Graphics Interface Conference 2015, 3–5 June, Halifax, Nova Scotia, Canada

AOI Transition Trees

Kuno Kurzhals and Daniel Weiskopf
University of Stuttgart, Germany

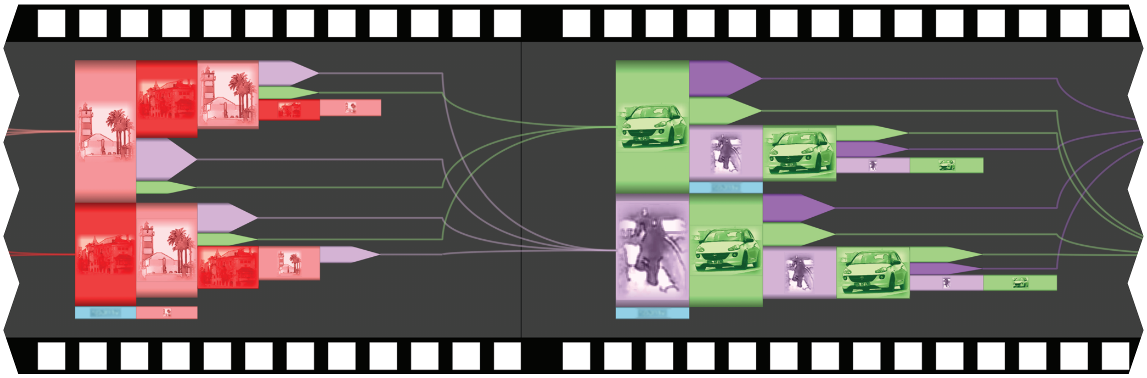

Figure 1: Transition trees of eye tracking data from multiple participants over two consecutive video shots taken from a long video sequence. Frequent sequential patterns of visits to areas of interest (AOIs) are displayed by extended icicle plots with AOI thumbnails. The correspondence of AOIs between the two shots is visualized by curved links that connect the transition trees of the two shots.

Abstract

The analysis of transitions between areas of interest (AOIs) in eye tracking data provides insight into visual reading strategies followed by participants. We present a new approach to investigate eye tracking data of multiple participants, recorded from video stimuli. Our new transition trees summarize sequence patterns of all participants over complete videos. Shot boundary information from the video is used to divide the dynamic eye tracking information into time spans of similar semantics. AOI transitions within such a time span are modeled as a tree and visualized by an extended icicle plot that shows transition patterns and frequencies of transitions. Thumbnails represent AOIs in the visualization and allow for an interpretation of AOIs and transitions between them without detailed knowledge of the video stimulus. A sequence of several shots is visualized by connecting the respective icicle plots with curved links that indicate the correspondence of AOIs. We compare the technique with other approaches that visualize AOI transitions. With our approach, common transition patterns in eye tracking data recorded for several participants can be identified easily. In our use case, we demonstrate the scalability of our approach concerning the number of participants and investigate a video data set with the transition tree visualization.

Index Terms: Human-centered computing [Visualization]: Visualization techniques—

1 Introduction

It is common practice to analyze human viewing behavior by eye tracking in scientific research and marketing analysis. Recorded eye gaze data can lead to a deep understanding of human cognitive processes [8], and can be applied to a wide range of research questions. The analysis of eye gaze data is generally performed using descriptive and inferential statistics [17] that support predefined hypotheses. To facilitate hypothesis building as well as statistical interpretation, visualization approaches can be applied. In general, visualization and visual analytics [3] have been gaining relevance for the analysis of eye tracking data because they can provide other insight into the data than the traditional statistical methods.

Recently, eye tracking has been applied more and more frequently to videos. Therefore, there is an increasing need for analysis and visualization techniques for eye tracking data recorded from dynamic stimuli [4]. For these stimuli, we have to distinguish between individually created stimulus content and content watched passively. Individual content is usually generated by interactions of participants with the environment. Passively watched stimuli comprise videos that are presented to the participants (e.g., movies [11]). Although the content of these stimuli is dynamic, recorded eye tracking data can be synchronized since no individual content is created and the length of a presented stimulus is constant.

Common visualization techniques in most of the analysis software packages from known eye tracking vendors include attention heat maps [41] and gaze plots or related scanpath visualizations [9]. Although attention maps and scanpath visualizations can be applied to dynamic stimuli, numerous analysis tasks—including the ones listed below—are hard to solve with these techniques. Nevertheless, attention maps provide valuable information about the spatial distribution of attention by performing kernel density estimation [29] on measured data points. They help identify areas of interest (AOIs) but cannot be used for an analysis of transition patterns between AOIs.

With AOI information, a scanpath can be represented as a string, consisting of different symbols that stand for respective AOIs. Single symbols can either represent a fixation on an AOI, a sampled gaze point, or just the AOI itself in a transition sequence. A sequential comparison of such sequences can provide insight into cognitive strategies [10]. There are automatic approaches to such analysis [39], but these techniques often bear problems with the visual interpretation of the results. A visually more accessible approach is the definition of AOI-based scanpaths by a hierarchical tree structure [34].

With a focus on dynamic stimuli and eye tracking data from multiple participants, we introduce a new approach to visualizing transition sequences: Our *AOI transition trees* employ icicle plots to represent AOI sequences modeled as trees. Shot boundary information is used to split the eye tracking data for the full stimulus into coherent timespans, leading to a good scalability with the number of AOIs to be displayed. Thumbnails are included as pictorial representations of annotated AOIs. With our approach, common patterns in transition sequences from multiple participants become visible, independently from their temporal order, and a quantitative interpretation of sequence frequencies is possible. Figure 1 shows an example of the AOI transition trees (see Section 3 for details). Although our approach is specified for dynamic stimuli, it is not only restricted to it and can be applied to any eye tracking data with only minor modifications.

Since current approaches for the analysis of scanpaths lack an effective representation of sequential scanning sequences [9] and it becomes more important for a strategy analysis "how" the AOIs are swept than "what" AOIs were visited [23], we aim to answer the following analysis questions related to a sequential analysis:

- *Which AOIs are frequently visited?*
- *What are frequent transition sequences between AOIs?*
- *How long are common transition sequences?*
- *What are the transition frequencies?*

Our main contribution is a summarizing visualization of all AOI transition sequences of variable length in a video that includes shot length information as well as a linking between transition trees of different shots. We enable the analyst to interpret the frequency of transition sequences of arbitrary length. Due to the scalability of our approach, even large numbers of participants can be summarized. With the representation of AOIs by thumbnails, we allow for the semantic interpretation of the stimulus without text labels.

2 Related Work

Considering the hierarchical structure of sequential visits on AOIs, generic visualization approaches for this type of data could be applied. An overview of such techniques is provided by Graham and Kennedy [12], Herman et al. [16], and Schulz et al. [27]. According to our visualization requirements (see Section 3.2), we focus on related work applying icicle plots to hierarchical/sequential data. Wongsuphasawat et al. [40] present an interactive icicle plot to display event sequences in hospital departments and millions of user action sequences from twitter. Trümper et al. [33] apply an icicle plot visualization similar to an AOI transition tree for the visualization of execution traces in software development. Telea and Auber [32] apply a cushioned icicle plot to visualize the evolution of source code. Two linked icicle plots can be applied to compare hierarchical structures such as folders in file systems [18]. A sunburst visualization [30] uses a circular layout of icicle plots and can be used to visualize hierarchical data in general. However, all of these approaches do not include multiple linked trees, they do not take advantage of the temporal division of the data, and are not designed to fit the changing information of AOIs in a video.

In eye tracking research, AOI-based scanpath visualizations that include absolute time provide information about the length of individual AOI visits. For static stimuli, Raschke et al. [24] represent scanpaths by parallel timelines for all AOIs and display the scanpath according to the visited AOIs over time, similar to a parallel coordinates plot. Burch et al. [5] show the temporal changes of AOI visits in a variant of the ThemeRiver [14] with additional flows for transitions between AOIs. Weibel et al. [38] use separate timelines for AOIs in a video stimulus and mark time spans when a participant looked at the AOIs. Another possibility is to use one timeline per participant and mark time spans according to different AOI colors [31]. This technique is also known as scarf plot [25]. Kurzhals et al. [20] use scarf plots to visualize scanpath similarity measures between multiple participants. All those techniques that visualize absolute time bear the problem that transition sequences that appear with different temporal extent or in another temporal order, cannot be detected efficiently. To overcome this problem, our AOI transition trees are designed to work with AOI subsequences that may have varying temporal length or position. To still retain the information about absolute time, we combine transition trees with a visualization that displays AOI-based scanpaths with absolute time (e.g., scarf plots) in multiple coordinated views. With the overview of common AOI transition sequences provided by our technique, corresponding time spans of selected sequences can easily be highlighted on the timeline of a scarf plot.

Tsang et al. [34] visualize fixation sequences with a Word Tree [37], using AOI text labels for sequences with a maximal length of 5 for dynamic stimuli. Our approach shares the same principal idea: sequences are represented by trees; branching into different AOIs along the timeline of the sequence corresponds to branching in the tree. However, there are several important differences as well. First, other than Tsang et al. [34], we visualize transitions between AOIs and not fixation sequences, to achieve a higher degree of data summarization. Second, we replace the Word Tree by an extended version of a space-filling icicle plot [19] that allows the integration of thumbnails for an intuitive mental linking between visualization and stimulus. With the icicles, quantitative assessment of transition frequencies is better supported than by the text font size in the Word Tree. Third, we introduce an overview representation of multiple transition trees based on shot boundaries, leading to better scalability with stimulus length and number of AOIs in the full stimulus. Additionally, Tsang et al. [34] and West et al. [39] focus on the analysis of scenarios that contained a static set of few AOIs. With our visualization approach, changing AOI constellations are handled with thumbnails and transition sequences of arbitrary length can be displayed.

Automatic approaches for the comparison of scanpath sequences are usually applied in two ways: Tools such as *ScanMatch* [6] provide information about the similarity of scanpaths; however, for many participants, the interpretation of the question *why* scanpaths are similar becomes problematic. The second approach is to identify common patterns automatically. In eye tracking research, this is often performed by the *Longest Common Subsequence (LCS)* [42] and *Sequential PAttern Mining (SPAM)* [15]. The results of these algorithms provide only information about the most frequent sequences. Less frequent subsequences that could be interesting to the analyst might be neglected. Also, the results of the algorithm still need to be interpreted by an analyst. With our approach, the visual interpretation of the most frequent sequences as well as the less frequent sequences becomes possible in a compact overview. Therefore, our visualization can be serve as an aid to support and interpret automatically extracted results, as well as an aid to derive new hypotheses from the recorded data that can further be analyzed by inferential statistics.

3 Visualization Technique

The focal point of this paper is the identification of frequently appearing transition patterns of variable length in eye tracking data recorded from multiple participants. In this section, we first describe the necessary data preprocessing for the annotation of AOIs (Section 3.1) and requirements for the visualization (Section 3.2). The visualization design is introduced by a general explanation of our approach (Section 3.3), combined with a comparison to other approaches for the visualization of transitions in eye tracking data. Section 3.4 describes how several transition trees can be composited to form a long sequence. Finally, we discuss the integration of AOI thumbnails into the transition trees (Section 3.5) and how interactions are handled (Section 3.6).

3.1 Data Preprocessing

Automatic approaches for the definition of AOIs (e.g., spatio-temporal clustering [21, 26]) can only provide information of important regions on AOIs, often not the complete silhouette of an object. If an AOI has to represent an object for semantic interpretation, the manual annotation of AOIs is often a necessary step of data preprocessing. In the case of synchronizable video stimuli with identical content, dynamic bounding regions can be defined around AOIs once, and gaze data can be mapped to the defined regions. For the annotation, we used *ISeeCube* [20], a software for the visual analysis of gaze data that includes an editor for the annotation of dynamic AOIs. Although we focus on videos with synchronizable content, our approach should also be applicable to stimuli with asynchronous video content and static stimuli (see Section 6).

3.2 Visualization Requirements

Depending on the analysis task, stimulus, and recorded eye tracking data, different requirements need to be met for an appropriate visualization of the data. In our case, the analysis task is to identify common transition patterns, the stimulus is an edited video, and the eye gaze data is from multiple participants.

We identify the following requirements and characteristics to be relevant for the visualization and analysis:

(R1) – Analysis of transition sequences and transition frequencies: The visualization needs to display AOI transition sequences, not just transitions between pairs of AOIs. The visual salience of important transitions and their frequencies should become accessible by the visualization.

(R2) – Subsequences of linear, ordinal scale time: The temporal aspect of the data in the case of transition analysis focuses on the ordinal time scale of visited AOIs, arranged along linear time [2]. The absolute time points are of lesser interest because we are interested in the transition patterns. Furthermore, identical patterns of linear transition subsequences in the data of multiple participants should become visible, regardless of their exact temporal position; i.e., subsequences of patterns should be identified anywhere along the timeline.

(R3) – Temporal division of the stimulus: We focus on the analysis of data recorded from dynamic stimuli, more precisely edited video stimuli. In contrast to unedited videos (e.g., from head-mounted eye tracking), edited material often contains intentional cuts that divide a video in scenes and shots that lead to abruptly changing AOI constellations over time. With these shot boundaries, a divide-and-conquer approach that splits the recorded data in semantic coherent sections can be applied. The advantage of this approach is that by dividing the data, consecutive transition sequences become shorter and therefore easier to interpret. In general, even unedited material can often be divided in parts of semantic coherence, e.g., by different tasks (see discussion in Section 6).

(R4) – Scalability: Since the visualization will be applied to video stimuli that can be divided, the scalability concerning the length of a video is not the critical aspect: video shots can be seen as individual units, and from the vast number of AOIs in a video, only those that exist in the current and the directly adjacent shots are important for the visualization of the transition tree. Scalability in our case concerns the number of recorded participants that have to be compared and the length of the transition sequences. Visualization techniques that display participants individually, (e.g., scarf plots) tend to become harder to interpret with an increasing number of participants. Therefore, the visualization needs an aggregated representation of the participants, independent from their number. Although the frequency of transition patterns decreases with an increasing sequential length, the visualization has to show transition patterns of variable length, until the patterns become unique.

(R5) – Semantic interpretation of AOIs: Video stimuli can contain a vast number of AOIs that appear at different time spans during the video. Mapping colors to AOIs is a common approach to make them distinguishable (e.g., [5, 34]). For a semantic interpretation of an AOI, additional labels are necessary. A visualization with text labels can be the best choice if only few AOIs exist and unambiguous labels can be given. In the case of edited video stimuli, however, a large number of AOIs can appear, making it tedious work to find an appropriate label name for every AOI.

To meet this requirements, we have developed an enhanced icicle plot visualization—the *AOI transition tree*—that displays the hierarchical structure of transition sequences (Section 3.3). The area of individual nodes can be utilized for the color coding (Section 3.3) and labeling with thumbnails (Section 3.5). The scalability of this visualization approach is improved by shot-based division of the stimulus (Section 3.4) and aggregation of sequence frequencies between participants (Section 3.3).

3.3 Visualization of Transition Sequences by Extended Icicle Plots

To visualize patterns of transitions between AOIs, a compression of the AOI-based scanpath strings is performed. The original string consists of symbols that represent visited AOIs, often composed of fixations driven by the sampling rate of the eye tracking device or the video. This leads to repeating symbols in the string and provides important information if the analyst is interested in how long individual participants visited an AOI. In that case, a scarf plot provides a good overview of the scanpath. For transition sequences, the analyst is interested in the order and frequency of AOI visits; thus, repeating symbols in the string have to be removed. Figure 2 illustrates the string reduction. The scarf plot shows a sequence

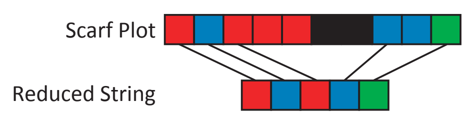

Figure 2: The original scanpath, represented by a scarf plot, is reduced by removing consecutive symbols of the same AOI. Black regions mark time spans with no AOI-relevant data. The resulting string is used for the analysis of transition patterns.

of AOI visits, sampled per frame with color symbols representing AOIs and black symbols for samples without AOI information. The reduced string represents only transitions between AOIs [39], independent from durations or multiple consecutive visits to the same AOI. Therefore, the string compression meets the linear, ordinal-scale time characteristics of requirement (**R2**).

Adopting Tsang et al. [34], we interpret the reduced AOI-based scanpath string as a tree. However, other than Tsang et al. [34], we convert any subsequence into the tree representation: regardless of when the subsequence occurs in the string, it is placed, beginning at the root of the tree. In this way, the requirement for subsequence analysis (**R2**) is met. In detail, transition sequences are represented by a multi-rooted tree, single nodes represent AOIs in a transition subsequence, the levels of the tree correspond to the length of a subsequence. In addition, nodes are enriched by a numerical attribute that represents the frequency of visits to the corresponding AOI. With this interpretation of the scanpath data, we face the problem of visualizing a tree of varying depth and with one numerical attribute; the attribute has the property that the sum of the children's attributes is equal or less than the value of the attribute of the node itself because there cannot be more visits to subsequent AOIs than to the current AOI of a sequence.

With this abstraction in mind, there are many potential visualization techniques (see Ward et al. [36] for a recent textbook presentation). We have chosen the icicle plot [19] from this list of candidate techniques because it best meets our requirements.

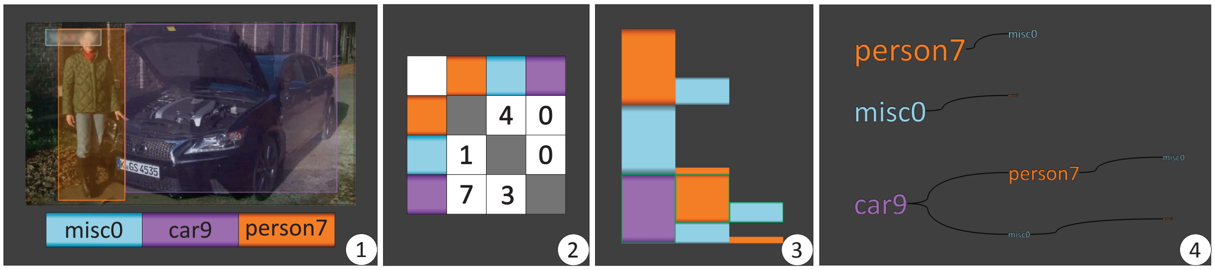

Figure 3: Transition visualization techniques for a video with three AOIs (1): transition matrix (2), AOI transition tree (3), and Word Tree (4).

Figure 3(3) shows an example of our visualization technique. The transition sequences are represented by an icicle plot with horizontal orientation, i.e., the time axis is along the standard left-to-right reading direction in English. Single nodes of AOIs are displayed by rectangular boxes in the icicle plot. The height of the box indicates the frequency of AOI visits. Data from several participants is easily aggregated by adding up transition frequencies for the respective icicle boxes. The boxes are sorted according to descending height. Finally, the boxes need to be visually associated with respective AOIs. We use a qualitative color scheme of 11 colors [13] to distinguish between the different AOIs. 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 resolved by looking at the AOI thumbnails (Section 3.5) and a video preview.

On the first level of the icicle plot, the height of the boxes is calculated by the total frequency an AOI appears in the sequence. With this approach, important AOIs that were frequently visited become visible. Figure 3(3) shows a transition tree with data from 16 participants that displays all subsequences of transitions between three AOIs in a video shot (see Figure 3(1)). Children of a node are ordered from top to bottom with decreasing frequency.

The interpretation of the transition tree in Figure 3(3) can be explained by traversing the icicle plot from left to right:

The *first level* of the tree (leftmost column of the transition tree) shows the distribution of attention to AOIs, aggregated from the full sequence. It can be interpreted as a vertically stacked histogram.

On the *second level*, transitions between two AOIs are displayed. The second level of a transition tree shows the same information as a transition matrix (see Figure 3(2)), representing the frequency of transitions between two AOIs. In the transition tree, the frequency can be read off from the height of the box in the second level.

Starting with *third level*, the power of the transition tree representation becomes clear: Sequences are interpreted identical to the second level, by traversing the transition tree from left to right. Other than with the transition matrix, sequences of arbitrary length can be identified efficiently. Since all appearing subsequences are displayed, patterns are possible to appear in other branches of the tree (see Figure 5), showing which AOIs where visited before the sequence started.

Our visualization design meets requirement (**R1**): it shows both the structure of the transition subsequences (by the icicle layout, along with color patterns) and the transition frequencies (by encoding by the height of the boxes). Through the height encoding, visually salient boxes are produced for important subsequences and, thus, important transitions are well represented. Furthermore, requirement (**R4**) is met because the aggregated visualization of transition frequencies allows for a summarization of data from an arbitrary number of participants.

We have chosen the icicle approach over other tree visualization candidates. One reason is the need for effective encoding of quantitative data. Height is an excellent visual representation for such data; according to Mackinlay [22], length is the highest ranked visual representation for quantitative data after position, and the latter is not available in our visual design because we also have to show the tree structures that needs spatial position. A related benefit of icicle plots is that the heights of child node representations add up: the additive structure of the transition-frequency attribute has a one-to-one visual correspondence to the spatial partitioning of the box' heights. The tree-map [28] is the other tree visualization technique that comes with an intrinsic additive structure. However, the tree-map does not separately show the quantitative value for internal nodes, which is important for our application because we want to highlight high frequencies of AOI visits in early stages of the sequence as well. Furthermore, paths from the root nodes downward the tree are harder to see and, thus, the tree-map is less suited for the visualization of AOI sequences. Finally, node-link type of diagrams were not chosen because they do not represent quantitative attributes and their intrinsic summation characteristics as well as icicle plots.

Our other design choice is to use color hue for associating icicle boxes to AOI labels. We chose color hue because it is excellent for showing nominal data [22]. In particular, it can be used to identify sequence patterns. One issue with categorical color coding is the restriction to a few easily distinguishable colors. Therefore, there are limitations in scalability with the number of AOIs. However, this scalability problem is quite uncommon in practical applications because we eventually split the stimulus into shorter time spans of coherent contents that typically contain just a few AOIs (Section 3.4). One shortcoming with color, however, is that it does not directly link to the contents or semantics of the AOIs. This problem is addressed by including overlay thumbnails of the AOI in the icicle box (Section 3.5).

The Word Tree approach [34] shares some of our design decisions, e.g., the display of the tree representation of AOI sequences and the ordering along a horizontal timeline. Figure 3(4) shows the Word Tree for our example data. In comparison to the AOI transition tree in Figure 3(3), the frequencies are better visualized by box size than font size. Through the text representation, more horizontal space is needed in Word Trees, restricting the depth of transitions sequences to be displayed. Moreover, icicle boxes may be scaled down to pixel size, yielding better vertical scalability. Finally, color patterns are more easily identified in the screen-filling icicle layout than in the sparse text layout of Word Trees. The main advantage of the Word Tree representation is the semantic labeling of nodes by text: the analyst can virtually read along the Word Tree, which is useful for detailed analysis of the semantics of paths through the AOI tree. In contrast, AOI transition trees are better suited for larger data sets and the analysis of transition frequencies.

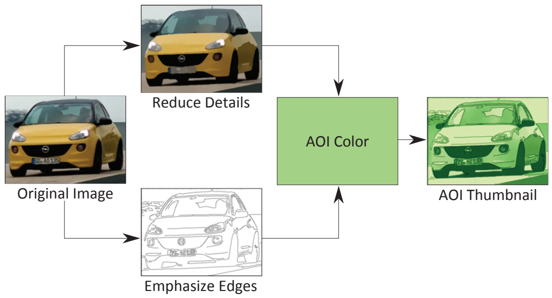

Figure 4: AOI thumbnail creation: the image is processed by multiple filtering steps, resulting in an abstracted representation of the AOI.

3.4 Sequence of AOI Transition Trees

So far, the transition trees are applied to a defined time span in a video. If the time span includes several shots in a video, the first level of the transition tree will include more AOIs that appeared in the video and individual scanpaths become visible with increasing length of transition subsequence. The temporal division of the stimulus (**R3**) allows for an approach that creates a sequence of smaller transition trees, instead of just a single, very large transition tree for the complete video. Shot boundary information provides a semantically motivated approach to divide the stimulus in separate time spans: the transition trees are created for individual shots of a video. Although we use shot boundaries to split the timeline, it should be noted that any other time division (e.g., task-related division, see also discussion in Section 6) is possible as long as it leads to an AOI-coherent division.

With this division, the visualization is extended from a single AOI transition tree to a *sequence of AOI transition trees*. Figure 1 shows an example of a sequence of transition trees. For an individual AOI tree, absolute time is not considered. Nevertheless, additional information about the duration of a shot is important to find out if long transition sequences in a tree result from a long shot, or from a dissimilar viewing behavior of the participants. Therefore, we apply a film strip metaphor to facilitate a qualitative assessment of the length of a shot. Film strips that represent the video shots are concatenated horizontally, forming a horizontal timeline summarization of the complete video stimulus. Logarithmic scaling is applied to ensure that transition trees fit even in short shots. The transition trees are then positioned on the film strip in the corresponding shot.

To connect the AOIs of two consecutive trees, all transition sequences are extended by an additional level of AOIs after the end of a sequence. These additional AOIs are the next elements in the transition subsequences that continue in the consecutive shot. Here, an arrow shape was applied, to emphasize the transition to the next shot. If a sequence was shortened due to filtering, no additional AOI is added since the shortened sequence is not continued in the next shot. Finally, connection lines are drawn between the corresponding AOI boxes. We chose curved connection lines to make them visually different from the graphical elements in the individual AOI transition trees, which only contain horizontal and vertical straight lines.

3.5 AOI Thumbnails

To this point, the transition trees consist of boxes with individual colors that represent the AOIs. To meet requirement (**R5**), a semantic interpretation of the AOIs has to be facilitated. Labels are a good way of building the link between the icicle box and the corresponding AOI. Text labels can provide detailed information about an AOI, but assigning meaningful text labels becomes tedious with

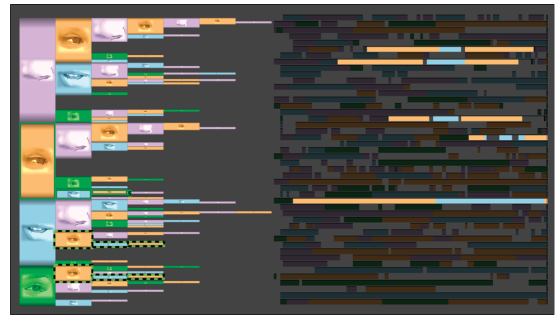

Figure 5: Sequences can be selected in the transition tree and their frequency is displayed (left, solid green border). They can appear as subsequence elsewhere (left, dashed green border). Since a visual search for this sequence in the scarf plots of multiple participants (right) requires much effort, the selected sequences are highlighted.

increasing number of AOIs and often depends on subjective interpretations of AOIs. Furthermore, text labels require relatively wide (horizontal) space.

As an alternative, we recommend using a pictorial "label". To this end, we introduce *AOI thumbnails*: small images that show the object of the AOI in an abstracted representation, and that preserve the color assigned to the icicle box of the AOI. The AOI thumbnail is placed inside the icicle box to illustrate the AOI's object. We chose an abstracted, non-photorealistic representation because that can be made readable even when shown as small picture. We use a schematic representation with enhanced feature lines, adjusted lightness contrast, less image detail, and color modification. The color is changed so that it matches the hue of the icicle box to maintain the color patterns of the AOI transition tree. With this approach, labels are less dependent on subjective annotations of the labeling person, and interpretations of the AOIs that are involved in a timespan become simpler, even without knowing the stimulus.

Figure 4 illustrates the image processing steps required to create an AOI thumbnail. First, a valid frame from the life span of an AOI is chosen. By mean shift filtering, image details are reduced and compositing with the AOI color can be performed on areas with a consistent lightness. We emphasize important edges in the resulting image to provide the analyst with enough structural information of an object for its recognition. To this end, Canny edge filtering is performed on the gray-scale version of the original image. For final compositing, the resulting images from mean shift filtering and Canny edge filtering are combined, taking into account the color of the AOI. Image compositing is performed in the perceptually linear CIE L*a*b* color space. The final image is obtained by compositing the lightness of the images and the AOI color. The resulting thumbnail is finally inserted into the corresponding boxes of an AOI in the transition tree.

3.6 Interaction

We integrated the visualization into the *ISeeCube* [20] system, a visual analytics tool for eye tracking data, consisting of multiple coordinated views with different visualizations for various analysis tasks. Hence, new possibilities for interacting with the visualization become possible.

Brushing and linking: The transition tree provides an efficient approach to identify frequently appearing sequences, but without taking their temporal position into account. In contrast, scarf plots provide this information, but are less suited to identify identical sequences in the viewers' scanpaths. By brushing and linking, the selected transitions from the trees become visible in the scarf plot visualization, combining the advantages of both techniques (see Figure 5). Vice-versa, a timespan can be selected in the scarf plots

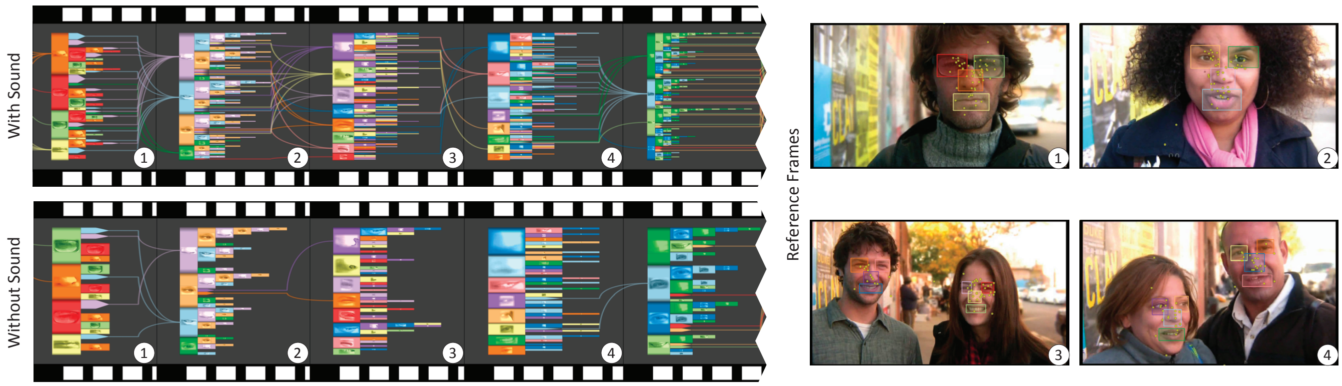

Figure 6: Video (50 Faces Brooklyn [1]): The AOI transition trees show the eye tracking data recorded for participants that watched the video with sound (top left) and without sound (bottom left). The faces of interviewed people in four consecutive shots of the video were annotated with dynamic AOIs for eyes, nose, and mouth (right). (Source: DIEM [7], original video images from "50 People One Question" [1].)

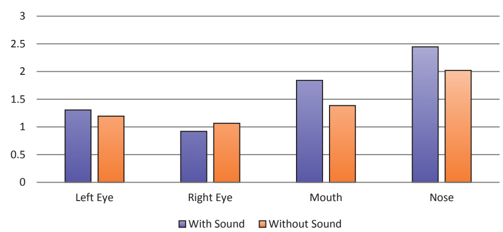

Figure 7: Average number of visits on different face regions over the whole dataset with two conditions: with and without sound.

and the corresponding, scalable transition tree is created below the filmstrips for better readability.

Filtering: To improve the scalability of our approach considering many AOIs and long transition sequences of individual viewers, we allow the user to filter the transition trees by setting the minimal transition frequency. Less important sequences can be removed and therefore, visual overload is reduced. By removing parts of the sequences, the tree is adjusted to fill out the free space to represent bigger AOI thumbnails.

4 Use Case

To show the applicability of our approach to typical eye tracking data for video stimuli, we use an example of the Dynamic Images and Eye Movements (DIEM) database [7]. The database contains stimuli from a variety of different genres. Gaze data recorded from multiple participants watching the stimuli is included.

Video: 50 Faces Brooklyn

This video is a sequence from the "50 People One Question" Brooklyn video [1], and the eye tracking data is from the DIEM database. The video shows various people interviewed to answer the question: *"Where do you want to wake up tomorrow?"* In a previous experiment, Võ et al. [35] showed the video to multiple participants to analyze the dynamic allocation of attention on moving faces (Figure 6). Two experimental conditions were applied to the stimulus: one half of the participants watched the video with sound, the other half without sound. The authors investigated the overall gaze distribution on the faces. Their results show a significantly lower percentage of the gaze distribution on the mouth when no sound was present, compared to the scenario with sound; sound had no effect on the relative gaze distribution on the eyes and nose.

We annotated the dynamic AOIs similar to the ones described by Võ et al.—with the difference that we annotated the two eyes of a person separately, whereas Võ et al. used a single AOI for a pair of eyes. Figure 7 displays the average number of visits of all participants on the annotated face regions, with the nose and mouth regions as the most frequently visited areas. The high number of visits on the nose regions results partially from their central position, resulting in visits when participants switch between regions by crossing the nose. Four examples of transition trees and corresponding images from the video are shown in Figure 6. For the sequential scanpath analysis, the questions mentioned in Section 1 become important at this point. We can answer the questions by interpreting the visualization:

- *Which AOIs are frequently visited?* In all shots except shot (1) without sound, the AOIs with noses of people were most frequently watched, since they are on top of each transition tree, independent from the sound condition. In the shots (2)–(4) the AOIs with mouth regions are on higher ranks for the sound condition, and at least one rank lower when no sound was played. This supports the results of Võ et al., indicating that not only the total distribution of attention on mouth regions was higher when sound was played, but also the number of visits on this region (see Figure 7). Notice that in shot (1), the mouth is on the last rank for both conditions; this results from the fact that the person in this shot was looking into the camera without talking.

The above question could also be answered with individual histograms for each shot. This information is represented by the first level of the transition tree. For answering the following questions, the histogram (Figure 7) cannot be applied, but the transition tree (Figure 6) needs to be used:

- *What are frequent transition sequences between AOIs?* Since AOIs with noses were most frequently visited (Figure 7), we can investigate how the transition trees branch when a sequence starts with a nose: With sound, the shots (2)–(4) show the most frequent transitions from the nose to the mouth and back. Without sound, the same can be noticed for the shots (3) and (4), but in shot (2), the eye was more frequently visited than the nose. With a transition matrix, we could identify that many transitions between nose and mouth regions appear, but such longer sequences would not be visible.
- *How long are common transition sequences?* By filtering out patterns that appear less than five times, long and individual

transition sequences disappear and the remaining sequences range from length 3 to length 7.

- *What are the transition frequencies?* By selecting an interesting transition sequence, the information about its frequency is displayed in the visualization. To see which scanpaths show these patterns and what their temporal context is, the linking between the transition trees and the scarf plot visualization is available (see Figure 5).

We can answer all these questions with one visualization approach. Except for the first question, a tedious analysis of all participants by scarf plots or video analysis would have been required to answer all questions. By including the interaction with the scarf plot visualization (see Figure 5), we can even show when and where selected sequences appear in the viewers' scanpaths. The last two questions could be answered by automatic analysis methods, too. However, such techniques would only result in a subset of the displayed sequences that still requires visual interpretation. With our approach, all sequences can be investigated in an overview while details on demand are still available for the interpretation.

5 Expert User Feedback

In separate sessions, we introduced the visualization to three visualization experts, all of them with advanced knowledge of eye tracking analysis. Each session took between 30–45 minutes. AOI transition trees were presented as static images on the screen that could be zoomed and panned by the experts.

First, a training data set was presented, explaining the structure of the transition trees and showing example patterns. Examples included the interpretation of frequently appearing AOIs in sequences, as well as explanations of how the links between trees can be interpreted. Here, the data set shown in Figure 1 was presented. This data set was already known to the participants for a better understanding of the new visualization.

Finally, the transition sequence of the data set from Section 4 was presented and the experts were asked to interpret the recorded eye tracking data, based only on the transition trees. No expert had prior knowledge of the video stimulus and the recorded data. Their analysis process was accompanied by the think-aloud method and their interpretations as well as comments on the visualization were noted.

The identification of frequently visited AOIs, and important transition sequences was achieved with the test data set. The experts were able to interpret the visualization and could identify patterns in the data. The most common findings, as described in Section 4, were the importance of the noses, followed by mouth regions, and the back switching to AOIs. Overall, the experts assessed the AOI transition trees as useful and promising alternative for the analysis for gaze patterns.

However, with the think-aloud method, limitations of the AOI thumbnails were identified as well: while a semantic interpretation of AOI thumbnails was performed without further problems, the interpretation of unknown AOIs became difficult for AOIs for which only few details were available in the original image. Especially the differences between eyes and mouths were hard to see in some shots of the video. As an improvement of this approach, the experts recommended that additional information such as a text label, and a representative image of the shot with AOIs marked on it (similar to Figure 6) may help in such situations. As a consequence, we included shot reference images below the filmstrips to help the user get oriented in the dataset. In a separate view, the video can be played back with annotated AOIs to help, if thumbnail pictures are difficult to interpret.

One additional problem that could be identified from the expert user feedback, is the interpretation of the first level in a transition tree. During the exploration of the data set, the AOIs in the first level of the transition tree were misinterpreted as representing all AOIs the shot begins with. After mentioning their misinterpretation once, the experts read the transition trees correctly. The film strips were only regarded as markers for separate shots, the interpretation of shot lengths was not considered by the experts.

6 Discussion

Since icicle plots are an established visualization technique, the interpretation of transition trees can be learned quite easily. Nevertheless, the qualitative expert feedback showed that misinterpretations of the root level can appear. Therefore, a training period is required to read the subsequence representation of transition trees correctly. The connection lines between transition trees could be interpreted without further problems.

As mentioned in Section 5, the application of AOI thumbnails facilitates good recognition of objects by the analysts if the original image provides enough detail for the generation of important feature lines. AOIs with little detail (e.g., due to a blurred image) and AOIs that cover only subregions of an object are difficult to interpret without the original stimulus. Since the interpretation of AOI thumbnails can be improved if further information of the stimulus itself is provided, an integration of the transition tree visualization in a system with multiple coordinated views enhances that readability. In combination with AOI attention histograms, scarf plots, video preview, and space-time cube visualization, provided by the *ISeeCube* visual analytics system, the interpretation of transition sequences can be supported by different views on the data.

So far, the presented approach has been tested on video stimuli with participants watching them passively. For actively created content (e.g., individual recordings with a head-mounted eye tracker), an application is also conceivable but would require additional annotation work: since each participant generates a unique video with gaze data, all video streams have to be annotated individually and semantically identical AOIs have to be linked. Due to the nature of the stimulus, shot boundaries are not available to create reasonable sub-trees. Nevertheless, other approaches could be applied to divide the stimulus in separate partitions: For example, in a shopping scenario where the participants buy different articles, a separation could be achieved by defining a boundary every time the participants pass a row of shelves. We plan to assess our approach for actively created content in future work.

Due to the color scheme that prioritizes the qualitative discriminability between AOIs that appear simultaneously, some colors can represent multiple AOIs. This happens if many persistent AOIs appear and reappear during the whole video sequence (e.g., a soccer game), or if long time spans are summarized with one transition tree. By including the AOI thumbnails into the visualization, we prevent ambiguities that may appear if only color would represent an AOI. In future work, additional information about AOIs could be shown on demand (e.g., by tooltips) to further improve the interpretation of sequences.

The scalability of our approach considering the number of possible AOIs is limited in the overview due to the restriction of space by the filmstrips. Since the space is divided between all AOIs on the first level of the tree, long sequences can become too small for a comfortable interpretation of AOI thumbnails. In the filmstrip overview, the most frequently appearing transition sequences receive more drawing space, reducing details in less frequent sequences. Therefore, the user can select a time span and create a new transition tree without size limitations on the canvas below the film strips where panning and zooming on these less prominent sequences is also possible.

The scalability of our approach considering the number of participants is not limited since all sequences become aggregated. With an increasing number of participants, more individual patterns will appear that can be filtered out to focus on common sequences.

7 Conclusion and Future Work

We presented a new approach for the exploration of AOI transition patterns in eye tracking data recorded from multiple participants while watching video. The presented transition trees provide a summarization of recorded data sets with semantic enrichment by AOI thumbnails. By applying our approach to various video stimuli, we showed that frequently appearing patterns could easily be identified even with a vast number of participants.

For future work, we plan to apply our approach to a wide range of eye tracking stimuli, including mobile eye-tracking data and visualization techniques. The latter case is of special interest, because in many visualization approaches, the definition of AOIs can only be performed rather on regions than on objects, which can lead to problems with the semantic interpretation of thumbnails. Also, we plan a comprehensive user study to compare our approach in detail with the others discussed in this paper.

Acknowledgements

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

References

[1] 50 People One Question. http://fiftypeopleonequestion.com, 2013.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, London, 2011.

[3] G. Andrienko, N. Andrienko, M. Burch, and D. Weiskopf. Visual analytics methodology for eye movement studies. *IEEE Trans. on Vis. and Comp. Graph.*, 18(12):2889–2898, 2012.

[4] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In R. Borgo, R. Maciejewski, and I. Viola, editors, *Proc. of EuroVis 2014 State-of-the-Art Reports*, 2014.

[5] M. Burch, A. Kull, and D. Weiskopf. AOI rivers for visualizing dynamic eye gaze frequencies. *Computer Graphics Forum*, 32(3):281–290, 2013.

[6] F. Cristino, S. Matht, J. Theeuwes, and I. Gilchrist. ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, 42(3):692–700, 2010.

[7] Dynamic images and eye movements (DIEM) database. http://thediemproject.wordpress.com, 2013.

[8] A. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Res. Meth., Instrum., & Comp.*, 34(4):455–470, 2002.

[9] J. H. Goldberg and J. I. Helfman. Visual scanpath represenation. In *Proc. of the Symp. on ETRA*, pages 203–210, 2010.

[10] J. H. Goldberg and X. P. Kotval. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24:631–645, 1999.

[11] R. Goldstein, R. Woods, and E. Peli. Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, 37(7):957–964, 2007.

[12] M. Graham and J. Kennedy. A survey of multiple tree visualisation. *Information Visualization*, 9(4):235–252, 2010.

[13] M. Harrower and C. Brewer. ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[14] S. Havre, B. Hetzler, and L. Nowell. ThemeRiver: visualizing theme changes over time. In *Proc. of the IEEE Symp. on Info. Vis.*, pages 115–123, 2000.

[15] P. Hejmady and N. H. Narayanan. Visual attention patterns during program debugging with an IDE. In *Proc. of the Symp. on ETRA*, pages 197–200, 2012.

[16] I. Herman, G. Melancon, and M. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Trans. on Vis. and Comp. Graph.*, 6(1):24–43, 2000.

[17] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, New York, 2011.

[18] D. Holten and J. J. van Wijk. Visual comparison of hierarchically organized data. *Computer Graphics Forum*, 27(3):759–766, 2008.

[19] J. B. Kruskal and J. M. Landwehr. Icicle plots: Better displays for hierarchical clustering. *The American Statistician*, 37(2):162–168, 1983.

[20] K. Kurzhals, F. Heimerl, and D. Weiskopf. ISeeCube: visual analysis of gaze data for video. In *Proc. of the Symp. on ETRA*, pages 43–50, 2014.

[21] K. Kurzhals and D. Weiskopf. Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Trans. on Vis. and Comp. Graph.*, 19(12):2129–2138, 2013.

[22] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. on Graphics*, 5(2):110–141, 1986.

[23] V. Ponsoda, D. Scott, and J. Findlay. A probability vector and transition matrix analysis of eye movement during visual search. *Acta Psychologica*, 88(2):167–185, 1995.

[24] M. Raschke, X. Chen, and T. Ertl. Parallel scan-path visualization. In *Proc. of the Symp. on ETRA*, pages 165–168, 2012.

[25] D. C. Richardson and R. Dale. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6):1045–1060, 2005.

[26] A. Santella and D. DeCarlo. Robust clustering of eye movement recordings for quantification of visual interest. In *Proc. of the Symp. on ETRA*, pages 27–34, 2004.

[27] H. Schulz, S. Hadlak, and H. Schumann. The design space of implicit hierarchy visualization: A survey. *IEEE Trans. on Vis. and Comp. Graph.*, 17(4):393–411, 2011.

[28] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. on Graphics*, 11(1):92–99, 1992.

[29] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, New York, 1986.

[30] J. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *IEEE Symp. on Info. Vis.*, pages 57–65, 2000.

[31] S. Stellmach, L. Nacke, and R. Dachselt. Advanced gaze visualizations for three-dimensional virtual environments. In *Proc. of the Symp. on ETRA*, pages 109–112, 2010.

[32] A. Telea and D. Auber. Code flows: Visualizing structural evolution of source code. *Computer Graphics Forum*, 27(3):831–838, 2008.

[33] J. Trümper, J. Döllner, and A. Telea. Multiscale visual comparison of execution traces. In *Proc. of the 21st International Conference on Program Comprehension*, pages 53–62, 2013.

[34] H. Y. Tsang, M. Tory, and C. Swindells. eSeeTrack: Visualizing sequential fixation patterns. *IEEE Trans. on Vis. and Comp. Graph.*, 16(6):953–962, 2010.

[35] M. L.-H. Võ, T. J. Smith, P. K. Mital, and J. M. Henderson. Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, 12(13):3, 1–14, 2012.

[36] M. O. Ward, G. Grinstein, and D. Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. A K Peters/CRC Press, Natick, 2010.

[37] M. Wattenberg and F. Viégas. The word tree, an interactive visual concordance. *IEEE Trans. on Vis. and Comp. Graph.*, 14(6):1221–1228, 2008.

[38] N. Weibel, A. Fouse, C. Emmenegger, S. Kimmich, and E. Hutchins. Let's look at the cockpit: Exploring mobile eye-tracking for observational research on the flight deck. In *Proc. of the Symp. on ETRA*, pages 107–114, 2012.

[39] J. M. West, A. R. Haake, E. P. Rozanski, and K. S. Karn. eyePatterns: Software for identifying patterns and similarities across fixation sequences. In *Proc. of the Symp. on ETRA*, pages 149–154, 2006.

[40] K. Wongsuphasawat and J. Lin. Using visualizations to monitor changes and harvest insights from a global-scale logging infrastructure at twitter. In *Proc. of the IEEE Conference on Visual Analytics Science and Technology (VAST) 2014*, 2014.

[41] D. S. Wooding. Fixation maps: Quantifying eye-movement traces. In *Proc. of the Symp. on ETRA*, pages 31–36, 2002.

[42] Y. Yesilada, S. Harper, and S. Eraslan. Experiential transcoding: An eyetracking approach. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pages 30:1–30:4, 2013.