Space-Time Visual Analytics of Eye-Tracking Data for Dynamic Stimuli

Kuno Kurzhals, *Student Member, IEEE*, and Daniel Weiskopf, *Member, IEEE Computer Society*

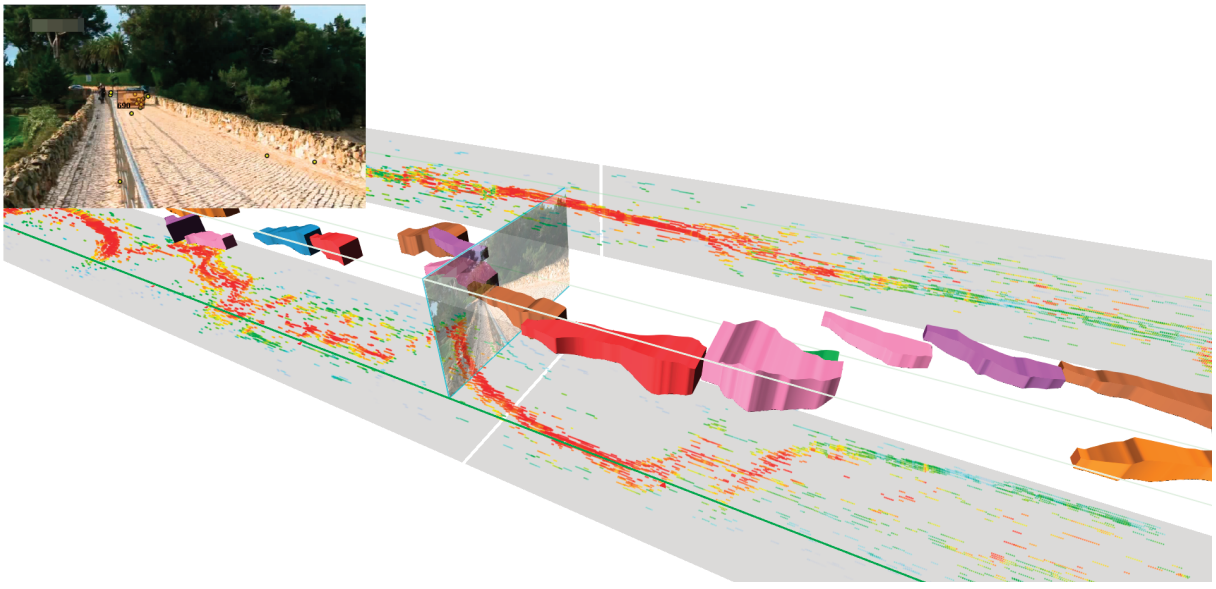

Fig. 1. Space-time cube visualization of eye-tracking data for a video stimulus, enriched by spatiotemporal clustering of eye fixations.

Abstract—We introduce a visual analytics method to analyze eye movement data recorded for dynamic stimuli such as video or animated graphics. The focus lies on the analysis of data of several viewers to identify trends in the general viewing behavior, including time sequences of attentional synchrony and objects with strong attentional focus. By using a space-time cube visualization in combination with clustering, the dynamic stimuli and associated eye gazes can be analyzed in a static 3D representation. Shot-based, spatiotemporal clustering of the data generates potential areas of interest that can be filtered interactively. We also facilitate data drill-down: the gaze points are shown with density-based color mapping and individual scan paths as lines in the space-time cube. The analytical process is supported by multiple coordinated views that allow the user to focus on different aspects of spatial and temporal information in eye gaze data. Common eye-tracking visualization techniques are extended to incorporate the spatiotemporal characteristics of the data. For example, heat maps are extended to motion-compensated heat maps and trajectories of scan paths are included in the space-time visualization. Our visual analytics approach is assessed in a qualitative users study with expert users, which showed the usefulness of the approach and uncovered that the experts applied different analysis strategies supported by the system.

Index Terms—Eye-tracking, space-time cube, dynamic areas of interest, spatiotemporal clustering, motion-compensated heat map

1 Introduction

The use of eye-tracking in various fields of research is a commonly accepted method to gain insight into how people look at certain stimuli. In psychology, the analysis of recorded eye-gaze data can lead to a deeper understanding of human cognitive processes [16]. For the analysis of visualization designs, eye-tracking can be used, for example, to gain insight into the viewers' exploration strategies of tree diagrams [10], or for the analysis of e-learning technologies [31]. The main focus in eye-tracking research in the past lies on the analysis of static stimuli such as images. Therefore, numerous methods exist to visualize fixations and scan paths of data recorded for static stimuli. For the analysis of dynamic stimuli such as video sequences, however, the number of available visualization methods is very limited and often, those techniques are not very effective. Animated heat maps, bee swarms, or just the recorded gaze paths are provided in the software packages of vendors such as Tobii [3] and SMI [2]. With the definition of dynamic Areas Of Interest (AOIs), common metrics such as fixations counts can be applied to time-dependent stimuli. For further details, see the survey of common metrics for eye-tracking data by Poole and Ball [30].

In general, the analysis of eye-tracking records from time-dependent data can either be achieved by watching the video with the afore mentioned visualization methods, or by statistical analysis of AOIs and gaze data. Watching the whole video to find interesting sequences can be time-consuming and exhausting for the analyst. Statistical analysis of AOIs requires either a reliable detection algorithm to find them, or tedious manual editing. Although future im-

- *Kuno Kurzhals is with the Visualization Research Center (VISUS), University of Stuttgart. E-mail: Kuno.Kurzhals@visus.uni-stuttgart.de.*
- *Daniel Weiskopf is with the Visualization Research Center (VISUS), University of Stuttgart. E-mail: Daniel.Weiskopf@visus.uni-stuttgart.de.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

provements in the field of computer vision could provide techniques to successfully identify all objects and regions of possible interest, human observation is still needed for semantic interpretation. For the analyst, it would be more efficient to look at a representation of the whole video at once and find interesting frame sequences without a sequential search through each frame.

An interesting sequence could be a shot from a movie that draws the attention of many viewers to one certain region at once—a phenomenon called attentional synchrony [39]. Automatic analysis could summarize these scenes, but it would lack semantic interpretation of the scene context. Our design provides the possibility to find such regions by interactive filtering and additionally shows motion patterns of traced objects. The Space-Time Cube (STC) helps identify and interpret these patterns in the spatiotemporal domain. Gaze direction changes to other positions after cuts in video scenes can also be interpreted easily with the provided visualization. Additionally, spatiotemporal clustering of data points identifies AOIs that can be filtered by the cluster size. This method helps find multiple AOIs in a time sequence and shows which objects received more attention.

We provide a method to analyze eye-tracking data recorded from static or dynamic stimuli within a space-time visualization with the focus on the analysis of multiple viewer recordings of videos. Our design combines an automatic algorithm to cluster gaze data of numerous viewers and a visualization that summarizes the whole data set. By interactive cluster and density filtering, users can identify time spans of high attentional synchrony as well as multiple regions of interest. The STC visualizes the spatiotemporal data in a static 3D representation. By interactive translation and rotation of the data set, data points and clusters can be interpreted in their original domain. To overcome problems of occlusion and depth perception, the design provides additional wall projections that can be adjusted individually.

Our interactive visualization approach is combined with computer-based analysis techniques from computer vision and data-mining. In particular, it includes low-level computer vision techniques like optical flow estimation, high-level analysis of shot detection, as well as spatiotemporal clustering.

Our key contribution is the unified spatiotemporal analysis of eye-gaze data in the spatiotemporal context of the dynamic stimulus. In this sense, our visual analytics approach is different from generic spatiotemporal analysis that does not incorporate the driving visual stimulus. Besides this fundamental contribution, we provide a couple of technical contributions as components of our visual analytics system: a motion-compensated version of eye-gaze heat maps, a variant of STC (see Figure 1) with improved visualization of spatial and temporal context, and spatiotemporal clustering that includes information from shot detection for better results.

2 Related Work

For the analysis of video material, the use of eye-tracking can provide valuable information to understand the viewing behavior of users. Tien and Zheng [41] measured gaze overlaps of a video that showed a surgical task to compare experts' gaze with the gaze of trainees. Goldstein et al. [19] examined similarities in the viewing behavior of several users to identify centers of interest in movie scenes. Marchant et al. [27] described an approach to investigate the influence of directorial techniques on film viewers' experience. Smith and Henderson [39] compared the degree of attentional synchrony between static and dynamic scenes. We provide visualization techniques that can be used to support the quantiative means of analysis from those papers, such as time-spans of high attentional synchrony.

Numerous methods exist to visualize eye-tracking data. Holmqvist et al. [21] provide a comprehensive guide to methods and measures. Generally, heat maps [7, 15, 47] are used to display aggregated eye-tracking data. Tsang et al. [43] provide a tree-like visualization for the exploration and comparison of sequential gaze orderings. Raschke et al. [32] introduced the parallel scan-path visualization to facilitate the comparison of eye-tracking data between several users. In the context of visual analytics, Andrienko et al. [5] provide a detailed methodology for eye-movement data. We adopt many of the standard visualizations; see Section 4 for more details on the components integrated in our design.

The STC is used in various fields of research. Gatalsky et al. [18] describe its application to event data in a geographical context. Chen et al. [12] and Botchen et al. [8] represent video content in 3D to depict individual objects and motion events. However, they do not include eye-tracking data in their representations. We adopt the 3D space-time video representation as context, but add the visualization of the eye-gaze data. In the context of eye-tracking, Li et al. [26] describe the use of the STC to visualize eye-trajectories. They focus on the analysis of static stimuli. For the application to dynamic stimuli, Duchowski and McCormick [14] describe a space-time representation of Volumes Of Interest for aggregated eye movement trajectories. We extend the concept for dynamic stimuli and provide different data representations in addition to the mentioned eye-trajectories.

Clustering of eye-tracking data is already used when fixations are identified in raw data. Salvucci and Goldberg [34] describe a taxonomy for different fixation identification algorithms. For the clustering of multiple user gaze data, Sawahata et al. [36] and Mital et al. [28] use a Gaussian Mixture Model. We use the mean-shift clustering approach for gaze data, according to Santella and DeCarlo [35] because it is robust to noise and does not require a preset number of clusters. However, we are not aware of any previous cluster method that would respect shot boundaries from shot detection algorithms.

3 Design Overview

Our design uses multiple coordinated views [33] to show the different aspects of spatiotemporal eye-tracking and stimulus data, as they are particularly helpful for analyzing this kind of data [4]. We provide a visual interface for analytics with adjustable and detachable components. Figure 2 shows a screenshot of the system. The main components are:

(a) **Viewer controls:** This control panel allows the individual adjustment of the visualization view, depending on the analytical task. The data point representation (Section 5.2) and the cluster representation (Section 5.4) can be enabled separately or together. The video preview, the projection walls, and the overview walls can be enabled and adjusted in size. Each projection wall can be set to show projections of data points or cluster.

(b) **Visualization view:** The visualization view consists of two components. The main component is the interactively explorable STC. It visualizes the eye-tracking and video data in their original spatiotemporal domain. The second component is the video preview that shows data points and AOIs as known from standard analysis tools.

(c) **Video controls:** To navigate through the STC, this panel provides a time slider, frame-wise navigation, and a play button. For analysis tasks related to cuts in the video, shot boundary frames can be jumped to directly, an approach that is also used in the work of Li et al. [25].

(d) **Parameter controls:** Interactive filtering and data drill-down can be performed by parameter adjustment in this panel. The first slider determines the scaling of the STC and the projection walls along the time axis. By adjusting the kernel size, data points can be filtered frame-wise by their distance to their center of mass. For the cluster representation, depicted clusters can be filtered by their size. A histogram shows the number of clusters in relation to the cluster size. Cluster size relates to the total amount of data points within a cluster, not to its spatial extent.

(e) **User list:** Each recorded viewer is listed and selectable for individual or multiple scan paths analysis. A qualitative scheme of 8 colors created by ColorBrewer 2.0 [20] is applied in a cyclic fashion to distinguish between scan paths of different viewers.

The design was developed in a formative process. In several short sessions with two visualization and eye-tracking experts, the analytical

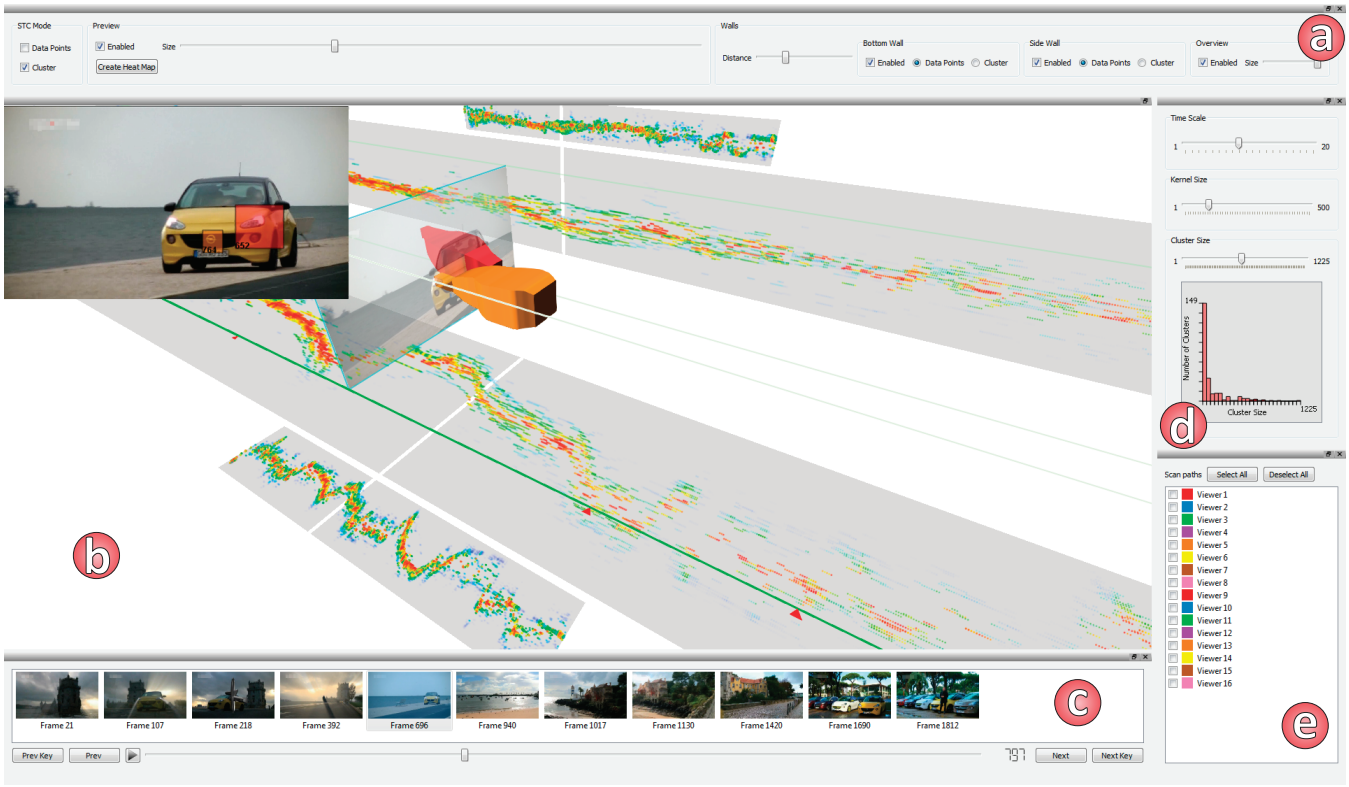

Fig. 2. Design components: (a) viewer controls; (b) space-time visualization view; (c) video controls with key-frames for single shots; (d) parameter controls; (e) user list for scan path visualization.

methods, the visualization design, and the usability in general were improved, according to their comments.

We combine standard eye-tracking visualizations (Section 4) with extended analysis methods that we have developed for dynamic stimuli (Section 5). According to the visual analytics mantra [24], we provide an automatic analysis of the gaze data by clustering and interactive filtering by cluster size. Data points can be filtered by their spatial density. In addition, automatic computer vision techniques such as optical flow estimation and shot detection allow us to off-load analysis work to the computer. The viewers' gaze direction can be influenced through abrupt cuts [11]. Therefore, it is necessary to define shot boundaries for further analysis. We use a shot boundary detection algorithm in order to support shot-based clustering and analysis related to shot boundaries.

4 Standard Eye-Tracking Visualization

In our visual analysis approach, we included established visualization methods for eye-tracking data. These methods allow for an easy adoption of the design, since the analyst is already familiar with several of its components. Also, the established methods are needed to cover a wide collection of different analysis tasks (see Section 7.5).

Heat Maps

Heat maps in eye-tracking research are commonly used to provide a qualitative impression of the users' gaze distribution. For static images, the aggregation of gaze positions over the observation time is expressive, resulting in a static heat map. For the generation of heat maps, we integrated the algorithm and color mapping described by Blignaut [6]. The principle of static heat maps can be applied to dynamic stimuli to summarize the distribution of attention, but video material with numerous cuts and various camera angles, such as Hollywood movies, can produce heat maps that bear little to no meaning. Ways to overcome this problem are either to create heat maps of very short sequences or to use dynamic heat maps. We provide the possibility to create a static heat map of a user-defined time-span as well as a dynamic heat map visualization during the playback of the video. Additionally, the user can generate a *motion-compensated heat map*, a novel technique that uses optical flow information to bundle data points at observed objects (Section 5.5).

Scan Paths

The individual history of each participant's gaze can be visualized by scan paths. 2D and 3D scan path visualizations can be found in different variants, as mentioned in Section 2. We integrated the scan paths of each viewer in our design. They can be enabled individually to investigate the viewing behavior.

Areas of Interest

With AOIs, statistical analysis of the data is possible. Common eye-movement metrics such as fixations per AOI or percentage of participants fixating an AOI can be used to retrieve objective information [30]. Generally, the analysts have to define regions that they want to examine with an appropriate metric. Applying clustering algorithms to recorded eye-tracking data of video material with unknown AOIs provides valuable information about the regions that have been examined by the viewers and might be of interest for the analyst. Figure 3 shows an example where the attention of many viewers was concentrated on a driving car. The number inside the AOIs (see also Figure 2) provides information about the cluster size (see Section 5.4).

All of these methods are integrated in our design and can be enabled individually.

5 Extended Analysis

This section describes modified, extended, or new methods that are included to facilitate the analysis of eye-tracking data for dynamic stimuli.

5.1 Space-Time Cube Visualization

Recording eye-movement data over time generates spatiotemporal data that can be analyzed in various ways. Commercial analysis tools

Fig. 3. Areas of Interest: Axis-aligned boxes represent regions of potential interest. Yellow dots (as marked by the white arrow) represent the gaze points of the viewers.

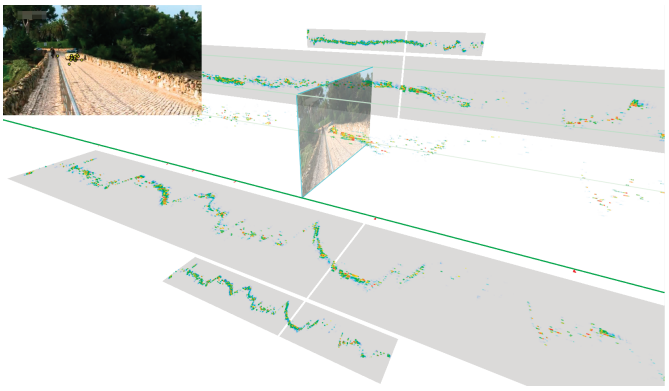

Fig. 4. Data point representation: Moments of high attentional synchrony are clearly visible and motion signatures can be recognized. Additionally, the wall projections provide 2D information of the data as well as an overview of the current position in the video.

provide visualization methods that are overlaid on top of the original stimulus and can be watched as a video. Especially the investigation of long video sequences becomes a time-consuming task with these methods. As an alternative, the static representation of spatiotemporal data within an STC reduces the effort to find time sequences of potential interest. Patterns of synchronous eye movement, as well as the existence and number of potential AOIs can easily be recognized. By providing a freely rotatable 3D visualization, the analyst can explore the data in its original domain.

A slice inside the STC represents the current video frame. Its position is freely rotatable and movable to investigate the data around it. With the video controls, the analyst can navigate through the video by using the time-slider, frame-wise navigation, shot boundary frames, or the playback function. Changing the frame position translates the STC relative to the video frame slice along the time axis, providing an easy method to analyze selected time-spans in the video. In the context of video analysis, the time axis typically shows the highest visual expansion. Therefore, zooming and scaling of the time axis enables the user to explore the data as an overview as well as in detail.

3D visualizations can be afflicted with perceptual issues resulting from occlusion, distortion, and inaccurate depth perception. To address these problems, we provide the user with the possibility to adjust the camera individually in order to resolve possible occlusions in the STC. We also adapted the idea of 2D wall projections (Figure 4) from ExoVis, introduced by Tory and Swindells [42]. With an adjustable scale and distance to the STC, the walls represent 2D overviews of the data without being occluded by the main visualization. A slider moves through the walls, to indicate the current position in the video. Each of the two walls can be adjusted to show either the data points (Section 5.2) or cluster projections (Section 5.4). Two small walls provide an overview of the whole video, independent from the current time-scale.

5.2 Data Point Representation

Eye-tracking data, provided as raw measurement data or as prefiltered fixations, can be mapped to video frames, corresponding to the time-stamp of the data. This provides a maximal set of data points per frame equal to the number of recorded viewers. However, due to saccades or measurement problems, a subset of the data points is usually not available. Bearing this fact in mind, the visualization is designed to analyze eye-tracking data of numerous viewers simultaneously, providing enough data points for interpretation.

Presenting the data points in the STC already gives an impression of the data distribution and especially of sequences with similar eye-movement. Attentional synchrony can indicate events of high saliency.

By determining the distance d of each point to the center of mass per frame, we can calculate a value $v(d) = e^{-0.5\left(\frac{d}{\sigma}\right)^2} \in [0,1]$. The value v defines the transparency and the color of a data point. The same color mapping as for the heat maps (Section 4) is used. By reducing the kernel size σ in the parameter controls, sparse data points in the space-time visualization fade out, facilitating the identification of dense regions.

Figure 4 shows an example scene. Data points with a red color indicate a distance close to their frame's center of mass. When many viewers simultaneously looked at a small area, a large number of data points appear red and remain even when the kernel size is reduced. This representation can also reveal motion patterns of objects tracked by several viewers. However, sequences with sparse data could also be interesting to examine. To this end, the cluster representation can be used (Section 5.4).

5.3 Shot Boundary Detection

Depending on the video material that has to be investigated, finding shots is useful before clustering to find special patterns, related to cuts. Manual annotation of shot boundaries would require the analyst to examine the complete video first, which would be very time-consuming. Therefore, we decided to include a computer vision technique that uses optical flow to find shot boundaries automatically. A shot is defined by a contiguous recording of one or more video frames depicting a continuous action in time and space [29]. Cuts between shots can either be manually marked or, as often preferred, detected automatically by an algorithm. Automatic shot boundary detection is an important pre-processing step in video analysis [38]. From the numerous different approaches that exist, we decided to use optical flow information similar to the method described by Bruno and Pellerin [9] because it is more reliable than the histogram-based approaches. In our implementation, a cut is detected by high disturbance in the optical flow. The optical flow is calculated by the method of Farnebäck [17], provided by OpenCV [1].

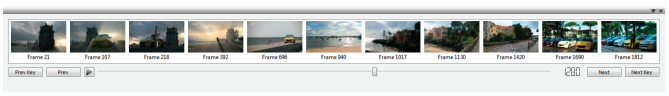

Fig. 5. Time controls: Key-frames allow direct navigation to shot boundaries.

A shot boundary is depicted by a red arrow on the time axis of the STC. In the video controls (Figure 5), a key-frame represents the boundary. By picking one of the key-frames, the space-time visualization jumps to the corresponding position on the time axis, providing an efficient method to examine shot changes. With the shot boundaries defined, the data can now be clustered to extract new information.

5.4 Cluster Analysis

Using clustering algorithms to identify interesting regions in a data set is common practice [23]. To find regions of potential interest in the

recorded gaze data, we choose a clustering algorithm that fulfills the following requirements:

1. **Unknown number of clusters:** The number of data points that have to be clustered can vary, depending on two factors: the number of participants for whom data was recorded; and the length of the stimulus presentation. Defining a proper number of clusters is not intuitive, even if these factors are known. Therefore, we decided to use an algorithm that requires no predefined number of clusters and uses more intuitive parameters.

2. **Parameterization:** A parameterizable clustering approach allows the user to define the granularity of the clusters. Therefore, the adjustable parameters have to be intuitively understandable. The algorithm should depend on two controllable parameters that determine the spatial and temporal extents of the clusters.

The mean shift algorithm performs without a preset number of clusters and can be parametrized in space and time independently. Therefore, it fits the requirements and is suitable for the clustering of the data. Mean shift clustering is widely used for feature space analysis in the field of computer vision [13]. Santella and DeCarlo [35] introduced its application to eye-tracking data.

We adopt their algorithm and extend it to take into account the shot boundaries. To this end, the spatiotemporal gaze points are separated for each shot. After this separation, mean shift clustering is independently executed for each shot to obtain cluster information. This method helps identify special behavior on shot boundaries of a video. As an example, it is known that a center bias exists that is related to cuts [44]. This short time-span of orientation to the center, as well as short periods of gaze reorientation after a cut, can only be investigated by taking cuts into account. Clustering the video without shot boundary information could falsely count these few data points after the cut to the cluster of the previous shot. With the shot boundary information, the few data points become a separate cluster that can be visualized to indicate the described behavior much better.

The found clusters are visualized in two different ways:

- **Cluster hull:** We create axis-aligned boxes around all data points of a cluster for every time-step because they represent the most common convention for AOI representation. The boxes are connected after applying an exponentially weighted moving average [22] to their size, in order to provide a smooth pursuit for the AOI representation. They create a hull around the cluster data points. The spatial extent provides information about changes in the spatial density of the data points in the cluster over time: "thick" cluster hulls correspond to a wide spread of points and, thus, low density—and vice versa.

- **AOI representation:** By projecting the cluster hull of a time step to the corresponding video frame, dynamic AOIs provide information about the distribution of attention on different regions or objects. The cluster size is measured by the number of data points it contains.

5.5 Motion-Compensated Heat Maps

With motion-compensated heat maps, we introduce a new approach to summarize eye-tracking data of dynamic stimuli. A motion-compensated heat map shows high values for observed objects in motion. For example, imagine an object moving through the video from the right to the left side. Assuming all viewers would always observe the object, the resulting heat map of this time-span would show a uniform distribution along the movement trail of the object. In contrast, the motion-compensated heat map would show high values only on the object that was observed, indicating the high amount of attention spent on this object.

The creation of a motion-compensated heat map can be described by a particle tracing in time-dependent fields [46] as follows:

1. The optical flow between consecutive frames in the video is calculated (here, the optical flow for shot detection can be reused, see Section 5.3). It is described by a time-dependent vector field.

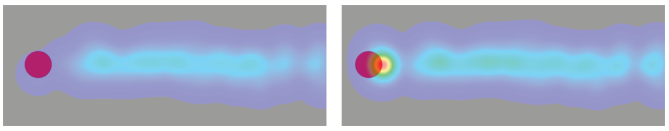

Fig. 6. A conventional heat map (left) and a motion-compensated heat map (right) of a red circle that moves from right to left.

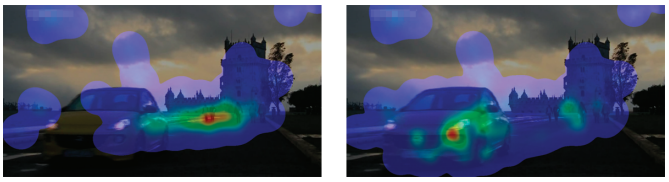

Fig. 7. Car driving from the tower to the left side of the screen: The conventional heat map (left) provides only little useful information about the dynamic AOIs and could lead to misinterpretations because the hot spot lies on two persons. The motion-compensated heat map (right) conveys the information on which object (the car) most of the attention was spent.

2. The analysts have to define a time-span they want to be summarized.

3. A key-frame within this time-span has to be picked. It defines the end for the particle tracing and serves as a representative for the sequence.

4. Each gaze-point within the time-span is traced along the flow until the key-frame position is reached.

5. The traced end positions are used to create a heat map that is blended together with the representative key-frame.

Figure 6 shows a comparison between a conventional and a motion-compensated heat map, created for the same frames of a video. In the video, a red circle moved from right to left. The viewers were asked to follow the circle during its movement. The measured data is distributed along the motion path and no high values remain on the circle itself. The motion-compensated heat map transports the majority of the data points along the optical flow, showing the hot spot with the highest value on the circle itself. The motion path can still be recognized, providing summarizing information about the movement and which object was attended to.

Figure 7 shows a real-world example: Both heat maps represent a short sequence (7 sec) with a driving car and five persons in the background. In this sequence, the car receives most of the attention. Due to the dynamic changes in the scene, the conventional heat map is hard to interpret and the existing hot spot seems to lie on two persons, which would be a misinterpretation. The motion-compensated heat map adjusts the data points along with the object movement, the hot spot lies on the car.

6 Use Cases

Our visual analytics framework provides different methods to identify spatial and temporal regions of potential interest. Our approach is generic and can be used with any image or video. Only the use for individual videos such as recordings of interactive tools or head-mounted eye-tracking devices is limited so far because the recorded data cannot be synchronized easily between viewers, which is a prerequisite to analyzing common eye-gaze behavior of groups of viewers.

6.1 Attentional Synchrony

With the data point representation, users can adjust the kernel size σ (Section 5.2) to filter the data interactively. Time-spans with sparse data fade out as the kernel size is reduced and only dense points remain. With this method, the video can be searched for time-spans

with dense data point distribution, indicating that the viewers' attention was drawn to the same region at the same time. This attentional synchrony can be interesting for various reasons. For example, commercials could be analyzed if the video draws attention to the intended object, or if another object in the scene receives too much attention. Figure 8 illustrates an example of high attentional synchrony. The video shows a commercial that presents pictures of different consumer products (see Section 7.1). The picture of a bottle of eau de toilette leads every viewer to concentrate on the small area of the label, in order to read it. In the data point representation, this short time-span is clearly visible, even when the kernel size is small.

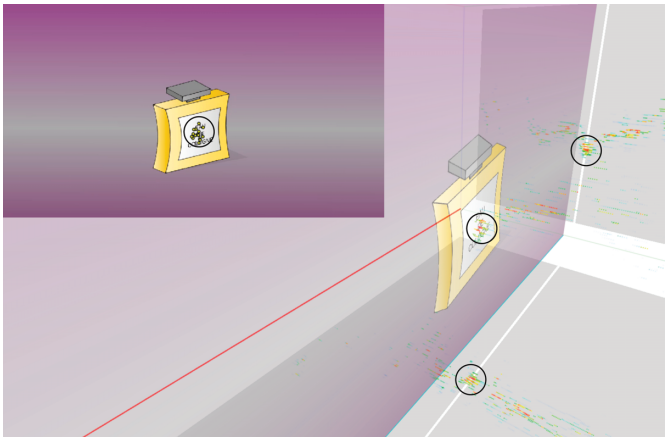

Fig. 8. Attentional synchrony: Dense regions in the data point representation indicate time-spans where all attention was concentrated on one area.

6.2 Multiple AOIs

Attentional synchrony leads to very few clusters during its occurrence. In contrast, no synchrony results in numerous, smaller clusters. This situation might happen, for example, when multiple objects appear on the screen and every viewer investigates each object in a different order. Detecting the objects as separate AOIs is therefore possible, cluster size information and a heat map can then be used to examine, which object is more interesting. Figure 9 shows an example situation. The video is the same as in the previous example; in this part of the video, however, three objects appeared at the same time on the screen. The time-span with these three objects is clearly visible in the cluster representation. Looking at the cluster size information and the corresponding heat map, we can assume that the camera and the coffee machine received more attention than the cell phone.

6.3 Shot Boundary Examination

With the shot boundary frames, the analyst can jump directly to the cuts in a video and examine if changes in the viewers' gaze direction are noticeable. Figure 10 shows an example from the second test video that was used for the analysis task (Section 7.1). The excerpt shows a small cluster of data points at the same position as the cluster before the cut. This indicates that most of the viewers' gazes remained at the old position for 16 frames, and then the viewers began to reorient their gaze in the new shot. This visualization can help identify such latencies, but it could also be used to visualize the center bias after cuts, reported by Tseng et al. [44].

7 Qualitative User Study

To evaluate our design, we performed separate testing sessions with 5 visualization experts; three of them have advanced knowledge of eye-tracking analysis. Each session took about 45–60 minutes to complete, including an introduction and an exploration phase. First, the different views and the design of the system were explained for an example scene and the participants explored it to make themselves familiar with

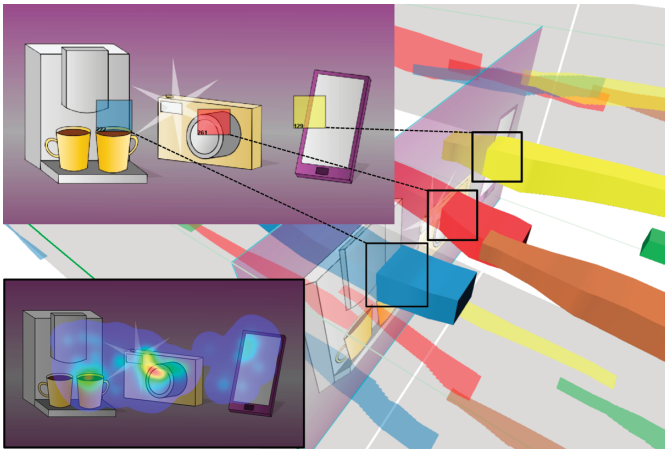

Fig. 9. Multiple AOIs: Three clusters in the same time-span indicate that three different AOIs exist. The corresponding heat map shows the distribution of attention.

Fig. 10. Shot boundary examination: After the cut (red arrow), the viewers' gazes remained at the old position (green cluster).

the design. Possible use cases as described in Section 6 were explained to prepare the participants for the following task:

- **Analysis Task:** To obtain useful and instructive feedback, the participants had to perform an analysis on a new, realistic data set in which they should find the 10 most interesting time-spans, based on their opinion. Attentional synchrony, distribution of attention on multiple AOIs, and viewing behavior at shot boundaries were mentioned as examples that could be of interest. The participants were free to switch between different representations as needed. Each of their findings was listed consecutively by the participants with a frame-span and a description of the discovered event.

During the task, the think aloud method [45] was used to gain insight into the analysis process of the participant. After the task, a questionnaire was used to obtain additional information on the different representations.

7.1 Test Data

The test videos were recorded from regular television program, providing a variety of different aspects to analyze. The recorded material comprised 13 clips from commercials, TV shows, and movies. No video clip was significantly longer than 90 seconds. Each video was upscaled to 720 pixels in height, the width was adjusted respectively. Upscaling was performed to show the videos on a Tobii T60 XL eye-tracker with a 24" screen (resolution: 1920×1200) at a distance of 65 centimeters from the eyes. In separate 20-minutes sessions, 16 volunteers watched the videos consecutively. As viewing behavior can be

related to a given task [40], we instructed all viewers to watch each video attentively and then summarize the main plot of each video. With this task, we reduced inattentiveness and encouraged an explorative viewing behavior.

As an introduction, we showed a credit card commercial where different consumer products appeared successively on the screen. Some of them turned up alone, others together. We used this video to explain the functionality of the visualization and to show the use cases (Section 6). Due to copyright issues, Figure 8 and Figure 9 contain illustrative images of the actual video. For the analysis task, we chose a promotional video for a new car (as seen in all figures except for Figures 6, 8, and 9). This kind of video aims to draw much attention to the product that is promoted. It can be assumed that the cuts and the arrangement of shots were carefully planned by the director. Analytical findings could show to which extent the director's intentions were reflected in the viewing behavior of the viewers.

7.2 Questionnaire

The questionnaire consisted of 9 items, concerning the representation of data points and clusters as well as the STC visualization in general. Additional comments could be listed at the end of the questionnaire. We used a 6-point Likert-scale from "I don't agree" (scale = 1) to "I agree" (scale = 6) with the option to give no answer. Each participant rated all statements. Concerning the data point and cluster representation, the following statements had to be rated:

- **Usability:** *The visualization was helpful to solve the task.*

  The data point representation (mean = 4.40, standard deviation = 1.34) was rated worse than the cluster representation (mean = 5.40, sd = 0.55).

- **Comprehensibility:** *The visualization was easy to interpret.*

  The data point representation (mean = 5.20, sd = 0.84) and the cluster representation (mean = 5.0, sd = 0.71) were rated similarly.

For the general use of the design, three statements were rated:

- **STC navigation:** *The navigation with the STC was easy to understand* (mean = 5.4, sd = 0.55).
- **Key-frame navigation:** *The key-frames were helpful to navigate through the video* (mean = 4.4, sd = 1.52).
- **Projection walls:** *The projection walls were helpful to understand the spatial distribution of the data* (mean = 5.6, sd = 0.8).

Comparing the representation of data points and clusters, the results indicate that the cluster representation was considered more helpful than the data point representation. All participants were able to interpret both representations without any problems.

According to the comments and ratings, the general use of the STC and the video navigation was easy to understand. Identification problems in the 3D visualization could be resolved by looking at the projection walls. Therefore, the projection walls were rated very good by almost all participants. Lower ratings for the key-frame navigation result from the fact that not every participant was interested in direct shot boundary investigation and used the key-frames at all.

7.3 Exploration Strategies

During the task, the participants were asked to think aloud what they wanted to find out and what they did to achieve this. This method provides essential information about the general usability of the design and insight into the individual strategies during the analytical process. For the following description, we refer to individual participants as P1–P5.

Given the same introduction, the participants started with a blank STC and could decide on their own, how to begin the analysis. We identify three different approaches for the first steps in the analytical process:

- **Sequential analysis:** P3 and P4 investigated the static representation only for a short time. Afterward, they began to proceed sequentially through the video. P3 started directly at the beginning and used the slider to navigate through video shot by shot. P3 claimed to be a regular user of video cutting software and that sequential analysis was the usual approach. During the task, P3 mainly concentrated on the bottom wall projections of nearly all clusters at once and the video preview. Only for further investigations, P3 reduced the number of clusters. P4 looked at the cluster overview first, than used the time slider to fast-forward through video. During the analysis, P4 tended to look mainly at the video preview with cluster AOIs activated. To find time-spans with distributed attention, P4 used the cluster representation. Using this strategy, P4 discovered and examined mainly the time-spans in the video that showed landscapes. P4 mentioned that the data point representation seemed very appealing, but during the task P4 almost forgot to use it. Both participants were familiar with eye-tracking, although their experience was mainly restricted to the analysis of static stimuli.

- **Large clusters first:** P1 and P2 investigated the cluster representation first and used the filter function to remove small clusters. Beginning with the largest cluster, they examined successively time-spans that contained them. P1 used the data point representation only for a short time, and concentrated mainly on the clusters. Although P1 had no further experience with the analysis of eye-tracking data, P1 was able to proceed very fast by cluster examination. P2 used the 3D cluster representation in combination with the projected data points on both walls. After investigating the 3 largest clusters, P2 began to search for time-spans with multiple clusters. P2 had experience with eye-tracking analysis.

- **Data point investigation:** P5 began with the data point representation. By looking at the data points in the 3D visualization and on the walls, P5 discovered the motion trails, resulting from the pursuit of the car while it was driving through the scene. After P5 had investigated these time-spans, P5 switched to the cluster representation in 3D but kept the data point projections on both walls. P5 proceeded by investigating time-spans where the data points were more widely distributed.

In general, each participant used the data point representation and the cluster representation during the task, but the main focus was on the clusters. This strategy reflects the results from the questionnaire, because the cluster representation was rated more helpful to search for interesting sequences. P3 and P5 reported that they used both representations to identify cuts and camera pans. The time-scale was mainly used for short time-spans; the participants preferred to see the STC completely during the exploration.

7.4 Findings

The participants were asked to identify the 10 most interesting time-spans. Independent from the individual exploration strategy, the most common findings were:

- **Introduction:** The promotional video was presented within a TV show. At the beginning, the host of the show can be seen for a short time-span, then the promotional video fades in (Figure 11) . The participants discovered a high concentration of the viewers' gazes on the face of the host shortly before the shot boundary. In the following shot, the gaze remained in this region and examined the station-logo first.

- **First appearance of the car:** The first appearance of the car is a very salient event (Figure 12). P1 described it as kind of a salvation from disorientation. In the previous shot, the camera pans to the right, leading to a gaze distribution on the edge of the screen, indicating an exploration of the new objects that appear in the scene. Then, the shot with the car fades in, concentrating all attention on the car surrounded by a halo.

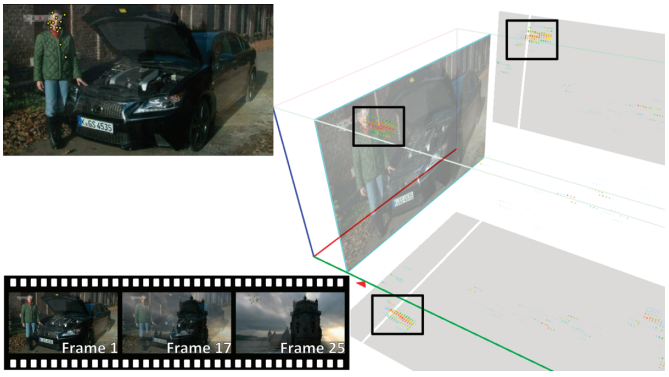

Fig. 11. Introduction: High attentional synchrony on the face and the logo.

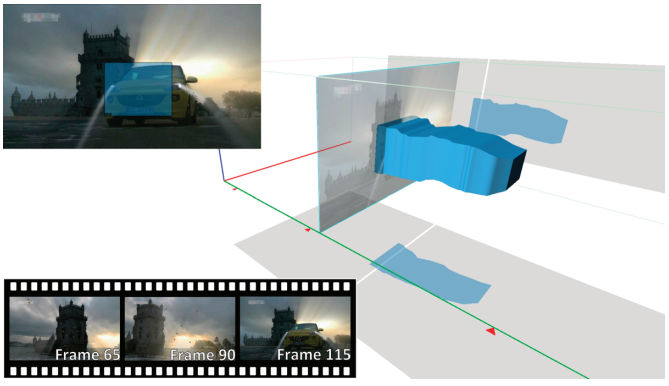

Fig. 12. First appearance of the car: The second largest cluster represents the AOI on the car when it appears for the first time.

- **Tracing the car:** The video contains 3 major shots that show the car driving from one side of the screen to the other. These shots result in a high concentration of the viewers' gazes primarily on the car. This yields a clearly recognizable motion signature in the data point representation (Figure 13). Likewise, some of the largest clusters indicate this motion.

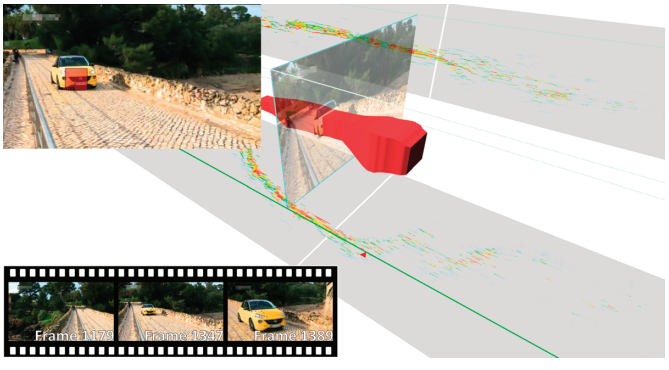

Fig. 13. Tracing the car: The largest cluster describes the tracing of the car. The data points on the walls clearly show the motion signature of this event.

- **Landscape exploration:** Between the shots with the car, Mediterranean landscapes are shown for short time-spans (Figure 14). To identify these scenes, data points and clusters were examined. The clusters were also used to identify the most important AOIs within these scenes.

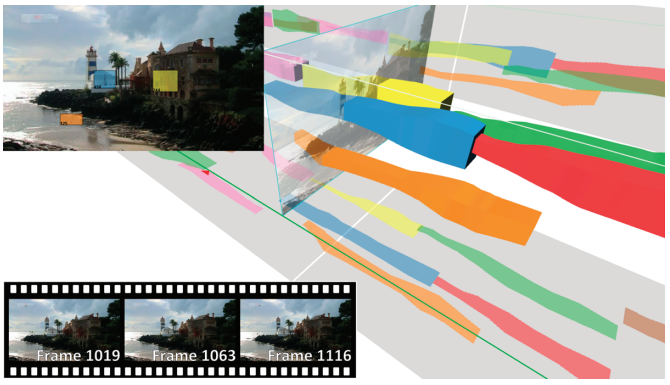

Fig. 14. Landscape exploration: Multiple AOIs were investigated. The lighthouse seemed to get more attention than the other objects in this scene.

- **Appearance of the spokesperson:** Another salient event appears towards the end of the clip (Figure 15). During the video, persons appear only in the background and faces are hard to recognize. Therefore, when the spokesperson finally appears, he attracts much attention.

Fig. 15. Appearance of the spokesperson: When the person fades in, nearly all attention is drawn to his face.

Visit *http://go.visus.uni-stuttgart.de/stva* to watch a supporting video.

7.5 Review Session

To compare the new eye-tracking visualizations with state-of-the-art methods and to identify task-related limitations, we hosted an additional group session with the three experts who had advanced knowledge of eye-tracking analysis. The review session was held after a time lapse (thirteen weeks after the testing session) and took about 45 minutes. The test data from the first session was presented to make the participants familiar with the system again. They were asked to assess the common existing eye-tracking visualizations for video analysis without predefined AOIs, as well as the new ones provided in our system, in terms of their suitability for different tasks. A predefined set of analysis tasks concerning the overview of the data [37], the identification of AOIs, the evolution of attention over time, and the extraction of general viewing strategies of multiple users [5] was used for an initial assessment. The participants were encouraged to state additional tasks along with an assessment of the different visualizations for those tasks. For the assessment, we included common existing visualizations techniques: bee swarm, dynamic heat map, and gaze replay. Motion-compensated heat maps, the data points display, and the cluster visualization represent the new methods for the assessment. Table 1 shows the results of the assessment.

Table 1. Suitability of different eye-tracking visualizations for various analysis tasks. Tasks marked by (*) were added by the participants. Each combination of visualization and task was rated either with '-' : not useful, '+' : useful or '++' : very useful.

Task \ Visualization	Bee swarm	Dynamic heat map	Gaze replay	Motion-compensated heat map	Data points	Cluster
Get a spatiotemporal overview of the data.	-	-	-	+	++	++
Estimate current distribution of attention.	++	+	++	+	-	++
Compare attention on different objects over time.	-	+	+	++	+	++
Find attentional synchrony of multiple viewers.	++	+	++	+	++	++
Find multiple AOIs in a video.	-	+	-	++	+	++
(*)Find multiple user groups.	+	-	+	-	-	-
(*)How long have objects been focused.	-	+	+	++	+	++
(*)In what order have objects been focused.	-	-	+	-	+	+

Compared to the common visualization techniques, the new methods provide a data overview that is most useful with the data point and cluster representations. For the estimation of the current distribution of attention in a specific frame, all visualizations were rated useful except for the data point representation. The participants mentioned that the intersection of points with the video plane in the STC was not sufficient for this task. For the comparison of attention on objects over time, all visualization techniques except for the bee swarm were considered to be suitable. The cluster representation was rated very useful for identifying attentional synchrony as well as multiple AOIs. Finding multiple user groups is a complicated task; here, the new methods were considered not useful. However, the existing visualization techniques provide only little support for this task, too. In this case, additional visualization techniques are required. For the two tasks concerning the focusing on objects, the new visualization methods were considered useful; from the common existing methods, only the gaze replay was considered useful for both tasks.

8 Conclusion

We have presented a new approach to analyzing eye-tracking data of videos or other dynamic stimuli with a space-time visualization in combination with computer vision algorithms. Our design provides multiple views that allow the user to focus on different aspects of the data. The data point representation or the cluster representation provides an overview of the whole video without the need to watch it completely. Filtering clusters by size and data points by spatial density is an effective method to find time-spans of potential interest.

With the expert feedback during the development of our system, we were able to improve the usability of the design. The following qualitative user study led to interesting insights that helped understand how our design was used for analysis tasks. It showed that the participants adopted different strategies for the analysis of the data. This could be related to analytical processes they are used to perform. Our design supports the different strategies and the results indicate that these individual approaches can lead to similar findings in the data. In combination with standard eye-tracking visualizations, the system extends the possibilities for the analysis of eye-tracking data of dynamic stimuli without predefined dynamic AOIs.

For future work, we plan to perform a longitudinal study. The investigation of exploration strategies over several sessions is of special interest. It could show if participants change their strategies, depending on their experience or the type of video material. We also plan to extend the analytical methods of our design. Including additional dynamic AOI information allows the use of common eye-tracking metrics as well as new possibilities for visual data representation. Audio analysis could help interpret events of short attentional synchrony that cannot be explained by visual aspects.

Acknowledgments

This work was funded by the German Research Foundation (DFG) as part of the SFB 716 / D.5 at the University of Stuttgart.

References

[1] OpenCV. http://opencv.willowgarage.com, 2013.
[2] SMI BeGaze eye tracking analysis software. http://www.smivision.com, 2013.
[3] Tobii Studio 3.2. http://www.tobii.com, 2013.
[4] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, London, UK, 2011.
[5] G. Andrienko, N. Andrienko, M. Burch, and D. Weiskopf. Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2889–2898, 2012.
[6] P. Blignaut. Visual span and other parameters for the generation of heatmaps. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 125–128, 2010.
[7] A. Bojko. Informative or misleading? Heatmaps deconstructed. In *Human-Computer Interaction. New Trends*, volume 1, pages 30–39. Springer, 2009.
[8] R. P. Botchen, S. Bachthaler, F. Schick, M. Chen, G. Mori, D. Weiskopf, and T. Ertl. Action-based multifield video visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):885–899, 2008.
[9] E. Bruno and D. Pellerin. Video shot detection based on linear prediction of motion. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 1, pages 289–292, 2002.
[10] 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 Transactions on Visualization and Computer Graphics*, 17(12):2440–2448, 2011.
[11] R. Carmi and L. Itti. Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46(26):4333–4345, 2006.
[12] M. Chen, R. Botchen, R. Hashim, D. Weiskopf, T. Ertl, and I. Thornton. Visual signatures in video visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1093–1100, 2006.
[13] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
[14] A. Duchowski and B. McCormick. Gaze-contingent video resolution degradation. In *Proceedings of Photonics West'98 Electronic Imaging*, pages 318–329, 1998.
[15] A. Duchowski, M. Price, M. Meyer, and P. Orero. Aggregate gaze visualization with real-time heatmaps. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 13–20, 2012.
[16] A. T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, 2002.
[17] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian conference on Image analysis*, pages 363–370, 2003.
[18] P. Gatalsky, N. Andrienko, and G. Andrienko. Interactive analysis of event data using space-time cube. In *Proceedings of the Eighth International Conference on Information Visualisation IV*, pages 145–152, 2004.
[19] 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.
[20] M. Harrower and C. Brewer. ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[21] 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.

[22] J. S. Hunter. The exponentially weighted moving average. *Journal of Quality Technology*, 18(4):203–210, 1986.

[23] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[24] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual analytics: Scope and challenges. In *Visual Data Mining*, volume 4404 of *Lecture Notes in Computer Science*, pages 76–90. 2008.

[25] F. C. Li, A. Gupta, E. Sanocki, L.-w. He, and Y. Rui. Browsing digital video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 169–176, 2000.

[26] X. Li, A. Çöltekin, and M.-J. Kraak. Visual exploration of eye movement data using the space-time-cube. In *Proceedings of the 6th International Conference on Geographic Information Science*, pages 295–309, 2010.

[27] P. Marchant, D. Raybould, T. Renshaw, and R. Stevens. Are you seeing what I'm seeing? An eye-tracking evaluation of dynamic scenes. *Digital Creativity*, 20(3):153–163, 2009.

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

[29] N. V. Patel and I. K. Sethi. Video shot detection and characterization for video databases. *Pattern Recognition*, 30:583–592, 1997.

[30] A. Poole and L. Ball. Eye tracking in HCI and usability research. In *Encyclopedia of Human Computer Interaction*, volume 1, pages 211–219. Information Science Reference, 2006.

[31] G. Rakoczi and M. Pohl. Visualisation and analysis of multiuser gaze data: Eye tracking usability studies in the special context of e-learning. In *Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies*, pages 738–739, 2012.

[32] M. Raschke, X. Chen, and T. Ertl. Parallel scan-path visualization. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 165–168, 2012.

[33] J. Roberts. State of the art: Coordinated multiple views in exploratory visualization. In *Proceedings of the International Conference on Coordinated and Multiple Views in Exploratory Visualization*, pages 61–71, 2007.

[34] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 71–78, 2000.

[35] A. Santella and D. DeCarlo. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 27–34, 2004.

[36] Y. Sawahata, R. Khosla, K. Komine, N. Hiruma, T. Itou, S. Watanabe, Y. Suzuki, Y. Hara, and N. Issiki. Determining comprehension and quality of tv programs using eye-gaze tracking. *Pattern Recognition*, 41(5):1610–1626, 2008.

[37] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[38] A. F. Smeaton, P. Over, and A. R. Doherty. Video shot boundary detection: Seven years of TRECVid activity. *Computer Vision and Image Understanding*, 114(4):411–418, 2010.

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

[40] B. Tatler, N. Wade, H. Kwan, J. Findlay, and B. Velichkovsky. Yarbus, eye movements, and vision. *i-Perception*, 1(1):7–27, 2010.

[41] G. Tien, M. S. Atkins, and B. Zheng. Measuring gaze overlap on videos between multiple observers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 309–312, 2012.

[42] M. Tory and C. Swindells. Comparing ExoVis, orientation icon, and in-place 3D visualization techniques. In *Proceedings of Graphics Interface*, pages 57–64, 2003.

[43] H. Y. Tsang, M. Tory, and C. Swindells. eSeeTrack – visualizing sequential fixation patterns. *IEEE Transactions on Visualization and Computer Graphics 16*, 6(6):953–962, 2010.

[44] P.-H. Tseng, R. Carmi, I. G. Cameron, D. P. Munoz, and L. Itti. Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, 9(7):1–16, 2009.

[45] M. W. Van Someren, Y. F. Barnard, and Sandberg. *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. Academic Press, London, 1994.

[46] D. Weiskopf and G. Erlebacher. Overview of flow visualization. In C. D. Hansen and C. R. Johnson, editors, *The Visualization Handbook*, pages 261–278. Elsevier, Amsterdam, 2005.

[47] D. S. Wooding. Fixation maps: quantifying eye-movement traces. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 31–36, 2002.