

Fixation-Image Charts

Kuno Kurzhals, Marcel Hlawatsch, Michael Burch, Daniel Weiskopf
University of Stuttgart*

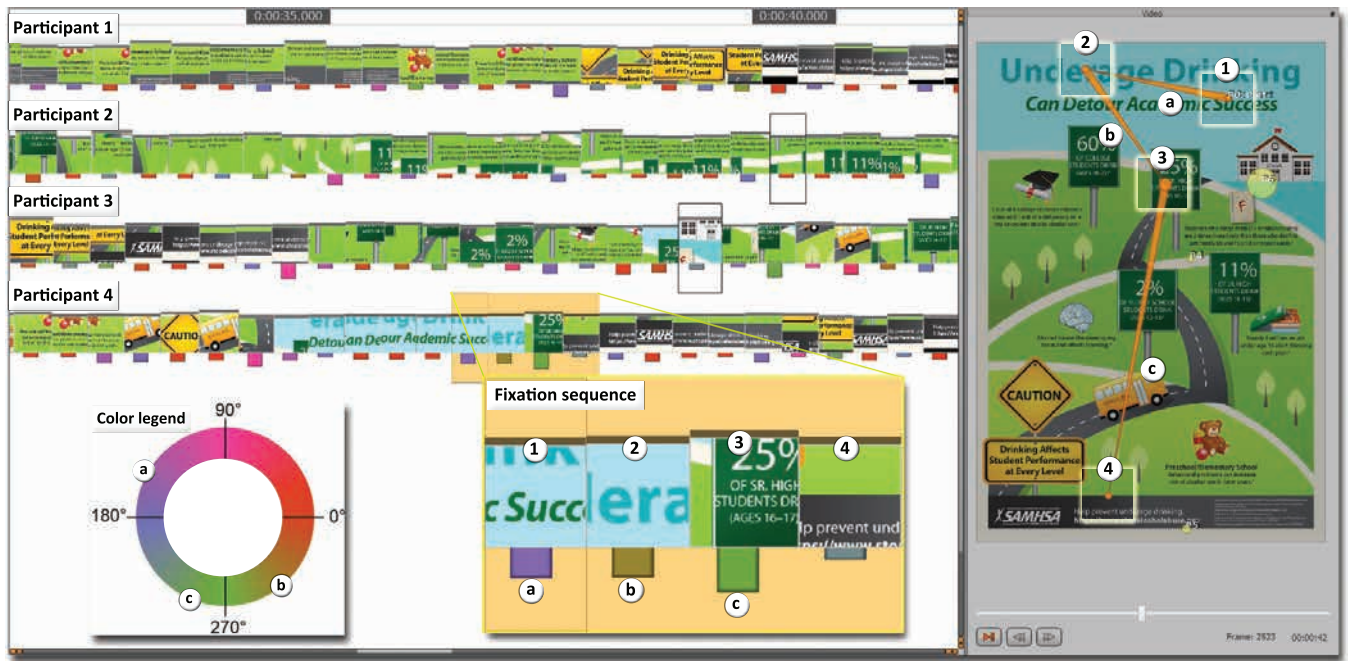


Figure 1: Fixation-image charts displaying the fixation sequences of four participants on timelines with information about fixation duration (height of the images), saccade lengths (length of colored bars), and saccade directions (according to the color legend). A subsequence consisting of 4 fixations (1–4) and 3 saccades (a–c) is highlighted in the closeup and depicted in the stimulus on the right.

Abstract

We facilitate the comparative visual analysis of eye tracking data from multiple participants with a visualization that represents the temporal changes of viewing behavior. Common approaches to visually analyze eye tracking data either occlude or ignore the underlying visual stimulus, impairing the interpretation of displayed measures. We introduce fixation-image charts: a new technique to display the temporal changes of fixations in the context of the stimulus without visual overlap between participants. Fixation durations, the distance and direction of saccades between consecutive fixations, as well as the stimulus context can be interpreted in one visual representation. Our technique is not limited to static stimuli, but can be applied to dynamic stimuli as well. Using fixation metrics and the visual similarity of stimulus regions, we complement our visualization technique with an interactive filter concept that allows for the identification of interesting fixation sequences without the time-consuming annotation of areas of interest. We demonstrate how our technique can be applied to different types of stimuli to perform a range of analysis tasks; and discuss advantages and shortcomings derived from a preliminary user study.

Keywords: Eye tracking visualization, static and dynamic stimuli, scanpath, fixation

Concepts: •Human-centered computing → Visual analytics; Visualization systems and tools;

1 Introduction

With the increasing popularity of eye tracking in various research fields [Duchowski 2002], many experiments are conducted and analyzed with established statistical and visual analysis techniques. Typically, statistical evaluation is applied for hypothesis testing, potentially showing significant differences between independent variables. Visualization is employed to convey and illustrate these results. Recently, visual analytics approaches [Andrienko et al. 2012; Blaschek et al. 2016] have become popular. They combine statistical methods and interactive visualization to enable the exploration of eye tracking data (e.g., from a pilot study) and support the understanding of the complex spatiotemporal nature of the data.

The general analysis problem we address is the interpretation of scanpaths in space and time. A scanpath, consisting of fixation sequences and saccades between the fixations, can be very complex, even for a single participant. The comparison of scan-

sonal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ETRA '16, March 14 - 17, 2016, Charleston, SC, USA

ISBN: 978-1-4503-4125-7/16/03

DOI: <http://dx.doi.org/10.1145/2857491.2857507>

*E-mail: name.surname@visus.uni-stuttgart.de

Permission to make digital or hard copies of all or part of this work for per-

paths from multiple participants is even more difficult. To address this issue, many approaches exist, some based on eye tracking metrics [Holmqvist et al. 2011], string-comparison algorithms [Duchowski et al. 2010; West et al. 2006], and visualization [Goldberg and Helfman 2010]. However, many established metrics, such as average fixation durations or average scanpath lengths, aggregate the data, neglecting its spatiotemporal structure. These approaches impair analysis tasks for which it is important to find out how viewing behavior, represented by fixations and saccades, changed over time.

If the temporal dimension of the underlying eye tracking data is included in the analysis, a visual representation of the measured data typically supports their interpretation. A simple and common approach is a plot of a measured variable over time (see Figure 2a). This visualization provides valuable insight into how the variable changed over time and multiple participants can be compared by overlaying their respective plots. Even string-based algorithms that calculate the similarity between scanpaths benefit from a visual representation of similar subsequences. However, gaze behavior is largely dependent on the visual stimulus. Without including the context of the stimulus, we will not be able to understand the reasons why certain changes in recorded measures occur. To address this issue, researchers usually have to investigate the recorded data as static and animated representation, overlaid on the stimulus. A typical example of such a visualization is the gaze plot (Figure 2b). Here, gaze points or fixations are represented by connected circles that maintain the temporal order of the scanpath. The radii indicate fixation durations. These circles are drawn on the stimulus, indicating where a participant was looking at which point in time. This approach suffers from occlusion due to the massive overdraw that occurs when many participants have to be compared over long timespans.

Other approaches such as attention maps aggregate the data and are therefore not suited for a comparative spatiotemporal analysis of multiple participants. If areas of interest are available, alternative timeline visualizations such as scarf plots [Richardson and Dale 2005; Kurzhals et al. 2014b] can provide a semantic scanpath representation of the sequential visits in the annotated areas. Especially for dynamic stimuli, this annotation is often a time-consuming pre-processing step requiring the definition of bounding shapes that can also change their position, size, and visibility.

Our main contribution is a visual analytics approach that combines descriptive statistics of fixation metrics, interactive filtering techniques, and a visual representation of fixation sequences. With our interactive visualization technique (Figure 1), we provide an alternative approach for scanpath comparison that retains the advantages of statistical and gaze plots. A detailed analysis of scanpaths from multiple participants over time is possible without the tedious annotation of areas of interest. For example, our approach can be applied for efficient labeling of fixations (Section 4.1) and for hypothesis building by investigating patterns in the visualization (Section 4.2).

2 Related Work

The number of new visualization techniques for eye tracking analysis increased significantly over the last years. Blaschek et al. [2014] provide an overview and a taxonomy of state-of-art techniques. The visualization methods can be categorized in two major classes: some techniques are point-based, i.e., the underlying data are either raw gaze points or aggregated fixations, and the other techniques are based on additional information about areas of interest (AOIs). Our technique can be classified as a point-based approach because AOIs are not required. The stimulus content is passive, i.e., participants cannot influence the presented stimuli. We

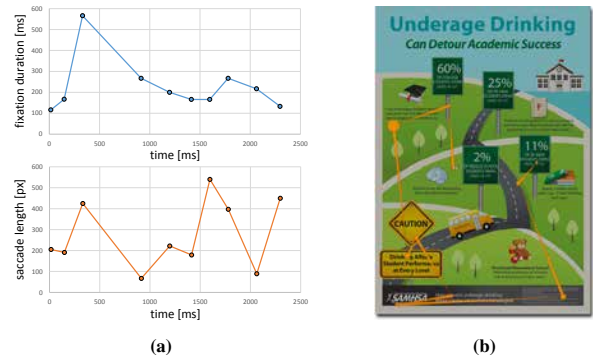


Figure 2: Typical examples of a visual analysis of the spatiotemporal aspects of eye tracking data: (a) statistical plots of fixation duration and saccade length, (b) a gaze plot drawn on the stimulus.

provide spatiotemporal information about the data from static and dynamic stimuli in a static, yet interactively adjustable visualization. Other techniques with similar properties either aggregate fixations [Duchowski et al. 2012] or scanpath trajectories [Duchowski and McCormick 1998; Hurter et al. 2014]. We do not aggregate the data over multiple participants or over time (except for the aggregation of raw gaze data to fixations), allowing for an easy comparison between participants. Furthermore, our visualization guarantees the stimulus context to be free from overlap, which is an advantage compared to gaze plots.

The concept of our visualization is based on combining eye tracking measures with the context of the stimulus by creating thumbnail images of fixated areas. A similar image-based approach to label fixation data is presented by Pontillo et al. [2010], who show images of fixated regions to the analyst for a semi-automatic classification of fixated areas. However, the authors applied this approach only to assign fixations to labels, further analysis with statistical or visualization techniques is still required for this annotated data. With our approach, the analyst can interpret the data directly in the visualization. Kurzhals et al. [2016] display fixation images for every measured gaze point in the data on timelines to compare eye movements of different participants over time. The authors represent the complete sequence of data samples, fixation information is only implicitly visible by duplicated fixation images on the timeline. We adopt their approach to a much more compact timeline by showing only one image per fixation. We compensate the resulting asynchronicity by an additional visual component: the time streams. Furthermore, in the work of Kurzhals et al., information about saccade directions and saccade lengths is not available. With our new approach, we provide a visualization that is much more scalable in the temporal dimension and provides additional fixation information. Additionally, we introduce an interactive filter concept that was not included in that previous work.

There are other concepts of our visualization that are visually similar to other techniques: the time streams and the filter dials. Burch et al. [2013] apply a horizontal river visualization on a timeline to display the temporal changes of attention on different AOIs, i.e., the “flow” of attention between AOIs. In contrast, we apply vertical rivers to visualize the temporal coherence of fixations between participants as time streams, i.e., we use a similar visual design for a very different data visualization problem that does not require any AOIs. Hlawatsch et al. [2015] apply filter wheels to find scanpath sequences based on saccade directions. Their approach is based on the scanpath trajectory alone, no visual connection to stimulus context is provided. We also allow the analyst to filter by saccade direction with our filter dial, but the filter results are instantaneously visible in context of the stimulus.

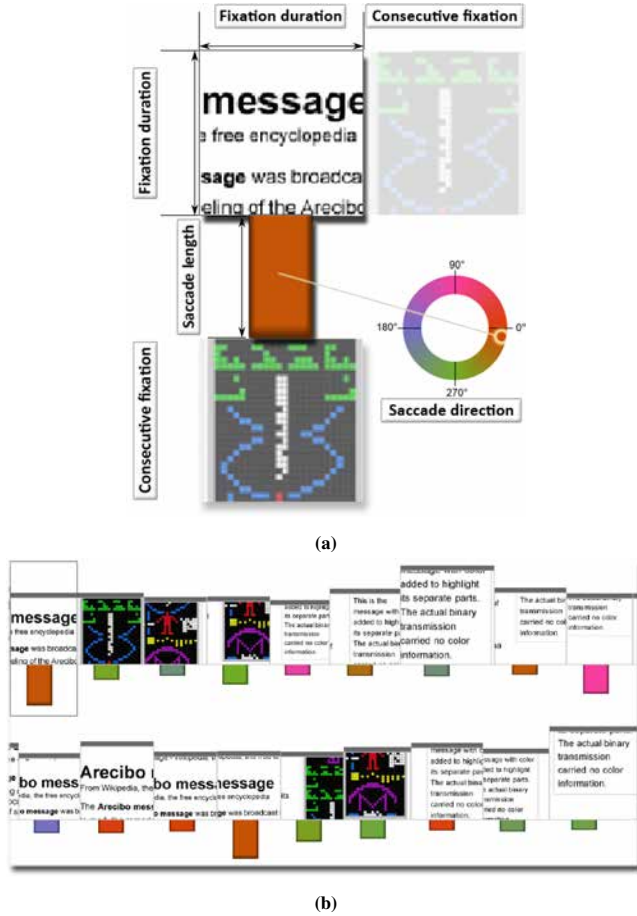


Figure 3: (a) Multiple fixation metrics can be encoded in one fixation image. (b) The sequences of fixation images represent the scanpaths of participants and can be compared with each other.

3 Visualization Design

A visualization for the analysis of sequential viewing behavior should convey the linear order and duration of fixations ①, display information about the position of fixations and the direction of saccades ②, and depict the attended context of the stimulus ③. Additionally, individual participants should be visible for comparison ④. Our visualization is designed to meet these requirements by introducing different visualization components.

3.1 Visualization Components

The visual design of our approach consists of three components (Figure 3a) that are combined on timelines for individual participants separately. Namely, *fixation-image charts* consist of **fixation images** (top) to display the properties of a fixation ① and the context of the underlying stimulus ③, **distance bars** (middle) as an indicator for saccade lengths ② between consecutive fixations (bottom, right), and **time streams** (Figure 4) to maintain the temporal synchronization between participants ④.

3.1.1 Fixation Images

The fixation images are generated based on the position of a fixation on the stimulus. Assuming a visual angle of 2° and a viewing distance of 64 cm in front of a 24" WUXGA screen, the im-



Figure 4: Time streams starting from fixed timeline intervals and comprising all fixation images that lie within the interval.

age is cropped around the fixation position (Figure 1), creating a 100×100 pixels image that approximately contains the foveated area [Nyström and Holmqvist 2010; Rajashekar et al. 2004].

So far, the fixation image provides us information where the participant was looking. To encode the duration of a fixation, the height and width of the image are adjusted accordingly (Figure 3a). We normalize the fixation duration by the longest fixation of all participants. This approach has the benefit that longer fixations also receive more space in the visualization. To emphasize the differences between low values, we use the square root of the values.

The fixation sequence of a participant is visualized by stacking the corresponding fixation images next to each other (Figure 3b). Hence, the consecutive fixation image in the sequence is always right to the current fixation. If single fixations are selected with our filtering methods (see Section 3.2.2), we show a compact representation of the results under the timelines. In this case, we display the consecutive fixation under the distance bar, indicating the target position of a saccade (Figure 3a, bottom).

3.1.2 Distance Bars

Below each fixation image, a distance bar indicates the distance between consecutive fixations (Figure 3a, middle). Although not calculated explicitly by algorithm, this distance is often applied as an implicit indicator for saccade length. The height of the bar describes the Euclidean distance between two consecutive fixations, normalized by the maximum distance within all participants' fixations. This representation is similar to a time plot of this metric (Figure 2a, bottom) that is common in eye tracking research and is easy to interpret by eye tracking experts.

The distance bars are color-coded by saccade direction to show even more information about the spatial relation between two fixations. The angle between the horizontal and the connection line between two fixations is used to obtain the color from the color-legend wheel (Figure 3a). By this approach, it is easy to detect bars of similar length and color, which can indicate repetitive behavior (e.g., reading from left to right). We decided against a glyph representation for directions (e.g., arrows) in favor of the better scalability of colors that is required in an overview of the visualization. However, the direction-to-color encoding requires some practice for interpretation. Therefore, we included the color legend directly to the corresponding filter dial (Section 3.2.2) to help with the interpretation of directions.

3.1.3 Time Streams

Since the number and length of individual fixations varies for each participant, there is a varying number of fixation images. For displaying the images on a timeline, there are two options: showing the images on an absolute timescale or showing the images stacked next to each other. The first approach preserves the synchronization

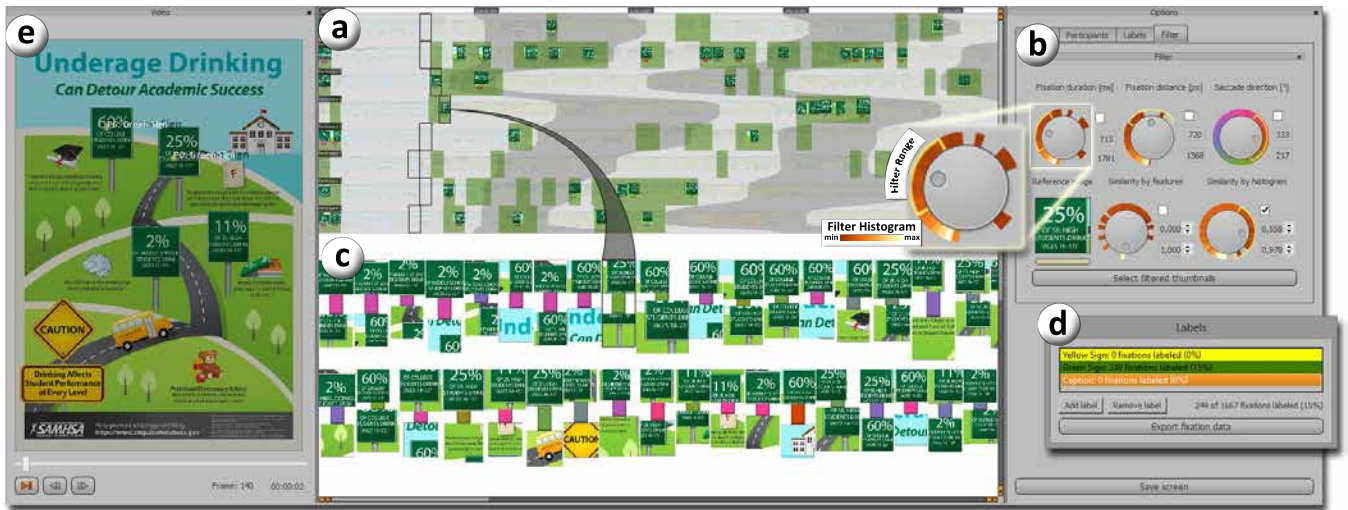


Figure 5: Visualization overview: (a) Fixation-image charts, (b) filter query interface, (c) query results, (d) label editor; (e) stimulus view.

between participants, but the resulting timeline would create gaps between fixations and the horizontal scalability would be impaired. Therefore, we decided for the second approach, neglecting absolute temporal position of fixations. This approach creates a dense representation of all fixations in their sequential order, requiring less space on the horizontal axis. However, it results in asynchronous timelines due to the varying number of fixations, impairing an efficient comparison between participants.

To compensate the asynchronicity between participants, we included time streams for equidistant time intervals. For each time stream interval, the stream passes the fixation images with the respective time stamps. The resulting segments are finally combined to a set of Bézier curves that mark the corresponding timespans for all participants. The time streams are rendered in the background of the fixation images with an alternating color scheme that can be adjusted individually.

Figure 4 shows an example of time streams from four participants looking at an infographic (Figure 6b). In the example, the number and duration of fixations for participant 5 and participant 4 are similar in the first two seconds, thus the stream is less distorted. After four seconds, participant 6 has two longer fixations leading to a stronger distortion in the time streams.

With these time streams, the asynchronicity can be compensated and the timelines remain comparable. We allow the user to adjust the selected timespan interval.

3.2 Interaction

The fixation image chart displays an overview of consecutive fixations without overlap. However, to improve the visual search for important events in the eye tracking data, additional interaction techniques are required. Therefore, we provide a linked view that displays the data with established scanpath visualizations, filter options to fade out fixation images unimportant for the current analysis, and a labeling function for selected fixations.

3.2.1 Linked View

The additional view shows the complete stimulus in the stimulus view (Figure 5(e)). A gaze replay of selected participants is

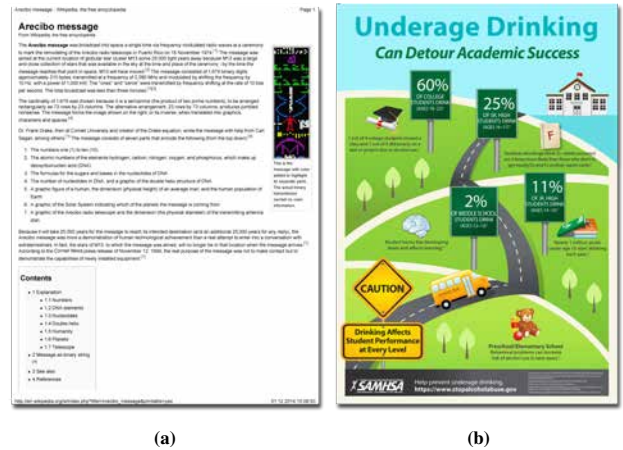


Figure 6: The two static stimuli *Arcicibo message*¹ and *UAD infographic*² are investigated with our visualization in this paper.

displayed during the playback of the recorded data. The linking between the stimulus view and the visualization is bidirectional. Changing the time in the stimulus view selects the respective fixation images in the visualization. Selecting a fixation image in the visualization sets the stimulus view to the according timestamp.

As it is common practice, the fixation duration is encoded by the radius of individual fixations. Additionally, labels are rendered into the visualization during playback, allowing for the verification of the labeling process. Looking at our example data from the infographic, we have two fixations from participant 0 and 6 in the current time step that are labeled as “Green Sign” (Figure 5(e)). By selecting consecutive fixations of a participant, the respective part of the scanpath is also highlighted in the stimulus view.

3.2.2 Filtering

The fixation-image charts provide a good overview of all fixations in a dataset. However, we need support for selecting and highlighting fixations according to different properties. In this way, it becomes easy to investigate fixations and label them. Therefore, we integrated an interactive query interface that allows us to filter according to the following categories:

¹ http://en.wikipedia.org/wiki/Arcicibo_message

² <https://www.stopalcoholabuse.gov/resources/Infographics/share.aspx?info=6>

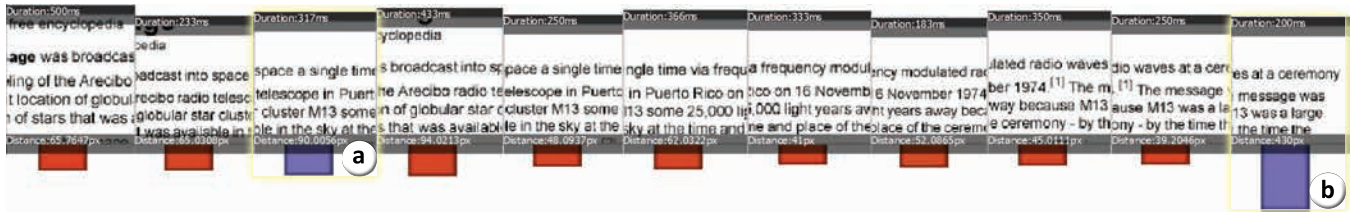


Figure 7: Visual pattern of reading behavior in our visualization. Many short saccades in reading direction (red) appear together with few (a) short and (b) long regressions against the reading direction (purple).

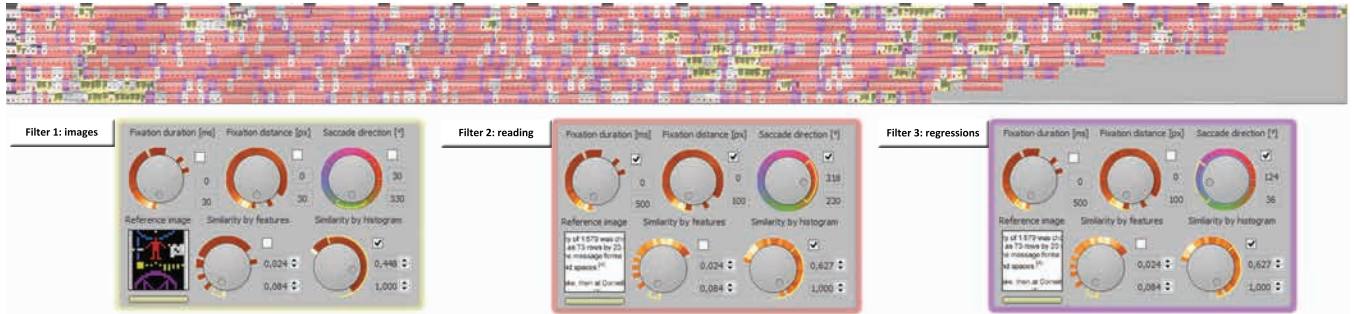


Figure 8: Three labels are defined for fixations on the picture (yellow), reading (red), and regressions during reading (purple). With three filter queries, about 80% of the fixations can be labeled very efficiently.

Filter by fixation data: The properties *fixation duration*, *fixation distance*, and *saccade direction* can be applied as filter criteria for selecting fixation images. Hence, knowledge about certain eye movements can be applied to highlight corresponding fixations. We will discuss this approach in more detail in Section 4.1.

Filter by image similarity: The analyst can select a reference image from the visualization and retrieve similar images. To this end, we provide two similarity measures. The first is based on the detection and matching of SIFT features [Lowe 1999] between the reference and the fixation images. To normalize the similarity measure, we use the number of matching features between the reference image with itself. The second similarity measure is based on a histogram comparison with the Bhattacharyya distance [Bhattacharyya 1946] between the images. We choose these two measures because of their applicability to different analysis tasks. Regions with similar structure can be identified by SIFT features, whereas regions with similar color can be identified by histograms.

All filters can be enabled and adjusted individually by separate dials (see Figure 5(b)). All enabled filters are concatenated by a logical AND connection. The analyst can choose a value and adjust the filter range with the mouse to select the values accepted by the filter. Fixation images outside of the selected ranges will be faded out in the visualization, highlighting only the currently relevant images. In our example, a reference image from a green sign is selected and the histogram similarity is activated in the search query. The query results show us the most similar fixations on green signs in the picture, sorted by their temporal appearance.

3.3 Labeling

Supporting the dissemination of analysis results, we allow the analyst to specify labels that can be assigned to fixations (Figure 5(d)). In contrast to intersection tests with AOIs, labels can describe more than just regions or objects, e.g., reading behavior or task-specific events (see Section 4.1). Labels are visible in the visualization through their assigned color in the background of the fixation images and in the stimulus view through their name.

3.4 Implementation

Our visualization prototype is implemented in C++ using Qt 4.8. Calculations on image similarities were performed with OpenCV 2.49. The source code of our prototype can be downloaded³.

4 Use Cases

To demonstrate our visualization concept, we apply the technique to static (Figure 6) and dynamic stimuli (Figure 9). The data was recorded with a Tobii T60XL and filtered with the Tobii fixation filter with standard settings (velocity threshold = 35 pixels/sample; distance threshold = 35 pixels), as recommended by Tobii⁴.

Our current implementation is designed for single stimuli with data from multiple participants. However, our concept can also be extended to data from multiple participants and multiple stimuli, as it would be the case in experiments that are recorded with mobile eye tracking glasses.

4.1 Efficient Labeling of Fixations

For the application to static stimuli, we investigate a test dataset recorded from 8 participants, looking at a Wikipedia page about the *Arecibo message* sent into space (see Figure 6a). A free-viewing task with a time limit of 90 s had to be performed.

In this use case, we want to show how our approach can be applied to label fixation data very efficiently. Labeled fixations are often necessary to evaluate the effects of changes in independent variables on measured dependent variables statistically. Labeled fixation data could also come from an automatic algorithm (e.g., for reading detection [Campbell and Maglio 2001]) and the analyst might be interested in verifying the results of the algorithm.

There are basically two approaches to label fixation data: Define areas of interest and check for each fixation if it lies inside one

³<http://go.visus.uni-stuttgart.de/stva>

⁴Tobii Studio 3.2, <http://www.tobii.com>

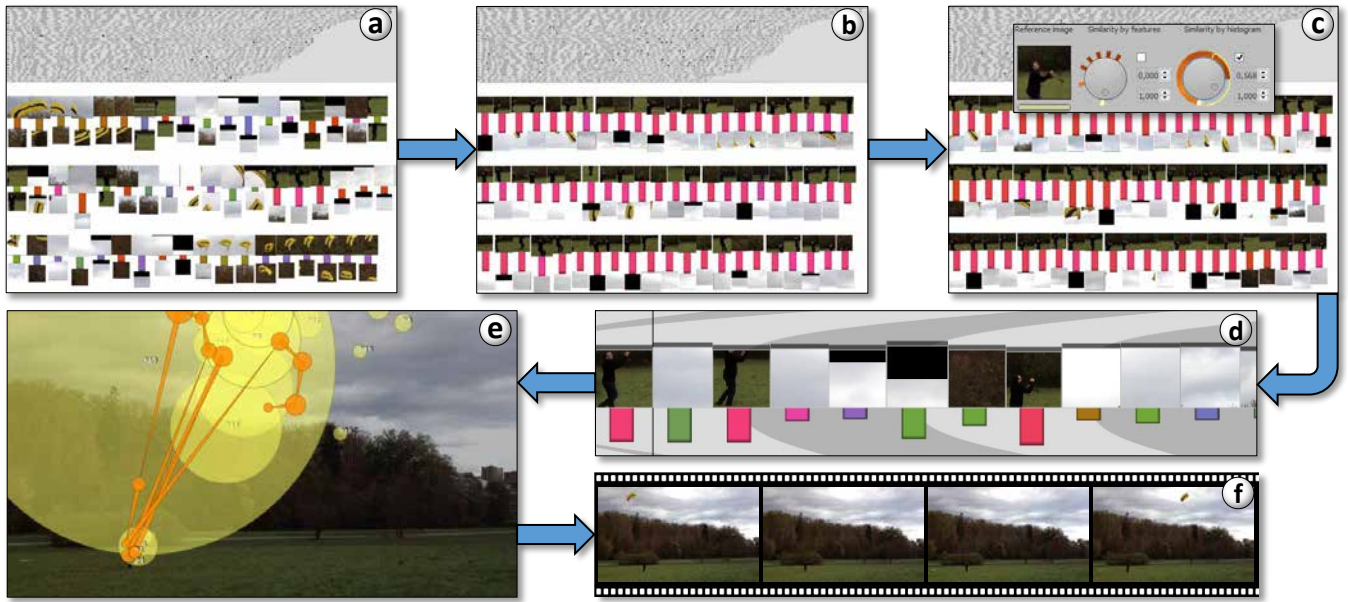


Figure 9: Results for the Kite dataset [Kurzahls et al. 2014a]. (a) Filtering for long fixations does not lead to a clear result. Nevertheless, we can see that many of the long fixations occur when the participants looked at the kite. Furthermore, their gazes moved in similar directions (visible in the color of the bars) when they looked at the kite at the same time. (b) Filtering for long saccades in upward direction provides clearer results. Most of these saccades start at the person controlling the kite. (c) To confirm this gaze behavior, we use the image similarity filter to obtain all fixations related to this person. We can see that all subsequent saccades go in upward direction. (d) We pick out one of these sequences to analyze this behavior further. The fixation images already show that the gaze moves between the person and the sky. (e) The respective scanpath confirms this. (f) Looking at the respective part of the video stimulus, we can see that the kite leaves the camera view in upward direction in this video sequence.

of these areas, or inspect each fixation individually and assign an appropriate label to it. Both approaches have their advantages and shortcomings. The main shortcoming for both approaches is that the definition of areas of interest or the assignment of labels can be very time-consuming. We aim for the second approach, since especially for dynamic stimuli, the annotation of areas of interest is a tedious preprocessing step. We take advantage of the filter query discussed in Section 3.2.2 to improve the annotation step.

Since our visualization already provides an overview of all fixations of the dataset, we can apply different filter settings to find similar fixations and assign them to a label. We choose three label categories: *image*, *reading*, and *regressions*. *Image* describes fixations that lie on the corresponding region of the stimulus (Figure 6a, top-right). Note that in this case, the same labeling could be achieved by defining areas of interest on the image region. However, the advantage of the individual labeling comes with the labels *reading* and *regressions*. The mean fixation duration during reading is approximately at 225 ms and the mean saccade length at 2° , depending on a number of factors [Rayner 1998]. Regressions typically occur during reading and are characterized by saccadic jumps opposite to the reading direction. Longer regressions typically appear when the participants jump back in the text, either to read the next line, or because they did not understand the text. Figure 7 shows a typical reading pattern in our visualization. The short fixations under 500 ms on the text are followed by short saccades under 100 px in reading direction (red). Regressions (purple) appear during the reading process either to (a) re-investigate words (in this case the word *Arecibo*) or (b) read the next line.

In the overview visualization (Figure 8, top), we can already see in the images that most of the 1793 fixations lie on regions with text and few fixations on the picture on the page.

Filter 1: By picking one of the fixation images with picture content, we obtain the similarities of all other fixation images to this reference. The main difference between the text and the *image* region lies in their histograms. Based on this, we can remove all fixation images under a certain threshold and mark the remaining fixations with the *image* label (yellow).

Filter 2: To identify *reading* fixations, we start with a histogram similarity for a fixation image containing mainly text. With these filter settings for the image similarity, we can now add the information about the fixation duration, saccade length, and saccade direction to identify fixations where a *reading* fixation appeared. Therefore, we adjust the filter dial for the direction to a span of about 90° to the right side and adjust the other dials accordingly. This leaves us with all the fixations for the label *reading* (red).

Filter 3: Finally, we keep the image similarity for text and just change the saccade direction filter to the left. By deactivating the other filters, we select long and short regressions. The resulting fixations are labeled as *regressions* (purple).

After this process, we have labeled about 80 percent of the fixations with only three filter queries in a very short timespan. For the remaining 20% of fixations, we can either decide individually which label they belong to, or perform new filter queries. The resulting labeled data could now, for example, be compared with the results of a classification algorithm for reading behavior.

4.2 Explorative Data Analysis

The use case in Section 4.1 required the analyst to apply previous knowledge from the literature about the properties of fixations for reading behavior to perform the labeling task. In our second exam-

ple, we assume that the analyst does not know what special behavior to look for. Hence, an explorative data analysis can help identify potentially interesting patterns.

In some cases, it is hard to obtain useful information by just looking at the overview visualization. In these cases, we can apply different filter methods to reduce the amount of visible information. Our filter methods cover standard eye tracking metrics like fixation duration or saccade length. Figure 9 demonstrates the utility of these filter methods. Filtering for long fixations does not directly lead to a clear pattern. However, it is visible that in many cases, the participants looked at the kite when a long fixation occurs. However, some of the long fixations also occur at the person controlling the kite. While the long fixations on the kite are followed by saccades in different directions, the long saccades following long fixations on the person seem to go all in upward direction. This can be verified by filtering for long saccades in upward direction. We can see that almost all of them start at the person. This can be further confirmed with the image similarity filter.

To find an explanation for this behavior, we can look at a respective sequence of saccades. We can see that the gaze seems to move between the person and the sky. This is also confirmed by the scan-path visualization. However, the kite does not appear in the fixation images. The respective sequence of the video stimulus provides an explanation for this. We can see that the kite left the camera view at the upper border and that the participants then looked at the person controlling the kite. From this, it can be assumed that the gazes primarily followed the motion of the kite. However, when the kite leaves the scene, the gazes move between the person and the sky. The participants seem to wait for the kite to reappear.

5 Discussion

In an initial qualitative user study, we presented our technique to three visualization experts from our institute with experience in eye tracking analysis. They were not involved in the development process of the visualization prototype. In separate sessions, they tested the application on three datasets. Each session took about 60 minutes to investigate the datasets. The first dataset, *Arecibo Message*, was used as test data, where the concept and functionality of the application was explained. For the second and third dataset, *UAD Infographic* and *Kite*, the experts were allowed to freely explore the data and search for noticeable events and patterns. During the exploration, the think aloud method was applied to protocol the experts' analysis strategies.

After the initial explanation, the visualization was interpreted easily by the experts. The overview of the visualization was investigated shortly before all experts applied the filter queries. Here, the strategies differed between the experts. We observed that the main strategies for the first search queries were the search for the longest fixation durations, the longest saccades, and direct search queries for images similar to fixation images the experts considered important in the visualization.

Although the visualization was received well by the experts, there are also potential shortcomings. Based on the expert feedback and general considerations, we identified the following issues.

Long fixations and smooth pursuit in videos: In contrast to static stimuli, video content can change drastically between consecutive video frames. By aggregating sampled gaze data into fixations, those changes are lost. In our current implementation we show the first video frame corresponding with the timespan of a fixation. Therefore, objects that moved through the fixated area within one fixation timespan can only be identified in the video re-

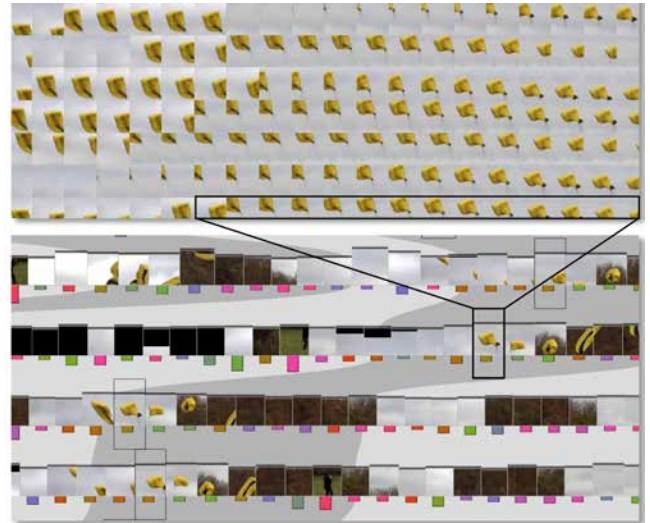


Figure 10: Comparison of gaze stripes [Kurzahls et al. 2016] (top) and fixation-image charts (bottom). The marked timespan in the gaze stripes is covered by a single image in the fixation-image chart.

play. Similar to this problem is the question how smooth pursuit can be handled by the visualization. Both scenarios require us to display more information about the underlying video frames. For example, the affected fixation images could be split into image sequences that describe the stimulus context more detailed. Those image sequences could be stacked on the timeline, or small animation loops could be applied, preserving the dense layout on the timeline.

Ambiguous labels: As with many automatic algorithms, in some cases the calculated similarity between images might be high, although the watched content is different. This might happen for example when two objects are close to each other and although the participant looked at one object, parts of the other object can be visible in the cropped image. In this case, the image can appear in both filter queries, if we search for the objects separately. Therefore, we suggest that the labeling process is performed by searching for the image similarity of both objects and finally, ambiguous fixations with more than one label can be investigated individually to solve the labeling problem. The efficiency of this approach could be improved by an additional view that lists ambiguous fixations. Additionally, new similarity metrics could be integrated into the system, providing a better accuracy for specific analysis tasks.

Comparison to gaze stripes: Figure 10 shows a comparison of gaze stripes [Kurzahls et al. 2016] and our proposed technique for the same part of the *Kite* dataset, as gaze stripes is the approach that is most similar to ours. We can see that temporal synchronicity is reduced in fixation-image charts. While it is easy to see in gaze stripes that all shown participants look at the same part of the stimulus, this information is more difficult to obtain from fixation-image charts. The time streams in the background alleviate this issue to some degree. Furthermore, fixation-image charts require more space in vertical direction. However, the space is used to provide additional information. We can see that for the selected point in time (black border), the gazes of all shown participants proceed into lower right direction (brown bars). This cannot be directly deduced from gaze stripes. Finally, fixation-image charts exhibit a much higher scalability with respect to the covered timespan. The figure shows that almost the full time range covered by the gaze stripes is represented with a single image in the fixation-image chart.

6 Conclusion

We presented a new visualization approach to display and compare fixation data from multiple participants. Our technique represents scanpaths by image sequences that convey additional information about fixation durations, saccade lengths and directions, and the local context of the stimulus. Through the compact representation of fixations, our approach scales well with respect to the duration of recorded eye tracking data. We demonstrated that our approach allows an efficient labeling of fixations. Furthermore, it supports the explorative analysis of eye tracking data and enables the detection of interesting fixation patterns. However, further evaluation considering the performance and acceptance of the technique in comparison to established methods will be required. So far, we applied our technique to static and dynamic stimuli with passive stimulus content. For future work, we plan to further evaluate our technique for active stimulus content from interactive applications and mobile eye tracking. For those scenarios, fixation patterns will be harder to interpret, as the underlying stimulus can vary extremely between participants due to changes in the environment (e.g., from different lighting).

Acknowledgements

This work was funded by the German Research Foundation (DFG) as part of the Cluster of Excellence in Simulation Technology (EXC 310/2) and the Priority Program SFB/Transregio 161.

References

- ANDRIENKO, G. L., ANDRIENKO, N. V., BURCH, M., AND WEISKOPF, D. 2012. Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and Computer Graphics* 18, 12, 2889–2898.
- BHATTACHARYYA, A. 1946. On a measure of divergence between two multinomial populations. *Sankhyā: The Indian Journal of Statistics* 7, 401–406.
- BLASCHECK, T., KURZHALS, K., RASCHKE, M., BURCH, M., WEISKOPF, D., AND ERTL, T. 2014. State-of-the-Art of Visualization for Eye Tracking Data. In *EuroVis - STARs*, 63–82.
- BLASCHECK, T., JOHN, M., KURZHALS, K., KOCH, S., AND ERTL, T. 2016. VA²: A visual analytics approach for evaluating visual analytics applications. *IEEE Transactions on Visualization and Computer Graphics* 22, 1.
- BURCH, M., KULL, A., AND WEISKOPF, D. 2013. AOI rivers for visualizing dynamic eye gaze frequencies. *Computer Graphics Forum* 32, 3, 281–290.
- CAMPBELL, C. S., AND MAGLIO, P. P. 2001. A robust algorithm for reading detection. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, 1–7.
- DUCHOWSKI, A., AND MCCORMICK, B. H. 1998. Gaze-contingent video resolution degradation. In *Proceedings of SPIE*, vol. 3299, 318–329.
- DUCHOWSKI, A. T., DRIVER, J., JOLAOSO, S., TAN, W., RAMEY, B. N., AND ROBBINS, A. 2010. Scanpath comparison revisited. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 219–226.
- DUCHOWSKI, A. T., PRICE, M. M., MEYER, M., AND ORERO, P. 2012. Aggregate gaze visualization with real-time heatmaps. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 13–20.
- DUCHOWSKI, A. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Method, Instruments, & Computers* 34, 4, 455–470.
- GOLDBERG, J. H., AND HELFMAN, J. I. 2010. Visual scanpath representation. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 203–210.
- HLAWATSCH, M., BURCH, M., AND WEISKOPF, D. 2015. Visual analysis of eye movements by hierarchical filter wheels. In *19th International Conference on Information Visualisation IV*, 107–113.
- HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., JARODZKA, H., AND VAN DE WEIJER, J. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- HURTER, C., ERSOY, O., FABRIKANT, S., KLEIN, T., AND TELEA, A. 2014. Bundled visualization of dynamic graph and trail data. *IEEE Transactions on Visualization and Computer Graphics* 20, 8, 1141–1157.
- KURZHALS, K., BOPP, C. F., BÄSSLER, J., EBINGER, F., AND WEISKOPF, D. 2014. Benchmark data for evaluating visualization and analysis techniques for eye tracking for video stimuli. In *Proceedings of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Visualization*, 54–60.
- KURZHALS, K., HEIMERL, F., AND WEISKOPF, D. 2014. ISeeCube: Visual analysis of gaze data for video. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 43–50.
- KURZHALS, K., HLAWSCH, M., HEIMERL, F., BURCH, M., AND WEISKOPF, D. 2016. Gaze stripes: Image-based visualization of eye tracking data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1.
- LOWE, D. G. 1999. Object recognition from local scale-invariant features. In *The proceedings of IEEE International Conference on Computer Vision*, vol. 2, 1150–1157.
- NYSTRÖM, M., AND HOLMQVIST, K. 2010. Effect of compressed offline foveated video on viewing behavior and subjective quality. *ACM Transactions on Multimedia Computing, Communications, and Applications* 6, 1, 4:1–4:14.
- PONTILLO, D. F., KINSMAN, T. B., AND PELZ, J. B. 2010. SemantiCode: Using content similarity and database-driven matching to code wearable eyetracker gaze data. In *Proceedings of the Symposium on Eye-Tracking Research & Applications*, 267–270.
- RAJASHEKAR, U., CORMACK, L. K., AND BOVIK, A. C. 2004. Point-of-gaze analysis reveals visual search strategies. In *Proceedings of SPIE*, vol. 5292, 296–306.
- RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3, 372–422.
- 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.
- WEST, J., HAAKE, A., ROZANSKI, E. P., AND KARN, K. 2006. eyePatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 149–154.