Evaluation of Fast-Forward Video Visualization

Markus Höferlin, Student Member, IEEE, Kuno Kurzhals, Benjamin Höferlin, Student Member, IEEE, Gunther Heidemann, Member, IEEE, and Daniel Weiskopf, Member, IEEE Computer Society

Abstract—We evaluate and compare video visualization techniques based on fast-forward. A controlled laboratory user study (n = 24) was conducted to determine the trade-off between support of object identification and motion perception, two properties that have to be considered when choosing a particular fast-forward visualization. We compare four different visualizations: two representing the state-of-the-art and two new variants of visualization introduced in this paper. The two state-of-the-art methods we consider are frame-skipping and temporal blending of successive frames. Our object trail visualization leverages a combination of frame-skipping and temporal blending, whereas predictive trajectory visualization supports motion perception by augmenting the video frames with an arrow that indicates the future object trajectory. Our hypothesis was that each of the state-of-the-art methods satisfies just one of the goals: support of object identification or motion perception. Thus, they represent both ends of the visualization design. The key findings of the evaluation are that object trail visualization supports object identification, whereas predictive trajectory visualization is most useful for motion perception. However, frame-skipping surprisingly exhibits reasonable performance for both tasks. Furthermore, we evaluate the subjective performance of three different playback speed visualizations for adaptive fast-forward, a subdomain of video fast-forward.

Index Terms—Video visualization, adaptive fast-forward, controlled laboratory user study.

1 INTRODUCTION

Video visualization, as an integral part of multimedia visualization and multimedia visual analytics [13], has recently been attracting much attention because it addresses today’s and future data-analysis challenges that come with the enormous amount of video contents from widely available low-cost cameras, internet-connected databases, or personal data collections; see Borgo et al. [5] for a recent survey of techniques for video visualization.

We focus on one aspect of video visualization: the effectiveness of fast-forward video presentation, where the output is a time-compressed video. Fast-forward video visualization is a typical example of animated visualization in which time of the original data (here, the raw video stream) is mapped to animation time; the time-to-time mapping may be nonlinear, for example, for adaptive control of the speed of fast-forward. We consider fast-forward video visualization as the first step of temporal aggregation or abstraction applied to the raw video. Therefore, it readily fits in visual exploration strategies such as the visual information-information seeking mantra [33] or the visual analytics mantra [22]. In top-down exploration, fast-forward video playback can be seen as a stage of data drill-down right before the raw video material is watched. It can also be employed in bottom-up processes, for example, in early stages of information foraging and filtering [30]. In this context, it is an early-stage abstraction of the video with a representation close to the raw data, such as the space-time volume rendering of video [11] or the VideoPerpetuoGram [7]. Besides video fast-forward, the visualizations we evaluate could also be applied to other animated content, such as interactive or animated visualizations (e.g., [19, 31]). Playback-speed of such animations, which usually show between 15 and 20 frames per second [2], can often be steered interactively, using slow motion or fast-forward [37].

In general, we can distinguish two types of video fast-forward: conventional fast-forward, which plays the video at constant pace, and adaptive video fast-forward, which adjusts the playback speed for each frame, according to some measure of relevance. By playback-speed adaption, annoying or irrelevant parts (e.g., static parts) of the video are reduced while emphasis is put on relevant parts (e.g., parts of high activity) — provided that the relevance measure matches the users’ expectations. Typical measures used to rate the relevance of video frames are: optical flow and visual complexity [27], similarity to target clip [28], motion, semantics, and user preference [12], information theory [20], and visual attention [21]. Contrary to conventional fast-forward, the users are more confused in estimating the speed of an object in videos accelerated by adaptive fast-forward [20]. The adaptive method is especially suitable for unedited material like video surveillance footage. In this paper, we focus on the evaluation of fast-forward visualization regarding video visualization for surveillance applications. In the following, we describe the main challenges in this context.

Due to the physical constraint of video display devices (usually 60–200 Hz), higher accelerations are typically achieved by discarding frames (frame-skipping). Frame-skipping at normal playback-speed (in this context termed time-lapse) affects the human perceptual performance. Keval and Sasse [24] experienced a strong decrease
of crime detection performance for time-lapse video presentation in their experiment and Scott-Brown and Cronin [32] pointed out that video in time-lapse format disrupts motion perception, and thus, increases change blindness. The authors of both works illustrate the importance of proper object/event identification and motion perception for video surveillance and the challenges that may arise by time-lapse video. However, the psychological influence of frame-skipping in fast-forward scenarios has not been investigated yet and remains an open research question.

To overcome the issues of frame-skipping a temporal blending method was proposed by Höferlin et al. [20]. From the perspective of computer graphics, temporal antialiasing by temporal blending is usually called motion blur [35] (see Navarro et al. [26] for a review of the state-of-the-art of motion blur rendering). In this paper, we compare the performance of frame-skipping, temporal blending of successive frames, and two new variants of visualization methods for video fast-forward: object trail visualization and predictive trajectory visualization.

The last one is inspired by storyboarding techniques [16] and Augmented Keyframes [10], which use arrows to show movements in still frames. The evaluation covers user performance and user experience according to the taxonomy of Lam et al. [25]. In the evaluation, we focus on the trade-off between motion perception and object detection and identification, to account for the main challenges of fast-forward visualization in the context of video surveillance. Although the perception of animation has been playing an important role in visualization research (see, for example, the textbook by Ware [35]), our work is—to the best of our knowledge—the first user study to evaluate different visualization techniques for compressed-time rendering of motion images.

Since playback-speed adaption of adaptive video fast-forward methods hinders correct object speed estimation in video, users must be made aware of the adaption factor used for adaptive video fast-forward. Communication of the current playback-speed should not distract attention from the video and should be presented in a way that matches users’ expectations. To convey the playback-speed information, a speedometer was utilized by Cheng et al. [12]. Besides a modified version of this speedometer, we introduce two visual representations of the playback-speed (color frame, analog VCR fast-forward) and evaluate their performance in terms of subjective effectiveness, level of distraction, and workload.

2 Fast-Forward Video Visualization Approaches

In general, fast-forward video visualizations have the objective to communicate the information of a certain number of frames from a source video, \( n_{src} \), within \( n_{dst} \) frames in the destination video. If we use the same frame rate for the destination video that we have in the source video, then the relation between these quantities depends solely on the fast-forward acceleration factors \( a_i \), which can vary frame-wise in the case of adaptive fast-forward: \( n_{dst} = \sum_{i=1}^{n_{src}} 1/a_i \).

2.1 Frame-skipping

The typical approach to boost the playback speed in video fast-forward is to discard as many frames as required to obtain the desired acceleration factor [20]. In detail: The \( j \)-th destination video frame \( f_{dst}^j \) is the \( i \)-th frame of the source video. The indices \( i \in I \) are determined by Eq. 1. All other frames are skipped (see Figure 2).

\[
I = \left\{ i \mid \exists j: j \leq \sum_{k=1}^{i-1} \frac{1}{a_k} \land j > \sum_{k=1}^{i} \frac{1}{a_k} \right\}
\]  

(1)

Since the original video frames are displayed in frame-skipping, the appearance of objects is preserved and should be as well observable as in the original video (see Figure 1a). Another advantage is the low computational cost. If an appropriate video codec is chosen and the number of keyframes is adequate, the video can nearly be accelerated without limits: the number of frames considered for visualization remains constant.

2.2 Temporal Blending

The temporal blending approach [20] was developed to alleviate the interruptions of frame-skipping including their perceptual issues. It is motivated by the human visual system, which summates visual stimuli over time [3, 17]. Fast moving objects thus appear blurred (see Figure 1b). This effect can be observed, for example, by swinging torches at night, as the integration time of the human eye depends on the luminance of the environment. Motion streaks from such blurring can support the perception of motion direction, as reported in psychophysical experiments [14]. Temporal blending emulates the blurring effect by integrating the frames in a similar way. The integration time for visualization depends on the acceleration factors \( a_i \).

In detail, each source frame \( f_{src}^i \) is added with the weight \( 1/a_i \) to the destination frame \( f_{dst}^j = \sum \frac{1}{a_i} f_{src}^i \) until the sum of weights is one \( (\sum \frac{1}{a_i} = 1) \). To satisfy this, the weight of a source frame \( 1/a_i \) may be split and the source frame will affect multiple destination frames (see Figure 3).

Typically, video frames are captured with gamma encoding, and are decoded by the monitor while displaying. To integrate the frames in a physically correct way, the original video frames are first gamma-decoded to a linear color space (i.e., from sRGB to linear RGB). After
blending in the linear color space, the resulting frames are encoded again, since the monitor devices require gamma-encoded input. The whole rendering process is depicted in Figure 3. In contrast to frame-skipping, temporal blending communicates object movement by integrating all available frames. Nevertheless, object identification may become difficult due to their blurred appearance (see Figure 1b).

2.3 Object Trail Visualization

Object trail visualization has been developed to combine the advantages of frame-skipping (object identification) with the support of motion perception of temporal blending: we visualize multiple entities of objects simultaneously with increasing transparency (see Figure 1c). The appearance is inspired by Dynamic Stills [9], where a whole video clip is summarized in one static image. The difference is that the duration we summarize is much shorter and used for each destination frame. This effect is also known from Microsoft’s pointer trails that can be used since Windows 3.1 or from high-density cursors [4] to enhance the visibility of mouse movement. We render older entities of an object with higher transparency, similar to the Salient Video Stills [34].

The main difference between object trail visualization and the two mentioned approaches (Dynamic Stills and Salient Video Stills) is that we do not summarize whole sequences in a single image. Object trail visualization creates many summarizations: one for each destination frame.

The rendering process has two stages (see Figure 4), which can be computed in parallel: frame blending and object enhancement. The first stage is dedicated to movement perception; the second stage facilitates object identification.

**Stage 1: Frame Blending.** A particular number of frames \((m-1)\) are blended. The distance \(k\) between the frames is determined according to the acceleration factors \(a_o\). The blended image is calculated by

\[
j^{\text{blend}} = \sum_{o=1}^{m-1} w_o \cdot j^o_{\text{frame}},
\]

where \(l = i - o - k\) and the normalized weights \(w_o\) of older frames are reduced according to \(w_o = \frac{2^o}{\sum_{o=1}^{m-1} 2^o}\) (see also Figure 4).

**Stage 2: Object Enhancement.** The blended image is superimposed by objects present in the recent frame. First, we extract the background image by the running Gaussian average method [29], then we calculate a foreground mask by background subtraction and thresholding, and finally apply morphological operators.

Besides that, we highlight the object movements by additionally blending the object masks of the previous frames semi-transparently: older entities receive a higher luminance here (see Figure 1c). Optionally, the movement emphasis can be disabled (as in Figure 4).

2.4 Predictive Trajectory Visualization

We have designed predictive trajectory visualization—similar to object trail visualization—to allow for good object identification and motion perception. It uses abstract trajectory illustration to improve the perception of moving objects. A special characteristic is that the visualization contains movement information from past and information about future object movements alike.

In detail, it consists of a keyframe that is chosen in the same way as in frame-skipping, a tail for each object that depictions motion from the past, and arrows to forecast the objects’ positions in future (see Figure 1d). The design decision to use arrows to indicate future movements is founded in related work of flow visualization: arrows are widely-used to indicate motion [36]. The rendering process (illustrated in Figure 5) extracts and tracks objects with their bounding boxes. This can be done by standard computer vision pipelines. Since object extraction and tracking is not in the focus of this paper, we refer to the survey of Yilmaz et al. [38].

In predictive trajectory visualization, objects are visible for a longer duration: their arrows are visible before they enter the scene. Thus, the observer can identify the objects earlier. However, this benefit comes at the expense of visual clutter, which may occur in videos with many moving objects.

3 Adaptive Fast-Forward Playback Speed Visualization

In adaptive video fast-forward, the playback velocity of each frame varies according to a measure of relevance. A user study by Höferlin et al. [20] showed that it is difficult to differentiate between variations in video playback speed and changes of the movement velocity of objects. To interpret video material correctly, it is indispensable to estimate the velocities of objects. An event including a moving person, for example, would be interpreted completely different if the observers realize that the person is running or if they imagine that the person is walking. To avoid such misinterpretations, the participants of the user study “suggested to add visual feedback to increase awareness of play- back speed” [20].

In this section, we discuss three possibilities to indicate the playback speed of adaptive fast-forward: the speedometer (Figure 6, left), the color frame approach (Figure 6, center), and the analog VCR fast-forward visualization (Figure 6, right).

3.1 Speedometer

The idea to communicate the velocity of adaptive fast-forward by a speedometer was introduced by Cheng et al. [12]. They show a needle that swivels to the current playback speed as well as the numeric value. We adopt this visualization and enhance it by color mapping (see Figure 6, left). Therefore, a heated body scale color map is applied, which is appropriate for ordinal data [6]. The applied variant is the Matlab’s color map hot, depicted in Figure 6 in the bottom center. The speedometer communicates the playback speed by a metaphor known from cars. Since the peripheral vision has low spatial resolution compared to the fovea [18], identifying the exact position of the needle or reading the numeric speed value requires visual focus of the observer. Since attention should be given to the video, we add a half-circular area at the center of the speedometer that shows the current speed by the color from the color map. The area is larger than the thin needle and the displayed digits, and thus, is better identifiable by peripheral vision.

3.2 Color Frame

In the color frame visualization, the video is enclosed with a thick frame that is inked according to the playback velocity (see Figure 6,
Fig. 5. Rendering process of predictive trajectory visualization. The current frame $i$ is selected as in frame-skipping. For tail rendering (left), the edge points of the objects' bounding boxes (green points) are used from the current frame $i$, and the two previous frames $i-k$ and $i-2k$. For arrow rendering, the center bottom points of the objects' bounding boxes (yellow points) are used from the current frame $i$, and the two future frames $i+2k$ and $i+4k$ (doubled frame distance). The distance $k$ can be adapted to the playback speed (here: $k = 5$). Each group of three temporally shifted control points is used to create a quadratic Bézier curve. The tail is rendered by two surfaces that show a semi-transparent fading white gradient. Each is defined by the two vertical Bézier curves coming from the top and bottom vertices of one of the bounding box sides. The arrow body (defined by its Bézier curve) is superimposed by an arrow head pointing in the direction of motion.

Fig. 6. The three proposed adaptive fast-forward playback speed visualizations. Left: speedometer; center: color frame; right: analog VCR fast-forward. The color scheme is depicted in the bottom center.

The applied color map is identical with the one used for the speedometer. We implemented and included this visualization in the user study since it does not require visual focus, and thus may reduce the required attention to assess the playback speed. In contrast to the speedometer, only a coarse impression of the playback speed can be provided by the color frame.

3.3 Analog VCR Fast-Forward

The third playback speed visualization, analog VCR fast-forward, is based on the characteristic horizontal distortion lines that occur for fast-forwarding analog video data (see Figure 6, right). We map the playback speed to the amount and thickness of the distortion rows, as shown in Figure 7. Horizontal distortion lines are generated by shifting the pixels inside the black rectangles of Figure 7 into a random direction (left, right) with a random magnitude (denoted by horizontal arrows): $\Delta_x = -\text{rand()} \mod 2 \cdot (\text{rand()} \mod \Delta_x)$, with the maximum amplitude $\Delta_x = 10$ px. The shifting direction and magnitude vary between frames. The playback speed is mapped in 10 levels to the distortion width (up to 3) and the amount of distortion rows (up to 3), where the absence of distortion lines indicates original playback speed.

In contrast to the approaches using color mapping, the users do not need to learn any color scheme. The visualization uses a known metaphor, and thus, can be interpreted without training. Similar to the color frame visualization, the playback speed can be recognized without focusing on particular screen regions. Nevertheless, the visualization can be characterized as qualitative visualization since it is difficult to identify the exact playback speed, and the video signal is distorted, which may negatively influence the perception of the video.

4 User Study

In the user study, we compare the four fast-forward visualizations in terms of object identification (measured objective and subjective) and motion perception (subjective) and the three playback speed visualizations in terms of playback speed feedback (subjective). We used three different videos in which we partially added artificial search targets. We use a within-subjects design.

4.1 Hypotheses

We expect a trade-off between support for object identification and motion perception for the fast-forward visualizations and a trade-off for the speed visualizations between conveying precise information and inducing less distraction. Thus, we test the following hypotheses:
HÖFERLIN ET AL: EVALUATION OF FAST-FORWARD VIDEO VISUALIZATION

Hypothesis 1 – Object identification. We expect that frame-skipping shows the best results in supporting object identification because it preserves the original video frames without distortion or superimposed information inducing visual clutter. This hypothesis is also motivated by findings from perception research that indicate that there is no active deblurring mechanism for motion perception in the human visual system [8]. Although object trail visualization and predictive trajectory visualization augment information and modify the video frames, we expect them to show good results: at high playback speed, the time an object is depicted at a particular position lasts longer for the object trail visualization (previous entities fade out), which should improve identification, and predictive trajectory visualization highlights objects, which attracts the attention of the users. However, too many objects may cause visual clutter. Temporal blending may show the worst results: motion blur at high playback speed may hinder correct identification of search targets.

Hypothesis 2 – Motion perception. By discarding frames, frame-skipping may disrupt motion perception. Both temporal blending and object trail visualization additionally merge objects of the current frame with previous instances, which may improve motion perception. Predictive trajectory visualization includes information about the past and future movement of objects. Therefore, we hypothesize that it shows the highest performance in motion perception.

Hypothesis 3 – Playback speed feedback. We expect the speedometer to be the most efficient visualization for playback speed feedback due to its detailed information representation. However, since it requires visual focus, it may be most distracting. The color frame visualization has least influence on the video scene and shows only marginal distraction. The color mapping can represent a wide range of possible fast-forward accelerations; thus, it may be the best trade-off between visual distraction and the accuracy of conveyed information. Estimating the pace of playback via analog VCR fast-forward does not require visual focus. However, this benefit comes at the expense of distorted video rendering and a rather coarse granularity of speed feedback. We expect it to be the least preferred visualization.

We separately tested each hypothesis, by a specifically designed task.

4.2 Stimuli and Tasks

In our experiments, the independent variable of interest is the visualization type. Object identification and movement perception were tested with the fast-forward visualizations frame-skipping, temporal blending, object trail visualization, and predictive trajectory visualization. Playback speed feedback was evaluated with the speed visualizations speedometer, color frame, and analog VCR fast-forward. Each participant had to perform each of the three tasks T1, T2, and T3. Each of the three tasks was designed to check one of the three hypotheses.

We used three videos in this study to compare the visualizations with each other. The videos are benchmark datasets used in different areas, such as tracking, and have in common that they show real world scenarios and originate from CCTV cameras. Video V1 (from the BEHAVE dataset\(^1\), resolution: 640 × 480 px, 25 fps) and video V2 (from the AVSS dataset\(^2\), resolution: 720 × 576 px, 25 fps) are used for the two tasks concerning the fast-forward visualizations (T1 and T2). We choose these videos to account for different amount of activity and complexity of movement. The first video (V1) is less complex and less populated than V2, it includes less perspective distortion, and covers a smaller area. Video V3 (from the multi-camera i-LIDS dataset\(^3\), resolution: 720 × 576 px, 25 fps) is used for the playback speed visualization task. This video is chosen since it was also used to evaluate

\(^1\)BEHAVE Interactions Test Case Scenarios
http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/

\(^2\)AVSS 2007

\(^3\)i-LIDS multi-camera tracking scenario dataset
http://www.homeoffice.gov.uk/science-research/hosdh/i-lids/

relevance measures for adaptive fast-forward [20, 21]. The tasks the participants had to perform were:

- T1 – Object identification (objective and subjective). The participants had to watch V1 accelerated by factor 20 and V2 accelerated by factor 10, using each of the four visualizations. These acceleration factors were determined during the pilot study. The stimuli had lengths of 45 s and 106 s, respectively. As search target, animated cartoon figures were artificially inserted into the videos, walking through the scene like normal persons (see Figure 8). The participants had to detect and recognize the cartoon figures. The brightness and contrast of the cartoon figures were adapted to the video to avoid pre-attentive perception. Comic figures have the advantage that they can be easily rendered into an existing scene without recapturing the video and feature a good recognition value. The detection had to be indicated by pressing a buzzer. A detection was counted to be correct, if the buzzer was pressed while the cartoon figure was present in the sequence or had left not earlier than two seconds before. Each appearance of the cartoon figure was only counted once. Additionally to this objective measure, the visualizations had to be rated by the participants. The spatial and temporal positions of the cartoon figure varied to avoid learning effects—in total, we used 4 different versions of each video (V1 and V2). The cartoon figure appears 13 times in each version of V1 and 8 times in each version of V2 (the appropriate numbers were determined in the pilot study). To counter-balance the experiment, the presentation order of the visualizations was permuted. To balance the video versions, version 1 of the video was always shown first, then version 2, and so on. This assured that each visualization was presented the same number of times with each version of the video and potential side effects arising from different difficulties of the variants were avoided.

- T2 – Motion perception (subjective). Each combination of pairs of the different visualizations (i.e., six visualization pairs) were presented and had to be compared by each participant. For each pair, two visualization stimuli of 10 s were successively presented, with a pause of 3 s in between. Three of the pairs showed a snippet of V1, for the other three pairs a snippet of V2 was used. The order and pair composition of the presented visualizations was permuted to counter-balance the experiment. The participants had to judge for each of the pairs which performed better in terms of motion perception. There was also the option to rate the performance of the two visualizations as equal. We used paired comparisons since the number of items from which to choose increases the cognitive burden and negatively influences the quality of the results [15].

- T3 – Playback speed feedback (subjective). The participants had to watch V3 in adaptive fast-forward three times, with each of the three speed visualizations applied. After the presentation of the stimuli, the participants had to rate each visualization in terms of effectiveness of playback speed feedback, perceived workload, and distraction from the video content. The presentation order of the visualizations was counter-balanced between the participants.


4.3 Pilot Study

Before conducting the study, we checked the study design by a pilot study with four participants. The pilot study allowed us to identify potential issues with the tasks and stimuli. It turned out that V1 with an acceleration factor of 16 and nine search targets was too easy for the first task. We increased the acceleration factor to 20 and inserted four additional search targets to the scene.

4.4 Environment Conditions and Technical Setup

The user study was conducted in a laboratory isolated from outside distractions. The room was artificially illuminated and only objects relevant for the study were contained inside. The participants were instructed to turn off their mobile phones.

The videos were presented in full screen on a laptop with a 16 inch monitor and a screen resolution of 1366 × 768 px. The distance between the participants and the screen was 50 – 80 cm. A video player was implemented that allowed us to record the user input during the first task, in which the participants used a buzzer to indicate a detection of the search target.

4.5 Participants

The study was designed as within-subjects study and was conducted with 25 participants. Due to the lack of interest of one participant, the test results of 24 participants are considered. Seven of those participants were female and 17 were male. Gender was not considered as relevant factor. The average age was 25.5 years; the youngest participant was 20 and the oldest 31. Most of the participants were students from our university. Six participants studied at other universities or had already graduated. Half of the participants were students of computer science or software engineering. The others had majors in humanities or other subjects of study. All participants were tested with a Snellen chart and had normal or corrected-to-normal vision. The experiment took 45 – 55 min, depending on the particular speed of each participant. Each participant was compensated with EUR 10.

4.6 Study Procedure

The participants were first asked to fill out a questionnaire about their gender, age, and field of study.

Then, a short tutorial (about 6 min) introduced the fast-forward visualizations using snippets of V1. Afterwards, a video snippet with a cartoon figure was shown to the participants to explain task T1. The video snippet was not reused in the tasks and it was presented in original playback speed. The first task was performed using V1. The participants watched all four visualizations and had to identify the cartoon figures. The start of the playback of each visualization was initiated by the participants. Hence, small breaks of individual length were possible between the visualizations.

After completing the task, the participants were asked to fill out a questionnaire about the effectiveness of the visualizations (subjective effectiveness), how comfortable they felt with the visualizations (comfort), how useful they considered each of the visualizations (usefulness), and the effort it made to watch them. For this rating, a 10-point Likert scale was provided. They were also asked which visualization they preferred for performing the task (forced choice). Finally, the possibility was provided to give free comments about the visualizations. The same procedure was repeated with V2.

For task T2, the participants were briefed to pay attention to object movement. The visualizations were presented pairwise, using snippets of V1 and V2. After presentation of a pair, the participants had to decide which visualization supported motion perception better, or if both performed similar. After T2, the subjects were asked to comment on the visualizations.

Before the experiment with the third task was conducted, the different playback speed visualizations were introduced in a short tutorial (about 4 min in length). Then, one stimulus for each of the three visualizations (calculated on V3) was successively presented to the participants. The duration of each stimulus was approximately 1 min. Afterwards, the participants were asked to fill out a questionnaire to judge the playback speed visualizations in terms of the effectiveness of the speed feedback, the effort required to interpret them, and the distraction from video they induced, on a 10-point Likert scale. In the end, they also provided their preferred visualization technique (forced choice) and could comment on all speed visualizations.

There was a “Give Up” option throughout the study; however, it was not used by the participants. The time to perform the tasks was limited by the duration of the videos. There was no time limitation between the videos and the participants decided when to continue.

5 Study Results

In the statistical analysis, we include the results of 24 participants. We test the significance of the results with non-parametric tests since not all results are normal distributed. For statistical computing, we used the software from the R Project [1].

5.1 Results of the Object Identification Task T1

The results of task T1 are divided into the measured effectiveness (objective results) and the results of the questionnaire (subjective results).

5.1.1 Results of the Measured Effectiveness

To measure the performance of the participants in task T1 (object identification), we calculate $F_1$-scores ($F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$) by determining the precision and recall using the buzzer input. The boxplots of the $F_1$-scores of all subjects for each visualization are depicted in Figure 9.

The non-parametric Kruskal-Wallis test shows a significant effect of the visualization type on object identification ($H(3) = 46.63, p < 0.005$).

Post-hoc pairwise U-tests confirm that frame-skipping (median of $F_1$-scores: 0.87) and object trail visualization (0.84) show significantly better results ($p < 0.005$, Bonferroni-corrected for multiple comparisons) than the other two visualizations (temporal blending: 0.67, predictive trajectory visualization: 0.70), respectively. According to the comments of the participants, the latter visualizations suffer from information overload.

Object trail visualization and frame-skipping do not show any significant differences. However, some subjects described the trails to be useful for identifying the objects.

With temporal blending and predictive trajectory visualization, it was more difficult for the participants to identify an object. The test showed no significant differences between these two visualizations. According to the participants’ comments, the objects in the video were too blurry when visualized by temporal blending. Some subjects also had problems with predictive trajectory visualization. They reported that the visualization provided too much information to focus on object identification. Nevertheless, they also mentioned that this visualization was useful in cases where only few objects were present.
As outlined in Hypothesis 1, frame-skipping showed the best results in object identification, and object trail visualization was able to produce similar results. However, we have to reject the hypothesis that predictive trajectory visualization can achieve results comparable to object trail visualization. According to the median, temporal blending performed worst, but it was not significantly worse compared to predictive trajectory visualization.

5.1.2 Results of the Questionnaire

Figure 10 depicts boxplots that summarize the questionnaire results of task T1:

- **Subjective effectiveness.** The Kruskal-Wallis test shows that the visualization type had a significant influence on the rating ($H(3) = 35.66$, $p < 0.005$), and the U-test reveals significant differences ($p < 0.05$, Bonferroni-corrected) between all visualizations except for two pairs: frame-skipping / object trail visualization and predictive trajectory / object trail visualization. The boxplots in Figure 10 (left) show that the subjective effectiveness and the measured objective effectiveness (cf. Figure 9) in object identification exhibit qualitatively similar results. However, in comparison to the other methods, the effectiveness of temporal blending was rated much worse than the objectively measured results indicate.

- **Usefulness.** The visualizations show significant differences (Kruskal-Wallis: $H(3) = 36.91$, $p < 0.005$) in terms of usefulness. Predictive trajectory visualization (median: 5.5) was considered significantly less useful than frame-skipping (median: 7.25, U-test: $p < 0.05$, Bonferroni-corrected). According to the comments of the participants, predictive trajectory visualization was only useful as long as a small number of objects were present in the scene. No significant difference was found between frame-skipping and object trail visualization (median: 7.0). Temporal blending was rated the least useful with a median of 4.0 and with significant differences to all other visualizations (U-test: $p < 0.05$, Bonferroni-corrected).

- **Comfort.** Significant differences (Kruskal-Wallis: $H(3) = 33.51$, $p < 0.005$) were found for the evaluation of comfort among the visualizations. Frame-skipping—as the familiar method to watch fast-forward videos—was considered the most comfortable to watch (median: 7.5). Frame-skipping performed significantly better than predictive trajectory visualization and temporal blending (U-test: $p < 0.05$, Bonferroni-corrected). A quite similar rating with significant difference to temporal blending was received for object trail visualization (median: 7.0; U-test: $p < 0.05$, Bonferroni-corrected). However, there was no significant difference between frame-skipping and object trail visualization. The least comfortable visualization was temporal blending (median: 3.5), which performed significantly worse than the others, except for predictive trajectory visualization (median: 5.5).

5.2 Results of the Motion Perception Task T2

A ranking of the pairwise compared visualizations was generated according to Kendall [23]: A visualization receives one point for each won comparison, and half a point for each draw. The resulting scores of the pairwise results are depicted in Table 1, summarized results in Figure 11. Following the test of Kendall [23] for paired comparisons with ties, the coefficient of agreement $\tau$ shows a significant but relative low accordance between the participants ($\tau = 0.07, \chi^2(6) = 26.90, p < 0.05$).

Predictive trajectory visualization won most of the comparisons, as Hypothesis 2 outlined, and received a score of 91. Surprisingly, frame-skipping is second with 86.5 points. Expected advantages of temporal blending (last rank, 51.5 points) in motion perception, and object trail visualization (third, 59 points) could not be confirmed.

5.3 Results of the Playback Speed Feedback Task T3

The Kruskal-Wallis test showed no significant effects of the visualization type on the given answers. Figure 12 shows the results of the questionnaire for this task. Based on the participants’ free comments...
Table 1. Scores of the pairwise motion perception ranking (task T2).

<table>
<thead>
<tr>
<th></th>
<th>Frame-skipping</th>
<th>Temporal Blending</th>
<th>Object Trail Visualization</th>
<th>Predictive Trajectory Visualization</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame-skipping</td>
<td>-</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal Blending</td>
<td>17</td>
<td>-</td>
<td>20</td>
<td>14.5</td>
<td>15.0</td>
</tr>
<tr>
<td>Object Trail Visualization</td>
<td>17.5</td>
<td>28</td>
<td>-</td>
<td>13.5</td>
<td>31.5</td>
</tr>
<tr>
<td>Predictive Trajectory Visualization</td>
<td>23</td>
<td>33.5</td>
<td>34.5</td>
<td>-</td>
<td>91</td>
</tr>
</tbody>
</table>

and the medians, we summarize these results with the following observations:

- **Effectiveness.** The participants attested analog VCR fast-forward as the highest effectiveness in playback speed feedback (median: 8.0). Many participants explained their rating by their familiarity with such visualization from VCR. In terms of playback speed feedback, the speedometer also showed reasonable performance (median: 7.0).

- **Effort.** The participants rated analog VCR fast-forward as also best according to the median in terms of effort while watching (median: 3.0). The effort required to keep track of the speed with the speedometer was higher according to its median (5.0). Following the comments of the participants, a reason for that originated from the frequent changes of focus between the video and the speedometer. The participants reported that the frequent color changes of the color frame visualization (median: 6.0) resulted in flicker, which led to stress.

- **Distraction.** The visualizations only marginally (according to the medians) differ in terms of distraction. The medians of the visualizations are 5.5 (speedometer), 5.0 (color frame visualization), and 4.5 (analog VCR fast-forward). The participants mentioned the frequent changes of focus between the video and the speedometer as a problem.

For playback speed visualization, 40% of the participants preferred the speedometer in the forced choice question, 36% the analog VCR fast-forward, and 24% the color frame visualization. As assumed in Hypothesis 3, the speedometer was the preferred visualization for speed estimation. However, the color frame visualization was not as good as expected, probably because of the induction of stress by flicker. Also remarkable is that the issue of distortion and coarse playback speed feedback for analog VCR fast-forward was not considered as problematic as we assumed. Nevertheless, the difference between the visualizations is relatively small, thus no significant differences could be detected by Kruskal-Wallis test as mentioned above.

6 **Discussion and Conclusion**

In this paper, we have evaluated different visualization techniques for video fast-forward. Four visualizations, of which two were introduced in this paper, have been compared with each other in terms of their support for object identification and motion perception. Additionally, we have evaluated the performance of three methods for visual playback speed feedback in the context of adaptive fast-forward. Two of the methods compared were introduced in this paper.

The user study exhibited some remarkable results concerning the performance in object identification of the different fast-forward visualizations. Frame-skipping appeared to be the preferred method for object identification and performed also well in motion perception. This result is contrary to the initial hypothesis that motion perception would be negatively affected by discarded frames. Temporal blending failed in both tasks: object identification and motion perception. It was especially surprising that most of the participants rejected this visualization also for the task of motion perception. The object trail visualization showed comparable results to frame-skipping in the object identification task. Nevertheless, the trail of an object did not improve motion perception. Although predictive trajectory visualization provides the most information on object motion (it was ranked best for this task), the additional information results in visual clutter and hinders object identification in crowded scenes. Hence we recommend, depending on the task, using either frame-skipping or predictive trajectory visualization as fast-forward visualization. Due to the dependencies of the visualization on different characteristics of the video stimuli, we plan to evaluate the video visualizations in detail according to the amount of action (i.e., crowded vs. empty scenes) and according to different playback speeds in future work. Especially the results of temporal blending indicated that this visualization depends on the acceleration factor; a reason could be that faster playback causes stronger blur. Another direction for future work is the adoption of predictive trajectory visualization according to the number of objects present in scene to reduce visual clutter. Moreover, switching automatically to the most promising visualization according to the current characteristics of video stimuli may support users.

The results of the evaluation of playback speed visualizations for adaptive fast-forward sum up to: feedback of playback speed was considered best for the speedometer visualization. However, the participants of the study criticized the need to constantly switch the visual focus between the video and the speedometer. The color frame, in contrast, induced stress by visual flicker, which limits its application to cases of non-permanent usage. Analog VCR fast-forward seems to be the best solution to perceive speed feedback while watching video in adaptive fast-forward despite its coarse feedback. Especially in the context of playback speed feedback, the results of the study point out the strong benefit of using established metaphors in visualization. Although analog VCR fast-forward distorts the video signal and the speedometer visualization requires the switch of visual focus, their acceptance was surprisingly high.

The conducted user study focused on real-world videos that originated from surveillance cameras. For the first task, we superimposed the video by an artificial object, a cartoon figure, which should be identified. Although we put much effort in adapting the brightness and contrast of the objects to appear as normal persons, we cannot fully rule out effects from this choice of stimulus. Therefore, future work should also feature real objects for the identification task. One direction for future work is also to generalize the results to completely different video footage, and visualization in general. As mentioned in the introduction, the visualization techniques can generally be applied to interactive and animated visualizations (e.g., [19, 31]) and other animated content. Due to interactive steering of the playback speed of such animations, the playback speed visualizations can be applied as well. For them, we also plan to evaluate further solutions, such as a progress bar or a rotating bar. Such visualizations seem promising due to good motion perception in the peripheral field of view.

**Acknowledgments**

We thank Michael Wörner for voice acting. This work was funded by German Research Foundation (DFG) as part of the Priority Program “Scalable Visual Analytics” (SPP 1335).

**References**


Fig. 12. Boxplots of the results of the playback speed visualizations (task T3).


