the upper five elements of the list can e.g. be described

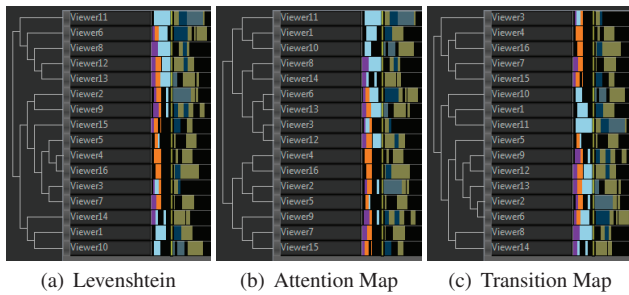


Figure 7: Clustering of the initial scene of the video containing a TV station logo (light blue), a woman (orange), and a car (purple). The scene is clustered using each of the three similarity measures.

as containing all scanpaths in which viewers were looking solely at the station logo (light blue) during the second half of the scene, while the second cluster comprises the rest of the scanpaths.

Figure 7(b) contains a clustering with the attention map measure, considering only the overall attention that each AOI received from a viewer. Here, it is also intuitive to find a good description of the two main clusters, with the one comprising the upper part of the list containing all scanpaths where the attention for the station logo (light blue) dominates the scanpath. This cluster is then further subdivided into scanpaths that contain solely the station logo (the upper three elements), and scanpaths with further AOIs. The lower part of the list contains scanpaths in which the woman (orange) and the car (purple) together received most of the attention.

The transition-based clustering is depicted in Figure 7(c). Here, a division is made between scanpaths that end with the woman (orange) and those that end with the station logo (light blue). The first cluster is further subdivided into scanpaths that start with the car (purple), and those that start with the woman. Viewer 3 can be easily identified as an outlier within this cluster. The scanpath starts with the car, continues with the logo, and finally ends with the woman thus containing three AOIs. The difference between the transition-based and the attention-based measure is illustrated by how differently they treat the scanpaths of Viewers 3 and 12. With the attention map measure, their respective clusters are merged very early, due to the very similar distribution of attention between all three AOIs. The transition-based measure, in contrast, merges their clusters in the final iteration because both scanpaths have only one identical transition: from the initial state to the car.

7 Case Study

To demonstrate how ISeeCube can be used to extract information from recorded gaze data, we re-investigate the data set described by Kurzhals and Weiskopf [2013]. We can reproduce their STC-related results and substantially improve the analysis of the data set with AOI-based methods. In a qualitative user study, they reported the strategies and findings of five participants for the analysis of a promotional video. The video (length: 1:23 min, resolution: 1280×720) was shown on a Tobii T60 XL eye tracker with a 24" screen (resolution: 1920×1200) at a distance of 65 cm from the eyes. The recorded data was preprocessed with the Tobii Fixation Filter (velocity threshold = 10 pixels/samples; distance threshold = 10 pixels), representing smooth pursuits as short fixations close to the raw data. In separate sessions, 16 viewers watched the video as one of 13 short clips in a random order. The viewers were instructed to watch attentively and summarize the main plot after each video. In the investigated video clip, a new car is presented and

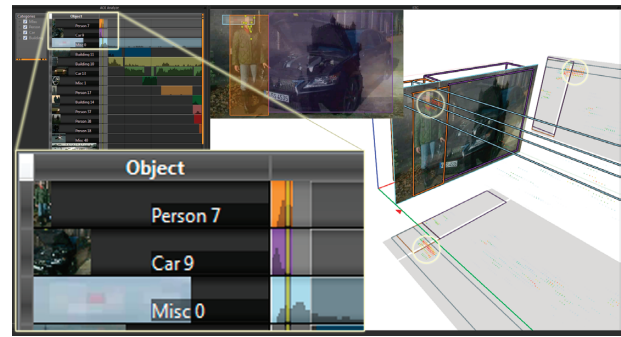


Figure 8: Attentional synchrony: shortly after the video starts, all viewers attended to the face of the person and the station logo. In the STC, high attentional synchrony is clearly visible (indicated by the circles).

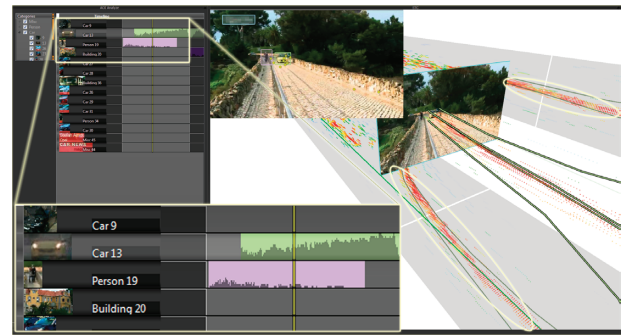


Figure 9: Motion signatures: high attentional synchrony on the car leads to characteristic motion signatures that are clearly visible on the walls (indicated by the ellipses).

various shots of the driving car along with views of Mediterranean landscapes are shown. An accompanying video and high resolution images of the case study can be found at <http://go.visus.uni-stuttgart.de/stva>. The following paragraphs discuss six different findings from the eye tracking data and how they were obtained by using ISeeCube.

1. Attentional synchrony: The video begins with a woman, standing next to a car. After a filter step to remove spatially sparse data, the data point representation in the STC (Figure 8) shows high attentional synchrony by densely located gaze points in the upper left corner of the video, the position where the face of the woman and the station logo were. Although attentional synchrony on an AOI leads to peaks in the attention histogram, this conclusion cannot be made from the AOI histogram alone: The attention histograms in the object timeline show that the viewers looked at the car (Car 9) first and then shifted their attention to the smaller region where the face of the woman (Person 7) and the station logo (Misc 0) were present. Since the size of Car 9 (purple box in the STC) covers more than half of the scene and the viewers' gazes were distributed in this area (see projection walls at the very beginning), attentional synchrony is not present. This example shows that AOI-only analysis can be misleading and that it should be accompanied by STC-based investigation.

2. Motion signatures: In the video, four shots show the promoted car driving through the scene. High attentional synchrony on the car is present in all of these shots. As a result, the data point representation clearly shows signatures of the motion of the car. With the timeline visualization, the additional AOI information can be

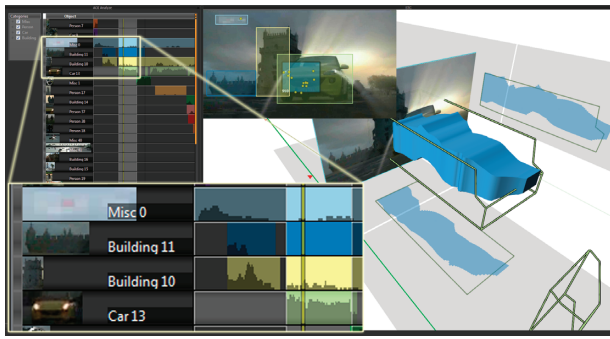


Figure 10: Dominant spatiotemporal gaze cluster: the cluster in the selected timespan corresponds well with the annotated AOI of the car.

used to investigate how changes in the properties of an AOI influenced the distribution of attention. As an example, Figure 9 shows the third shot where the car drives over a bridge. The attention histogram (Car 13) shows that an increasing number of viewers attended to the car while it was moving toward the camera. The size histogram in Figure 6 shows this shot after the first long timespan without the object. We can see that the object size increased significantly toward the end of the shot, resulting in higher values in the attention histogram.

3. Dominant spatiotemporal gaze cluster: By filtering out clusters that contain few gaze points, one of the biggest clusters can be found in the timespan in which the promoted car is presented for the first time (see Figure 10). The AOI representation of the car shows that the identified cluster corresponds well to the annotated AOI; this supports the approach of using cluster information for the definition of dynamic AOIs. The attention histogram confirms that Car 13 received most of the viewers’ attention in this timespan.

4. Distribution of attention: At the end of the video, a spokesperson (Person 34) appears talking about the car (see Figure 11). By filtering out small clusters, this event becomes visible in the STC. With the additional AOI information, we can investigate how long different objects received the viewers’ attention while they existed: With the attention histograms in the timeline, as well as with the gaze points on the projection walls, we can see that the attention of many viewers lies on the face at the beginning, but shifts to the lower left part of the video for a short time (red cluster). At this point in time, the name of the person (Misc 45) blends in, the viewers’ attention is drawn to the text and then back to the face. With the timeline visualization, the interpretation of the temporal distribution of attention can be achieved more efficiently, compared to common techniques such as a bee swarm or an animated heat map.

5. Outlier detection and individual viewer behavior: Figure 12 shows a timespan in the video with a beach and several boats in the water. By investigating the scarf plots in this timespan, we can see that Viewer 10 was the only one who looked mainly at the logo (Misc 0) in the upper left corner during the shot. The clustering with the Levenshtein distance confirms that Viewer 10 is an outlier in this timespan. This event could also be detected by investigating the 3D scanpaths of all viewers simultaneously. However, when the AOIs are close to each other, the 3D scanpaths would cause visual clutter that impairs the identification of such an outlier scanpath, while the scarf plots would show a result similar to the presented finding.

6. Multiple viewer groups: In the timespan that is shown in Figure 12, the sun reflects on a boat on the left (Misc 40). The clustering with the Levenshtein distance shows that two main groups of



Figure 11: Distribution of attention: the biggest clusters appear at the head of the person. The red cluster indicates the name (Misc 45) that blends in after the person begins to talk.

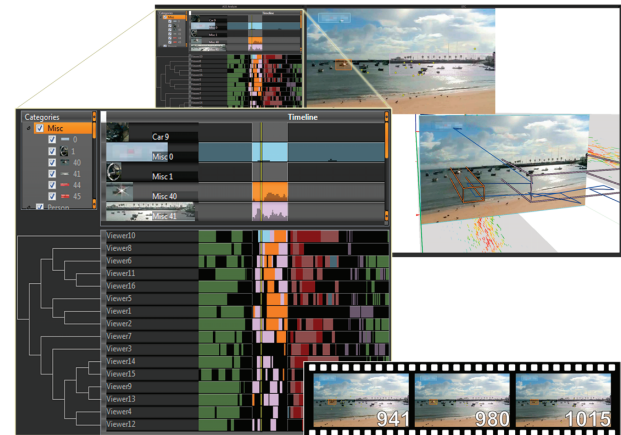


Figure 12: Outlier detection and individual viewer behavior: the attention of seven viewers was drawn to a reflection on a boat, while the others ignored it. Viewer 10 (scanpath represented by the blue line) was the only one that attended to the station logo (Misc 0).

viewers can be identified for this scene: one group that looked at the boat shortly after the blinking reflection appeared (Viewers 8–2); and a second group that ignored this reflection (Viewers 7–12). Viewer 7 looked at the boat after the reflection disappeared; in this case, it is not clear if the reflection was the cause for the attention shift.

8 Conclusion

We presented a new visual analytics method for the analysis of gaze data, recorded from video, that combines the advantages of spatiotemporal analysis of the STC with AOI-based methods for the temporal analysis of viewers’ scanpaths. In multiple coordinated views, the user can explore the data interactively. In combination with automated algorithms (e.g., spatiotemporal gaze clustering, scanpath clustering), the user is supported in various analysis tasks. The presented case study shows that previous findings could be confirmed and new findings were made possible with the integration of AOI-based timeline visualizations and string comparison methods.

For future work, we plan to improve the definition of AOIs. Gaze clusters provide valuable information about the position of important objects that can be used to support automatic object detection and tracking. To improve the precision of AOI-based measures, we plan to include more complex structures for dynamic AOIs with

higher-order interpolation methods, a separate handling of smooth pursuits in the recorded data, and compensation methods for possible inherent delays of the eye tracking device. For further evaluation of the visualization design, we plan to conduct additional user studies with several analysis tasks as well as longitudinal studies with multiple analysis sessions to gain insight into the development of different strategies of users while they use ISeeCube.

Acknowledgements

This work was funded by the German Research Foundation (DFG) as part of the Priority Program “Scalable Visual Analytics” (SPP 1335).

References

- ANDRÉ, P., WILSON, M. L., RUSSELL, A., SMITH, D. A., OWENS, A., AND SCHRAEFEL, M. 2007. Continuum: designing timelines for hierarchies, relationships and scale. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, 101–110.
- ANDRIENKO, G., ANDRIENKO, N., BURCH, M., AND WEISKOPF, D. 2012. Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and Computer Graphics* 18, 12, 2889–2898.
- BURCH, M., KONEVTSOVA, N., HEINRICH, J., HÖFERLIN, M., AND WEISKOPF, D. 2011. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Transactions on Visualization and Computer Graphics* 17, 12, 2440–2448.
- DOERMANN, D., AND MIHALCIK, D. 2000. Tools and techniques for video performance evaluation. In *Proceedings of the 15th International Conference on Pattern Recognition*, 167–170.
- DUCHOWSKI, A., AND MCCORMICK, B. 1998. Gaze-contingent video resolution degradation. In *Proceedings of Photonics West’98 Electronic Imaging*, 318–329.
- DUCHOWSKI, A. T. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4, 455–470.
- GOLDSTEIN, R., WOODS, R., AND PELI, E. 2007. Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine* 37, 7, 957–964.
- HARROWER, M., AND BREWER, C. 2003. ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1, 27–37.
- HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. 2009. *The Elements of Statistical learning: data mining, inference and prediction*, 2 ed. Springer.
- HOLMQVIST, K., NYSTRM, M., ANDERSSON, R., DEWHURST, R., HALSZKA, J., AND VAN DE WEIJER, J. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- KURZHALS, K., AND WEISKOPF, D. 2013. Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2129–2138.
- LESSING, S., AND LINGE, L. 2002. IICap: A new environment for eye tracking data analysis. *Master’s thesis. University of Lund, Sweden*.
- LEVENSHTEIN, V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady* 10, 8, 707–710.
- LI, X., ÇÖLTEKIN, A., AND KRAAK, M.-J. 2010. Visual exploration of eye movement data using the space-time-cube. In *Proceedings of the 6th International Conference on Geographic Information Science*, 295–309.
- MARCHANT, P., RAYBOULD, D., RENSHAW, T., AND STEVENS, R. 2009. Are you seeing what I’m seeing? An eye-tracking evaluation of dynamic scenes. *Digital Creativity* 20, 3, 153–163.
- MITAL, P., SMITH, T., HILL, R., AND HENDERSON, J. 2011. Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation* 3, 1, 5–24.
- POOLE, A., AND BALL, L. J. 2006. Eye tracking in HCI and usability research. In *Encyclopedia of Human Computer Interaction*, C. Ghaoui, Ed. Idea Group Reference, 211–219.
- RICHARDSON, D. C., AND DALE, R. 2005. Looking to understand: The coupling between speakers’ and listeners’ eye movements and its relationship to discourse comprehension. *Cognitive Science* 29, 6, 1045–1060.
- RISTOVSKI, G., HUNTER, M., OLK, B., AND LINSEN, L. 2013. EyeC: Coordinated views for interactive visual exploration of eye-tracking data. In *Proceedings of the 17th International Conference on Information Visualisation*, 239 – 248.
- SANTELLA, A., AND DECARLO, D. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, 27–34.
- SMITH, T., AND HENDERSON, J. 2008. Attentional synchrony in static and dynamic scenes. *Journal of Vision* 8, 6, 773–773.
- STELLMACH, S., NACKE, L., AND DACHSELT, R. 2010. Advanced gaze visualizations for three-dimensional virtual environments. In *Proceedings of the 2010 Symposium on Eye Tracking Research & Applications*, 109–112.
- THOMAS, J. J., AND COOK, K. A. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press.
- TIEN, G., ATKINS, M. S., AND ZHENG, B. 2012. Measuring gaze overlap on videos between multiple observers. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, 309–312.
- TSANG, H. Y., TORY, M., AND SWINDELLS, C. 2010. eSeeTrack-Visualizing sequential fixation patterns. *IEEE Transactions on Visualization and Computer Graphics* 16, 6, 953–962.
- WEIBEL, N., FOUSE, A., EMMENEGGER, C., KIMMICH, S., AND HUTCHINS, E. 2012. Let’s look at the cockpit: exploring mobile eye-tracking for observational research on the flight deck. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, 107–114.
- WEST, J. M., HAAKE, A. R., ROZANSKI, E. P., AND KARN, K. S. 2006. eyePatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, 149–154.