Benchmark Data for Evaluating Visualization and Analysis Techniques for Eye Tracking for Video Stimuli

Kuno Kurzhals, Cyrill Fabian Bopp, Jochen Bässler, Felix Ebinger, and Daniel Weiskopf
University of Stuttgart, Germany

ABSTRACT

For the analysis of eye movement data, an increasing number of analysis methods have emerged to examine and analyze different aspects of the data. In particular, due to the complex spatio-temporal nature of gaze data for dynamic stimuli, there has been a need and recent trend toward the development of visualization and visual analytics techniques for such data. With this paper, we provide benchmark data to test visualization and visual analytics methods, but also other analysis techniques for gaze processing. In particular, for eye tracking data from video stimuli, existing datasets often provide few information about recorded eye movement patterns and, therefore, are not comprehensive enough to allow for a faithful assessment of the analysis methods. Our benchmark data consists of three ingredients: the dynamic stimuli in the form of video, the eye tracking data, and annotated areas of interest. We designed the video stimuli and the tasks for the participants of the eye tracking experiments in a way to trigger typical viewing patterns, including attentional synchrony, smooth pursuit, and switching of the focus of attention. In total, we created 11 videos with eye tracking data acquired from 25 participants.

Categories and Subject Descriptors

Human-centered computing [**Visualization**]: Empirical studies in visualization

General Terms

Evaluation, benchmark

Keywords

Eye tracking, visualization, evaluation methods

BELIV'14, November 10, 2014, Paris, France

ACM 978-1-4503-3209-5/14/11...$15.00
http://dx.doi.org/10.1145/2669557.2669558

1. INTRODUCTION

The popularity of eye tracking for the analysis of human viewing behavior in various research fields increased remarkably over the last years. In addition to the statistical analysis of the recorded eye tracking data, visualization and visual analytics techniques can be applied to support these results or even discover new aspects in the data. Simple visualization techniques such as attention maps [17] and scanpath visualizations [11] provide only a coarse overview of the data. Hence, eye tracking data analysis has become increasingly advanced and enriched by new visualization techniques and interactive analysis methods over the last years [3]. For the development of new visualization techniques and visual analytics systems that support typical eye movement analysis tasks [2], researchers often have to produce their own datasets, or rely on available datasets created for other purposes. Unlike some other fields of visualization where there are common test datasets, we have been witnessing that research papers on eye tracking visualization and visual analytics do not use common test data that would allow the research community to easily compare between the different visualization and analytics techniques. With this paper, we want to fill in this gap by providing publicly available benchmark data that supports the replicability of results and a comparison between different techniques.

In general, the stimuli presented in eye tracking studies can be separated in static—such as text documents or pictures—and dynamic—such as videos and user-influenced content. To this point, the main focus was on the development of new techniques for the analysis of data recorded from static stimuli. Due to the additional problems of dynamic stimuli, such as dynamic areas of interest (AOIs) and the synchronization of multiple participants for scanpath comparison, the development of new techniques for the analysis of dynamic stimuli is still an open field for current and future research. Therefore, we have chosen to support this line of research by providing gaze data for dynamic video stimuli.

This benchmark serves as input data for visualization research. With the provided data, new analysis techniques can be developed to answer open research questions. As an example, the data could be used to develop a new technique to efficiently identify viewers with different tasks, based on their viewing behavior. All in all, the data analysis of spatio-temporal eye tracking data in combination with video content is still missing efficient approaches for several analysis tasks [3]. With this benchmark, we provide data for future research in this field to the visualization community.

The curated data with controlled conditions covers relevant viewing characteristics. The actual evaluation of new visualization techniques based on this benchmark has to be done by means like task performance, or qualitative evaluation.

Our dataset focuses on eye tracking data from multiple participants while watching video. Since the participants did not interact with the stimuli, a set of synchronizable recordings for the comparison of multiple participants could be created. We designed the video contents and the tasks to induce typical viewing patterns (see Section 3). In this way, the benchmark data is designed to test a variety of analysis goals that one typically wants to perform with dynamic gaze data. To evoke these patterns, the content of the stimuli and the viewing tasks were controlled and a user study was conducted to create a dataset of eleven videos with gaze data from 25 participants.

For the analysis of eye tracking data, the annotation of AOIs is a often necessary step for various analysis methods. Approaches that solely rely on the spatio-temporal eye tracking data provide only rudimentary information. Especially for dynamic stimuli, semantic information about object boundaries can be used to apply new metrics to the data (e.g., fixation counts on objects or transitions between objects). Hence, we included annotation data for important objects in the dataset. In Section 5, we demonstrate in a use case, how the data can be displayed in a space-time cube visualization, to achieve an overview of the recorded patterns.

In summary, our data suite consists of three parts (see Section 4): the dynamic stimuli in the form of video, the eye tracking data (spatio-temporal data), and AOI annotations with semantic naming.

The benchmark is available at:
http://go.visus.uni-stuttgart.de/eyetrackingBenchmark

2. RELATED WORK ON EYE TRACKING DATA

Outside the visualization and visual analytics community, there are some publicly available collections of datasets that contain stimulus material along with eye tracking information. Winkler and Subramanian [20] provide an overview of such data, summarizing 28 datasets of which 11 include video data with correlated eye tracking data. Figure 1 displays how our dataset can be compared to these other video datasets regarding the number of stimuli and participants.

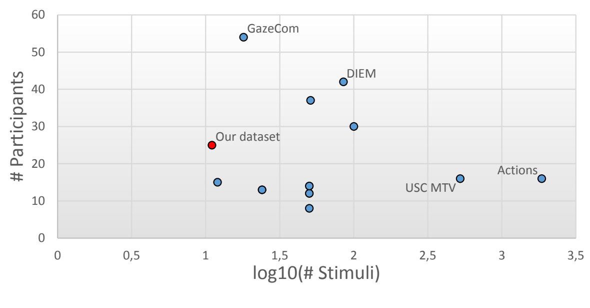

Figure 1: An overview of available datasets that include video stimuli and corresponding eye tracking data (according to Winkler and Subramanian [20]).

These datasets were created with different intentions and objectives than ours, e.g., research on visual saliency [5, 14], video quality [1], and the natural viewing behavior for everyday video material [10, 15]. The tasks in these datasets include either one specific task per dataset and/or a free-viewing task. All datasets include gaze data from between 8 to 54 (*GazeCom* [9]) participants. One exception is the *DIEM* dataset [7]: since the overview of the datasets [20] was created, the number of records increased for some videos to more than 200 participants. The datasets *Actions in the Eye* [14] (1857 videos) and *USC CRCNS MTV* [5] (523 videos) contain the largest number of stimuli.

In the visualization and visual analytics community, we are only aware of a single data set with gaze data that would have been made publicly available [19]: eye tracking data for a user study on the readability of tree diagrams [4]. However, this data is restricted to static stimuli and covers only one class of stimuli, showing hierarchy visualizations.

Our benchmark data is in the medium range in terms of the number of participants and at the lower end of the number of stimuli: it includes 11 videos with different tasks and eye tracking data from 25 participants. In fact, our goal was not to push the boundaries in terms of number of stimuli and participants. Instead, we target a controlled set of stimuli and task combinations for benchmark purposes. Our data was designed to induce similar viewing behavior of multiple participants, covering a comprehensive set of viewing patterns. Additionally, we provide dynamic AOI annotations of the data that can be included for a statistical and visual analysis of the data. None of these objectives are pursued by any of the previous data collections.

3. VIEWING PATTERNS

The videos and tasks for the eye tracking experiment were designed to evoke 3 different patterns that are most common in dynamic stimuli. These patterns either emerge from attention distributed between different AOIs, or from attention focused on individual AOIs.

- **Switching focus**: The participants have to attend to various AOIs simultaneously. Since the task can require to distribute the visual attention, the participants continuously switch their focus between AOIs. Although tracking of multiple objects is possible [6], retaining of visual features is limited in this case and for identification tasks, the participants still have to focus on single objects.

- **Attentional synchrony**: The stimulus contains time spans with one AOI that attracts the attention of all participants, even if their eyes have been tracked separately. Attentional synchrony has been investigated for static and dynamic stimuli [16]; due to the high saliency of movement, it can appear frequently in dynamic scenes.

- **Smooth pursuit**: A moving AOI in the stimulus attracts the attention of the participants, causing them to follow its movement. In these time spans, smooth pursuit eye movement [11] can be present.

The above viewing patterns are canonical patterns that may incur in gaze data for dynamic stimuli. Since the induced patterns are often guided by the tasks given to the participants of the experiment, we also include the effect of task dependency:

Table 1: The recorded stimuli with a description of the stimulus settings, the given tasks, and eye movement patterns that could be observed. Stimuli marked with (*) included different tasks for the 2 participant groups.

ID	Stimulus	Setting	Task	Induced Patterns
S1	Car Pursuit (0:25 min)	Panning camera follows a red car while it was going through a roundabout.	Follow the red car.	Potential **smooth pursuit** with long time spans of **attentional synchrony** on the red car.
S2	Turning Car (0:28 min)	Camera follows turning car. The movement of the car describes the shape of an eight.	Recognize the shape that is described by the movement of the car.	**Attentional synchrony** on the car with potential **smooth pursuit** eye movement.
S3	Dialog (0:19 min)	Two persons talk to each other in front of the camera.	Follow the dialog attentively.	**Switching focus** between the faces of both persons. Label on shirt (right person) attracts additional attention.
S4	Thimblerig (0:30 min)	A thimblerig with three cups and a marble.	Find the cup with the marble.	**Attentional synchrony** mainly on the cup with the marble.
S5	Memory (2:28 min)	A 4×4 memory game. Pairwise flipping of cards is performed until all pairs are found.	After one card is flipped, focus on the corresponding card of the pair.	Increasing attention on matching cards after several turns and **switching focus** during the search.
S6	UNO (2:01 min)	Two persons play UNO card game until the right player wins.	For each player's turn, focus on the playable cards on the hand.	**Switching focus** and attention mainly distributed between both hands and the stack of played cards.
S7	Kite (1:37 min)	Person on a meadow steers a kite. The kite repeatedly leaves the field of view.	Follow the flight path of the kite if possible.	**Smooth pursuit** if the kite is visible. Otherwise, the participants either tried to estimate the position of the kite, or focused on the person.
S8	Case-Exchange (0:27 min)	Various persons crossing the field of view while a text ribbon in the lower part is showing further information.	Task is provided by the text ribbon: Look for metal case.	**Attentional synchrony** on the text ribbon until the metal case appears and the task is readable.
S9	Ball Game* (0:31 min)	Three players with orange shirts and one player with a white shirt pass a ball around.	Task group A: Count ball contacts of the white player. Task group B: Count passes between orange players.	**Attentional synchrony** often on the ball, independent from the task.
S10	Bag Search* (2:13 min)	Various persons carrying different bags are crossing the field of view.	Look for a specific bag. Two groups (A,B) with two different search targets, presented before the video started.	**Switching focus** on new bags in the scene. Depending on the group, the search targets attract more attention.
S11	Person Search* (2:52 min)	People with different clothing cross the field of view.	Task group A: Find the person with a hooded sweater. Task group B: Find the person with a red shirt and a headgear.	**Switching focus** on new persons. After identification, search targets become less important than new persons.

- **Task groups**: The participants can be separated in groups of similar viewing behavior by assigning them different tasks. This type of pattern is of special interest for new methods that compare scanpaths to identify clusters of participants. Differences between groups may also be based on the participants' background or condition (not included in our datasets); for example, one could investigate differences in the scanpaths of healthy and mentally disordered persons.

To induce these viewing patterns, we designed video scenarios with according viewing tasks. Table 1 summarizes the stimuli settings and tasks, as well as the patterns produced.

4. DATA ACQUISITION AND FORMAT

This section describes the technical details of the videos recorded as stimuli, the acquisition of the gaze data, and the annotation of the AOIs.

4.1 Data Format

We provide three components of data: the video stimulus, the recorded eye tracking data, and AOI annotations. These components are given for each of the 11 scenarios summarized in the previous section.

Video stimulus: With the exception of stimulus 7 (Kite), all videos were captured with a Panasonic HDC-SD5 camcorder. Stimulus 7 was captured with an Apple iPhone 4S at 30 frames per second (fps) since the camcorder was not available at this point in time. The other videos were recorded at 25 fps and with a tripod for stabilization. Except for stimulus 3 (Dialog), the audio track was removed from the videos since it was negligible for the tasks. Stimulus 3 has a stereo MP3-coded audio track at 128 kBit/s. All videos were converted to Xvid-coded AVI files with a frame rate of 25 fps and a maximum data rate of 12 MBit/s. The videos have a resolution of 1920×1080 pixels and were displayed centered on the screen with their native resolution. These technical parameters were chosen to ensure the compatibility with the eye tracking software.

Eye tracking data: The data was recorded with a Tobii T60 XL eye tracker, with a sampling rate of 60 Hz and a 24" screen with a resolution of 1920×1200 pixels. We provide the complete data from the recordings (except for the absolute timestamps, to protect the privacy of participants by anonymization) in separated TSV files, exported from the Tobii software. The data files include raw gaze data with coordinates and relative timestamps, as well as fixation indices extracted by the Tobii fixation filter with standard settings (velocity threshold = 35 pixels/samples; distance threshold = 35 pixels). We recommend using the raw gaze data for best accuracy and reliability of the data; this is especially true for videos with smooth pursuit because this type of eye movement is not supported by the fixation filter. A complete description of the extracted file format is available in the Tobii T60 XL manual [18].

Dynamic AOIs: To support the application of advanced analysis methods based on AOIs, we included sets of manually annotated, dynamic AOIs for every video in the dataset. With this additional information, various AOI-based eye tracking metrics can be applied to the data. AOIs are annotated by dynamic and axis-aligned bounding boxes. We provide the data in an XML format that is compatible to the well documented ViPER file format [8]. Hence, an import to other visualization or analysis systems can be performed by simple XML parsing. Mapping of gaze data to the AOIs has to be performed by an appropriate approach.

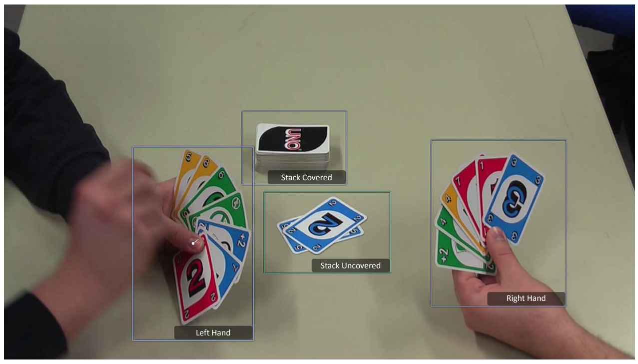

Figure 2: Dynamic AOIs are annotated by axis-aligned bounding boxes names that reflect the semantics associated with the AOIs.

Filename convention: The video filenames are coded by the stimulus ID and the name of the video. AOI annotations have the identical name for the XML file. The names for files with exported gaze data are coded by the order of participant ID, followed by the group, and the stimulus name. For example, the file *"P3A-01-car pursuit.tsv"* contains the eye tracking data for stimulus **S1** (Car Pursuit) from participant **P3** in group A.

4.2 Eye Tracking Experiment

The eye tracking data was acquired with a user study with $n = 25$ participants (60% male, 40% female; mean age of 24 years). All participants had normal or corrected-to-normal sight (tested with Snellen chart). No color vision deficiency (tested by the Ishihara color test) was detected for any participant. Except for 3 participants without academic background, all other participants were BSc or MSc (or Diplom) students, 9 of them from computer science, the others with different technical majors. The participants were divided into two comparison groups to solve different tasks for the videos: group A (13 participants) and group B (12 participants). The study procedure took about 45 minutes and each participant was compensated with 7 EUR.

At the beginning of the study, the participants were informed about the conditions of the participation and the study procedure. Then, the eye tracker was calibrated using a 9-point calibration pattern. For the main part of the study, the video stimuli (see Figure 3) were presented, counterbalancing their order to prevent bias from fatigue or learning effects. Before a video stimulus was shown, the corresponding upcoming task was described; the participants were then allowed to decide when to start the task. Since stimulus **S6** required the participants to know the rules of the card game UNO, we provided a short explanation of the rules at the beginning of the task if necessary. For stimulus **S10** and stimulus **S11**, we presented images and textual descriptions of the search targets to the participants before the video started.

4.3 AOI Annotation Process

The annotation of dynamic AOIs in the video stimuli was performed with the ISeeCube system [12]. The annotation

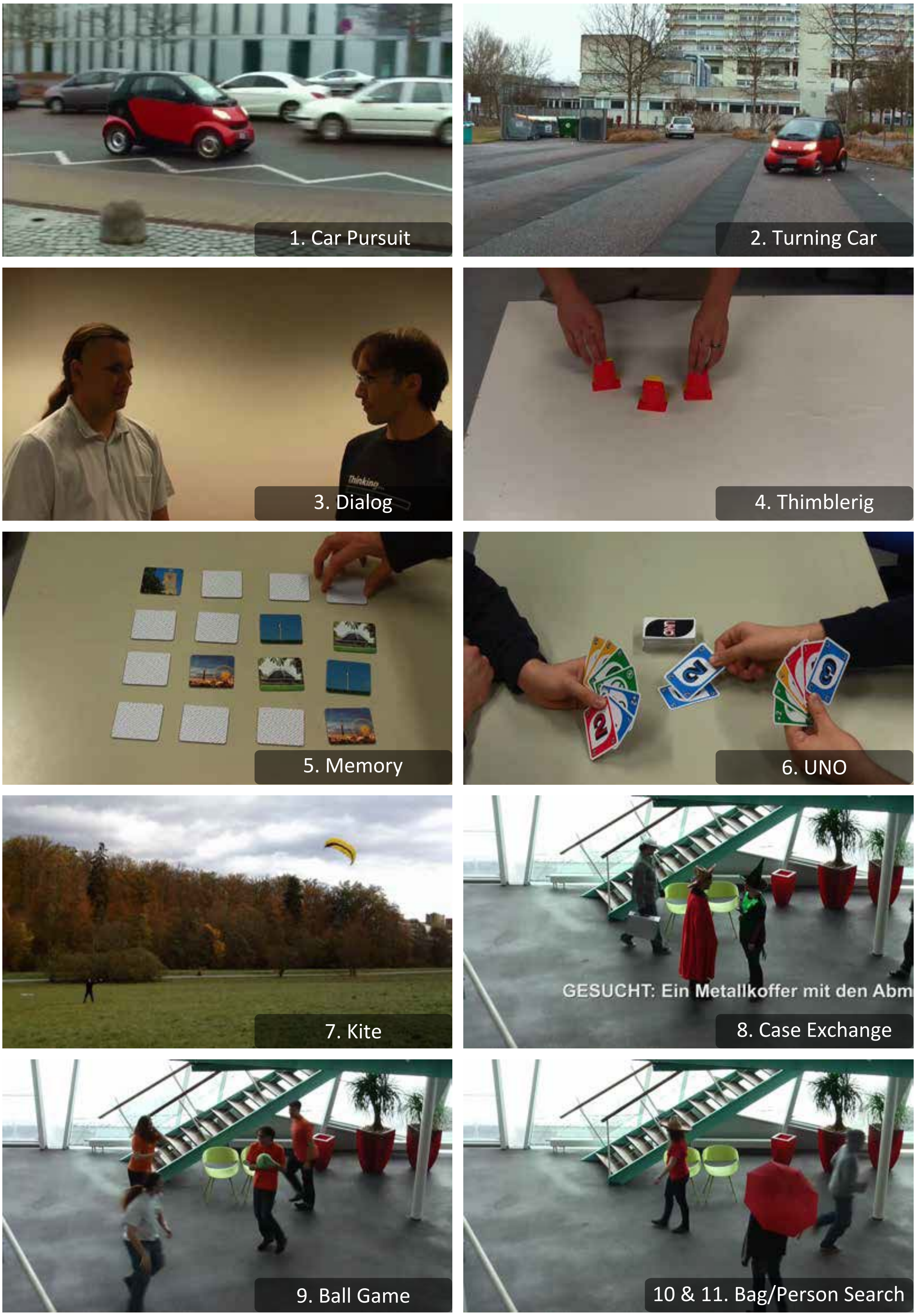

Figure 3: Video stimuli in the order described in Table 1. The stimuli comprise various settings with different eye movement patterns induced.

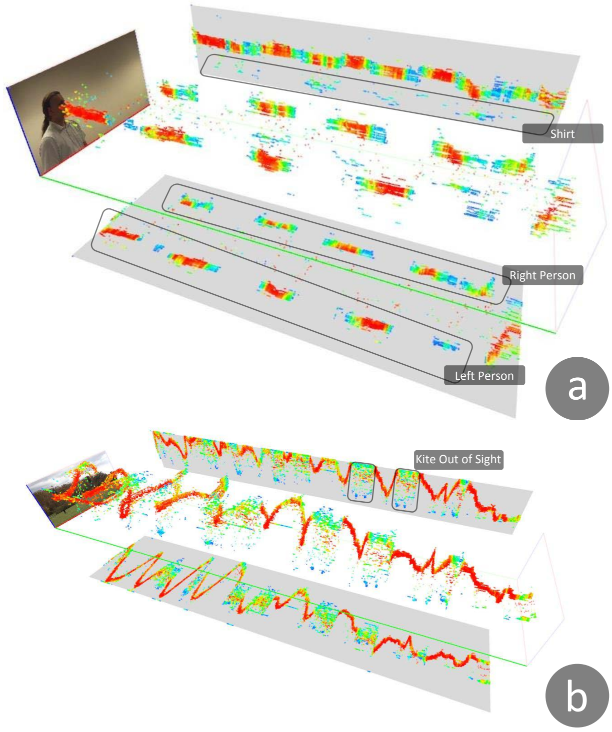

Figure 4: Space-time cube visualization of the data from S3 (a) and S7 (b). Gaze points in regions of high attentional synchrony are colored in red.

process was performed manually by 2 annotators, in order to increase the quality of the annotation. Axis-aligned bounding boxes were defined to cover the silhouette of an AOI. The bounding boxes were set at key frames, positions in between were interpolated linearly. Depending on the object movement, the distance between key frames was varied for a good approximation of the object's trajectory and size.

Only task-relevant AOIs were annotated, including semantically matching names. Figure 2 shows an annotation example of stimulus **S6**. Here, four AOIs were defined with the names that reflect the semantics associated with the AOIs: *Left Hand, Right Hand, Stack Covered, Stack Uncovered*.

The annotation took about 10 hours for the complete data, 7 hours were spent by the first annotator for the main annotation, 3 hours by the second annotator for additional annotations and refinements of the existing annotations. Depending on the research question, the definition of new or alternative AOIs might be required. Since AOIs in the videos did not have complicated silhouettes that would require complex polygonal bounding shapes, we decided for rectangular bounding boxes to facilitate the interpretation of the data. This approach is limited to stimuli where only such AOIs of simple shape exist.

5. EXAMPLE USE CASE

To demonstrate some of the possible applications of our benchmark, we visualized the recorded eye tracking data with a space-time cube (STC) [13]. In this spatio-temporal overview (see Figure 4), the gaze points can be displayed in 3D, the gray walls show 2D projections of the gaze points to reduce perceptual issues resulting from occlusion, distortion, and inaccurate depth perception. Gaze points in regions of high attentional synchrony are colored in red.

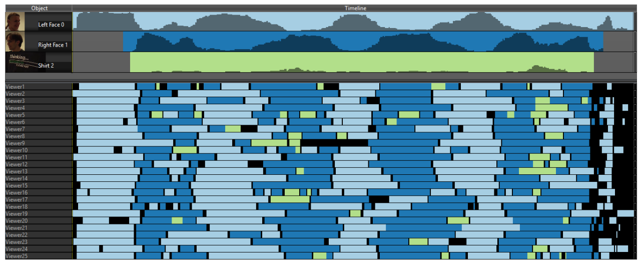

Figure 5: AOI timeline and scarf plots of the dialog (S3) show one example of how AOI information can be used to investigate how the attention of all participants was distributed over time individually.

Figure 4(a) shows the data from stimulus **S3** (Dialog). The switching focus between the two persons is clearly visible in the STC. Since all participants attended to the dialog, attentional synchrony on the talking person can be seen in the respective time spans. Additionally, the shirt of the right person attracted the attention of some participants in certain time spans.

Figure 4(b) shows the data from stimulus **S7** (Kite). Attentional synchrony on the kite is present in time spans where it is in the field of view. When the kite flies out of sight, a different viewing behavior can be noticed: some participants tried to estimate the position where the kite would re-enter the scene, whereas others switched their focus to the person steering the kite.

With the AOI annotations provided, even more analysis methods can be applied to the data. Figure 5 shows an example of a timeline visualization from Kurzhals et al. [12] where AOI information is used to visualize the distribution of attention of each participants with scarf plots. The scarf plots in the lower part show time spans colored by the corresponding color of the AOI that was focused. Additionally, the upper part shows aggregated attention histograms of the individual AOIs over time where the changing focus of attention becomes also visible.

These are two visualization examples applied to the same data, providing different visual representations. Therefore, the effectiveness of new techniques should be investigated by established means of evaluation. For the visual analysis of eye tracking data, an overview of existing techniques is provided by Blascheck et al. [3]. With this benchmark, new techniques can be developed to answer open research questions. Especially the comparison of the different task groups and the identification of visual reading strategies provide challenging analysis tasks that require the development of new techniques to gain deeper insights into the data. We envision this dataset mainly as test data for visualization approaches due to the included AOI information that is not provided by other datasets. The additional information facilitates research on various open questions that cannot—in our opinion—be solved efficiently by purely statistical analysis. Therefore, we aim to support techniques for visual exploration and visual analytics.

6. DISCUSSION

Considering the generalizability of our benchmark data, we have to distinguish between the generality of participants and the generality of the stimuli.

Regarding the participants, the dataset includes eye movement data from persons with a mean age of 24 years. A possible extension of the dataset could include data from infants or elderly people, where the viewing behavior might differ. Also, no participants with color vision deficiency took part. Extending the dataset with recordings from impaired participants could provide additional possibilities for comparison between different groups.

Regarding the stimuli, only videos with fixed camera position and without cuts were employed to synchronize data between participants. For future extension, videos with different filming techniques and videos with individual content (e.g., from eye tracking glasses) could be included. Also, the dataset includes only videos with content from everyday scenarios. No abstract content (e.g., from animated visualizations) was used.

To this point, we can provide comparable data between participants including a variety of common eye movement patterns under controlled conditions with this subclass of possible stimuli. Since the main focus of this work is on providing the aforementioned patterns, the included stimuli are well suited, since they cover these patterns adequately.

7. CONCLUSION

We designed a set of 11 videos with different settings, tasks, and dynamic AOI information. In a user study, we collected eye tracking data from 25 participants. Since the tasks were designed to induce certain eye movement patterns, the resulting data is more suitable for a benchmark application than datasets with a free-viewing task. By providing dynamic AOIs, we spare visualization experts a tedious preprocessing step.

We provide this dataset as a benchmark for the development of new visualization and visual analytics methods for eye tracking data from videos. Due to the variety of information included in our dataset, it is conceivable that it will even find applications outside of evaluating visualization techniques, e.g., in fully automatic gaze analysis methods.

8. ACKNOWLEDGMENTS

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

9. REFERENCES

[1] H. Alers, J. A. Redi, and I. Heynderickx. Examining the effect of task on viewing behavior in videos using saliency maps. In *IS&T/SPIE Electronic Imaging*, pages 82910X–82910X. International Society for Optics and Photonics, 2012.

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

[3] 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.

[4] M. Burch, N. Konevtsova, J. Heinrich, M. Höferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2440–2448, 2011.

[5] R. Carmi and L. Itti. The role of memory in guiding attention during natural vision. *Journal of Vision*, 6(9):4, 2006.

[6] P. Cavanagh and G. Alvarez. Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7):349–354, 2005.

[7] DIEM-Project. http://thediemproject.wordpress.com, visited 2014-09-12, 2014.

[8] D. Doermann and D. Mihalcik. Tools and techniques for video performance evaluation. In *Proc. of the 15th International Conference on Pattern Recognition*, pages 167–170, 2000.

[9] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth. Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10):28, 2010.

[10] H. Hadizadeh, M. Enriquez, and I. Bajic. Eye-tracking database for a set of standard video sequences. *IEEE Trans. on Image Processing*, 21(2):898–903, 2012.

[11] 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, 2011.

[12] K. Kurzhals, F. Heimerl, and D. Weiskopf. ISeeCube: visual analysis of gaze data for video. In *Proc. of the Symposium on Eye Tracking Research and Applications*, pages 351–358, 2014.

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

[14] S. Mathe and C. Sminchisescu. Dynamic eye movement datasets and learnt saliency models for visual action recognition. In *Computer Vision–ECCV 2012*, pages 842–856. Springer, 2012.

[15] P. K. Mital, T. J. Smith, R. L. Hill, and J. M. Henderson. Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1):5–24, 2011.

[16] T. Smith and J. Henderson. Attentional synchrony in static and dynamic scenes. *Journal of Vision*, 8(6):773–773, 2008.

[17] O. Špakov and D. Miniotas. Visualization of eye gaze data using heat maps. *Electronics and Electrical Engineering*, 2(74):55–58, 2007.

[18] Tobii T60 XL manual. http://www.tobii.com/en/eye-tracking-research/global/library/manuals, visited 2014-09-12, 2014.

[19] Visual analytics methology for eye movement studies data. http://geoanalytics.net/and/papers/vast2012em/data/index.html, visited 2014-09-12, 2014.

[20] S. Winkler and R. Subramanian. Overview of eye tracking datasets. In *Proc. 5th Int. Wksp. on Quality of Multimedia Experience*, pages 212–217, 2013.