Exploring the Need and Design for Situated Video Analytics

Fouad Alallah
University of Manitoba
Winnipeg, Canada
umalallf@cs.umanitoba.ca

Yumiko Sakamoto
University of Manitoba
Winnipeg, Canada
umsakamo@cs.umanitoba.ca

Pourang Irani
University of Manitoba
Winnipeg, Canada
pourang.irani@cs.umanitoba.ca

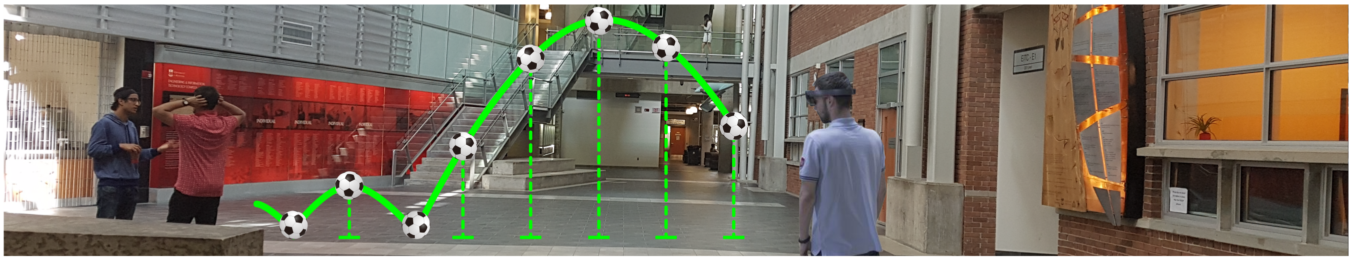

Figure 1: Future prototype design, utilizing situated visualizations based on one example of a participant's drawing.

ABSTRACT

Visual video analytics research, stemming from data captured by surveillance cameras, have mainly focused on traditional computing paradigms, despite emerging platforms including mobile devices. We investigate the potential for situated video analytics, which involves the inspection of video data in the actual environment where the video was captured [14]. Our ultimate goal is to explore the means to visually explore video data effectively, in situated contexts. We first investigate the performance of visual analytic tasks in situated vs. non-situated settings. We find that participants largely benefit from environmental cues for many analytic tasks. We then pose the question of how best to represent situated video data. To answer this, in a design session we explore end-users' views on how to capture such data. Through the process of sketching, participants leveraged being situated, and explored how being in-situ influenced the participants' integration of their designs. Based on these two elements, our paper proposes the need to develop novel spatial analytic user interfaces to support situated video analysis.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; • **Situated Visualization**; • **Information Visualization**; • **User Study**; • **Video Analysis**; • **Sketching**;

KEYWORDS

Situated Analytics, Video Analysis, Augmented Reality, Input Modalities, Head-Worn Displays, In-Situ Data Analytics.

SUI '20, October 31-November 1, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7943-4/20/10…$15.00
https://doi.org/10.1145/3385959.3418458

ACM Reference Format:
Fouad Alallah, Yumiko Sakamoto, and Pourang Irani. 2020. Exploring the Need and Design for Situated Video Analytics. In *Symposium on Spatial User Interaction (SUI '20), October 31-November 1, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3385959.3418458

1 INTRODUCTION

Surveillance cameras are commonly used in public spaces. Video analysis of this footage enhances understanding of scenes captured in terms of temporal, spatial, sequential, relational, and interactional features. Based on the analyses of these pieces, users draw practical/real-world conclusions (e.g., sentencing) about the event seen in the video. Currently, an analyst would sit in front of a monitor and examine the events of interest with video analysis tools. To date, video visual analytics and visualization research has predominantly focused on using non-situated computing paradigms (i.e., analysis on a computer while using classical desktop interfaces) to explore data [18, 19, 25, 28, 30, 33, 38]. However, there is great potential in the benefits of quickly advancing platforms such as Augmented Reality (AR)/ Virtual Reality (VR).

While technologies associated with video analyses are advancing rapidly (e.g., technologies used to make criminal allegations), the video analysis methods/techniques remain unchanged (e.g., using traditional desktop interfaces in office setting). Thus, we propose an alternative video analytical method which integrates video analysis tools and tasks into the real environment of an event in question. Our study explores the potential of this proposed approach, and whether it could improve the reliability of video analysis. The first study compares users' performances in situated video analytics and non-situated analytics. *Revisiting the scene* where the events in question took place should expand the understanding of said events even further. For instance, one can answer some questions only by revisiting scenes to collect clues that video analysis tools may not capture adequately (e.g., accurate spatio-temporal information and details of events that took place partially off camera). In fact, a revisitation step is essential to make accurate judgements about any events. For this, emerging mobile technologies and computing paradigms could possibly amalgamate video analysis and exploration of the actual environment; this could further enhance the outcome of Situated Analytics (SA) by enabling the visualization of actual events in the actual location.

Little is known about the benefit of in-situ exploration of video data since SA is still an emerging area [16, 39]. Results from Study 1 confirm overall improved accuracy when participants perform video analysis in-situ, compared to the non-situated settings. This leads us to the next question. What design considerations are needed to develop situated video analytics interfaces? Thus, we next conduct a design session with end-users to explore situated video visualization potential design. An analysis of participants' sketches present some design take-aways for situated video visualizations that can aid with common video analytic tasks.

Our contributions are threefold. We; 1) conduct an empirical study comparing Situated and Non-situated Video Analytics, 2) report user-generated visualization design and interfaces to support in-situ video analytics tasks. Based on the analysis of visualization, we use our generated knowledge for how to exploit the user's immediate environment to place and represent visualizations [5, 12] to 3) share our takeaways for designing novel user interfaces for situated video analytics technology.

2 RELATED WORK

2.1 Exploratory Video Data Analysis

Video data analyses often start out with an exploratory search to formulate tentative and broad queries. These queries often narrow down the data to be examined. These primary explorations then lead to further questions, hypotheses generation, and ultimately answers to the questions [10, 30]. Several exploratory search tasks have been investigated thus far; e.g., tasks such as person identification [8], object movement measurement [27, 33], motion and pattern analysis [20, 37], and human and object measurement [2, 3, 9].

2.2 Video Visual Analytics and Visualization

Visual Analytics (VA) is defined as "the science of analytical reasoning facilitated by visual interactive interfaces" [40]. Visual video analytics and visualization tools aim to create a visual representation of raw video footage, to visually highlight important spatio-temporal data [4, 31]. Current video visual analytics tools include a variety of approaches consisting of video summarization [24, 32, 34], video content visualization [27, 30, 33, 37], and video interaction and navigation [20–22, 30]. The major purposes of these tools are to: 1) remove redundancies in the analytical step; 2) visualize a summary of the video data; 3) provide alternative 3D representations of individuals and object spatio-temporal information, and 4) introduces novel interaction techniques.

2.2.1 Video Summarization: Video summarization aims at facilitating large-scale video data browsing processes via the selective presentation of meaningful components of an original video [32]. Several examples of video summarization studies highlight the state-of-the-art, along with opportunities for further study and exploration. For example, Video Synopsis tool produces a summary video of the events locations and objects from the original input video by removing spatio-temporal redundancies, and condensing multiple activities, which occurred at different times, into a single, short output video [34]. Another video summarizing technique was proposed to visualize the trajectory data of a moving object [24]. This technique automatically divides a video into segments (clips) based on the detected movements of objects, then, extract trajectory data from each object. Finally, trajectory paths of moving objects are visualized on top of each video segment.

2.2.2 Video Content Visualization: Video summarization techniques are limited to displaying either the original frame or fast playback. Thus, video content visualization has emerged as an alternative solution to this limitation [30]. Video content visualization techniques use visual elements and effects, or a modified version of the original video frame to visually summarize video content. VATAS is a visualization system that enables automated analysis and annotation of movement trajectory events in videos [27]. Video Summagator is a video summarization and navigation tool which visualizes object movement data in a space-time cube [33]. Another method used to visualize object movement data is to extract movement data from a video, then lay the results of the extraction on top of the same video [37].

2.2.3 Video interaction and navigation: Video interaction and navigation techniques aim at supporting video visual analytics via interactive exploration of video content. Several examples of video interaction and navigation demonstrate browsing and exploration techniques for improving the ability to interact with video data. The authors in [21] proposed an interactive video visual analytics tool that visualizes object movement trajectories from a given video. In [22] and [20], the presented tool uses clustering and schematic visualization techniques to visualize moving object trajectories in 2D. In [30], the authors have proposed an interactive visualization tool for surveillance video. Their tool extracts object movement data from videos, then provides three interactive exploration visualizations of the movement data.

2.3 Situated Analytics (SA)

SA takes advantage of virtual spaces by mapping them to the physical environment [16, 17]. SA can be defined as "the use of data representations organized in relation to relevant objects, places, and persons in the physical world for the purpose of understanding, sense-making, and decision-making." [39]. Researchers highlighted important design guidelines related to situated data visualization and interactions [15]. For example, information should be registered on physical objects in such a way that the location and appearance of the information adapts to the changes in the object. Two different types of interactions were highlighted [15]: 1) AR to allow for interaction (e.g., select/deselect to display data) with physical objects, in context with a query; and, 2) analytical interaction to allow to control the exploration and analytic tasks performed on the data registered to a physical object. Unique and novel video analysis tools are likely to see increased research focus as developers take advantage of new case scenarios, made possible through new computing paradigms and modalities.

Video visual analytics and visualization tools were designed primarily for video analysis on non-situated computing paradigms. So far, however, no research has empirically investigated how Situated vs. Non-Situated video analysis differs in terms of users' performance and experiences.

3 STUDY 1: SITUATED VS. NON-SITUATED VIDEO DATA ANALYSIS

We explore whether SA can facilitate video analytic tasks. For this, we conducted a study that observes video-analysis activities in two settings, in-situ (i.e., situated) vs. as traditionally done, at the desk (i.e., non-situated). The dependent variables for each video scenario were based on the exploratory search tasks found in video analysis tools in the literature [2, 3, 7, 9, 20, 26, 27, 27, 29, 33, 35, 37]. Additionally, for an exploratory purpose, we assessed participants' confidence levels for their own judgements.

3.1 Participants

We recruited 40 participants (M = 18 , F = 21 , Other = 1), aged between 18 and 41 years (*M* = 24.70, *SD* = 6.59), from a local university. They were randomly assigned to either Situated or Non-Situated condition. 20% of the participants reported English as their first language, and none of the participants had any language issues throughout the study. All participants reported normal or corrected-to-normal vision.

3.2 Apparatus

For visual analysis, a Microsoft Surface Pro 2 was used in both the Situated and Non-Situated conditions. Its screen size was 10.8 inches (27 cm) by 12 inches (30 cm) and the resolution was 1920 x 1080. To support the participants' mobility during the study in the Situated condition, they were able to switch between laptop mode and tablet mode via a detachable keyboard. Participants were allowed to switch between the two modes as needed. For example, when a participant wanted to explore the scene physically, they would choose tablet mode for a better video viewing experience. Furthermore, participants in both groups were provided with two measurement tools (a ruler and a measurement tape), pen, paper, and a stopwatch to help them answer the analytic tasks. To develop video stimuli (i.e., video clips), a video camera (the Canon HF-M52) was used (See Section 3.4 for detailed video scenarios).

3.3 Method

A two by two mixed between (1: Situated vs. 2: Non-Situated) within (1: On-Camera vs. 2: Off-Camera) design study was conducted. The scenario clips and questions were uploaded on an online survey system, Qualtrics. The general procedure was explained by a research assistant after participants signed the consent form. Their main task was to perform visual analytic to answer questions. Thus, participants were asked to watch video clips on the Microsoft Surface Pro 2 first. Participants were instructed to be as fast and accurate as possible, but no time limit was set. Measurement tools, pens, paper, and a stopwatch were provided to the participants to help them in completing the tasks. The participants in the Situated condition were instructed to walk around and gather information that could help them to answer questions (See Figure 2). The questionnaire consisted of three main sections. First section was a practice session. Next section gathered participants' demographic and vision data. The last section provided stimuli and questions: There were five blocks in this section, and each block had its own purpose (i.e., scenario types): 1) projectile trajectories [6, 25], 2) key changes in the environment [30], 3) movement direction [21, 24, 25], 4) movement/action duration [30], and 5) absolute measurements [3, 6]). Further, each block contained four questions focusing on its scenario type. For completion time, the time is captured, in both groups, from the moment participant start playing the video until the time he/she submit the answer. The selection of scenario types was made based on common video analysis activities found in [3, 6, 21, 24, 25, 30]. Participants were presented with one scenario block at a time, each containing video clips, questions related to an event in each video, and questions about participant confidence. To minimize the order effect, scenario order was counterbalanced for all the participants. Participants were allowed to analyze freely: replay, pause, rewind, and frame-step, as well as watch the video multiple times. Completion time, clicks at coordinates on images, answers to questions, and participants' confidence were collected. Participants used the imperial or metric system when reporting their measurements, based on their preference. Each study lasted roughly 75 minutes and each participant received a $15 gift card for their participation.

Figure 2: After a Situated group participant watched video clips and read questions related to the scenario, the participant walks to the location of events in the video to find (a,b) the ball's location when it touched the ground and reached its apex, (c) the marker's color used in the vandalism act, and (d) the ball's locations after it left actors' hands, (e) the time it took the actors to reach a predefined location, and (f) the newspaper stand height.

3.4 Video Clips

Five scenarios were acted out by three actors (two males and one female). Two conditions for each scenario were produced: video clips in on-camera condition captured all of the action within camera Filed of View (FOV), whereas, in the off-camera condition, some parts of the action occurred outside of the camera FOV. Videos were shot in a local university building atrium while others were not present. We positioned our camera at a height akin to that of the actual security cameras there. All videos were recorded in 1920 x 1080 resolution and 30 frames per second. We used an open source software called "Tracker" [6] to identify the accurate/correct responses (i.e., answer keys) from the video to be shown to participants. Twenty-six unique video clips were recorded. There were 4 to 8 clips in each scenario block. 44.125 seconds is the mean duration of all videos. 5 to 215 seconds was the range of duration for all videos. The experiment was comprised of original and unaltered footage (no added special effect or modification was present). The following scenario types were used in the study:

3.4.1 Projectile Trajectories: Participants viewed eight videos. Four variations of this scenario were created using two types of camera optical axes (perpendicular or parallel) and two types of FOV (on- or off-camera) (Figure 3). In the on-camera clip, an actor threw a soccer ball and the trajectory ends within the camera FOV (see Figure 3-a and c). However, the trajectory ends on the outside of the FOV in off-camera clips (see Figure 3-b and d). Questions asked regarded the ball's trajectory; 1) after it left the actor's hands, 2) at its apex, 3) at contact with the ground.

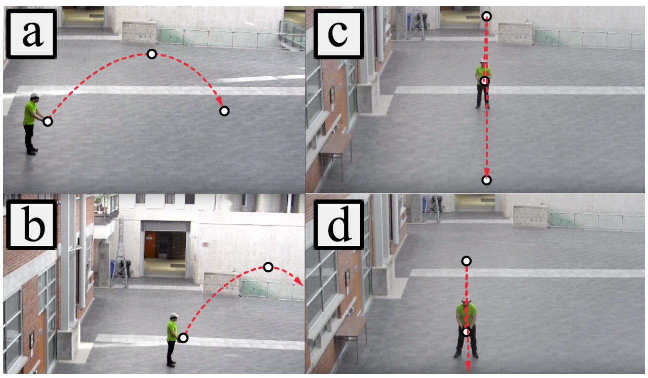

Figure 3: (a) and (b) show the ball being thrown perpendicular to the camera, with (a) in FOV and (b) not entirely in FOV. Image (c) and (d) show the ball being thrown parallel to the camera, with (c) in FOV and (d) not entirely in FOV. Note the red arrows and white dots were not shown to participants.

3.4.2 Key Environment Changes: Four videos explored questions related to physical changes that happen within a scene's environment (in this case, a vandalism event). Four white sheets of paper were posted in four different locations to simulate this event: two were placed on-camera while others were placed off-camera (see Figure 4). An actor walked into the camera FOV, sat down at a table next to the wall for a few seconds, walked toward one of the white papers, draws a shape with a color marker, then left the scene. Each sheet was marked with a different color (black, red, green, or blue). The participants' was asked to report the color.

3.4.3 Movement Directions: Four videos explored participants' abilities in tracking object movements and directions. In on-camera clips, three actors stood in a circle facing each other and threw a soccer ball between themselves (five passes in). Participants' tasks were to; 1) click on the location of the ball after it left an actor's hand, and 2) draw the ball movement and direction on a piece of paper. For off-camera clips, a red bag was placed on a table: the first actor picks up the bag, walks in a straight-line path and leaves the camera FOV for a few seconds. The actor then passes the bag to

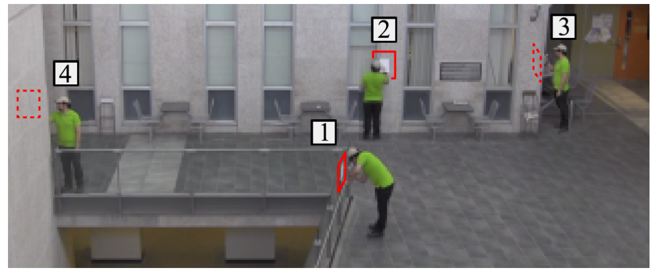

Figure 4: A stitched image of the actor acting in the four video clips. The figure shows different locations of posted white papers. The locations (1) and (2) in the camera FOV whereas (3) and (4) are outside the camera FOV. Please note that the red boxes were not shown to participants.

the second actor, who walks in straight-line into camera FOV for several seconds, then exits the camera FOV. Then the bag is passed to the third actor. The third actor then walks in a straight line into camera FOV for few seconds and stops. Participants were asked to identify the location and direction of the bag exchanges, using mouse clicks.

3.4.4 Duration of Movement/Action: Four videos explored the participants' measurement perception and estimation of event duration. For the on-camera material, three meeting points were predefined. In the video clips, three actors meet at each location for a certain amount of time, then disperse. Participants' tasks were to 1) measure the duration of each meeting. For the off-camera material, two actors sit at a table and talk for several seconds, then stand up to walk out of the camera FOV. Participant tasks were to 1) estimate the time the actors spent to arrive at a predefined location, outside of the camera FOV.

3.4.5 Absolute Measurements: Participants' perception regarding absolute measurements was explored with four videos. For the on-camera material, an actor places either a piece of duct tape on the floor or on a newspaper stand next to the wall (two scenarios), then leaves the camera FOV. Participants' tasks were to 1) report either the length of the duct tape and height of the newspaper stand. For the off-camera material, one actor begins on a lower floor, then walks halfway up a flight of stairs, stands for a few seconds, then walks back down to the lower floor. The upper half of the actor's body was shown to the camera FOV and was visible within the clip. Participants' tasks were to 1) report the height of the actor. The same action was repeated in another video clip, by an actor of a different height.

3.5 Results

After checking assumptions, mixed two by two ANOVAs were conducted throughout Study 1, to explore the effect of condition (Between: Situated vs. Non-Situated) and the analysis type (Within: On-camera vs. Off-camera) on each dependent variable. Participants' familiarity with the building did not differ across conditions, $F(1, 38) = .28$, $p = .60$. Participants' responses were compared against the answer keys we had generated. We report a significant value when $p <= .05$.

3.5.1 Projectile Trajectories.

Response error: The distance between the participants' responses (i.e., where the participants clicked on the monitor) and the correct response (in pixels where 1 pixel is equal to 0.0264 centimeter) were computed to indicate the magnitude of the participants' response error. A significant condition effect emerged, $F(1, 38) = 16.66$, $p = .0002$, $\eta p^2 = .31$. Scene type effect was also significant, Wilks' Lambda = .73, $F(1, 38) = 14.14$, $p = .0006$, $\eta p^2 = .27$. Further, a significant interaction effect emerged, Wilks' Lambda = .38, $F(1, 38) = 7.25$, $p = .011$, $\eta p^2 = .16$. Simple main effect analyses confirmed that in for both on- and off-camera scenes, participants in Situated condition made less error($ps < .05$, see Figure 5 for the means). Being in the scene allows participants to view the the video clips from different view-points, and this reduced their levels of errors. Hence, interestingly, participants in Non-Situated analysis made greater errors, compared to Situated condition, even when the video clip captured the entire scene within the camera FOV.

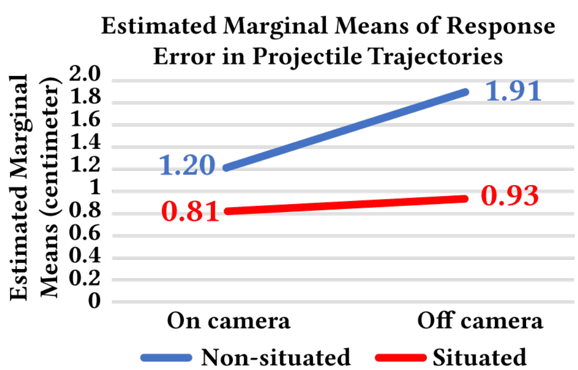

Figure 5: The interaction effect on the response error for projectile trajectories ($p = .01$).

Completion time: There was a significant condition effect, $F(1, 38) = 21.79$, $p = .00004$, $\eta p^2 = .36$. Compared to the Situated condition, participants spent less time when they were in the Non-Situated condition to complete their tasks. Further, the scene type effect was also significant, Wilks' Lambda = .30, $F(1, 38) = 88.73$, $p < .000001$, $\eta p^2 = .70$. There was a significant interaction effect, Wilks' Lambda = .56, $F(1, 38) = 29.41$, $p = .000004$, $\eta p^2 = .44$ (See Figure 6). Pairwise comparisons indicated that in both on-camera and off-camera materials, participants in the Situated condition spent longer time than their counterparts did.

3.5.2 Key Changes In the Environment.

Response accuracy: For the dependent variable, the percentage of the participants' correct responses in a color detection task was used. No interaction effect nor scene type effect emerged ($ps > .14$). A main effect of condition emerged, however; $F(1, 38) = 667.45$, $p < .000001$. As predicted, participants in the Situated condition responded perfectly ($M = 1.00$, $SD = .00$) on color detection task while participants in the Non-Situated condition did poorly ($M = .18$, $SD = .14$).

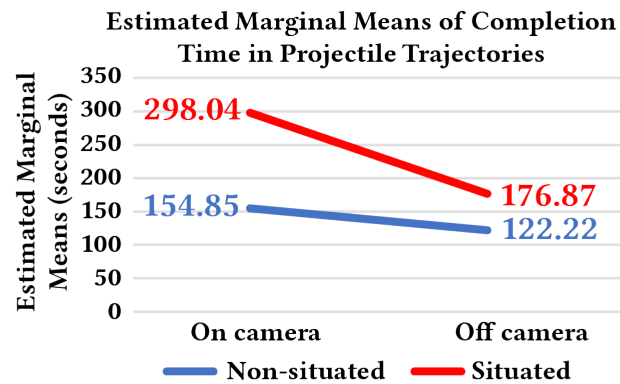

Figure 6: The interaction effect on the completion time in projectile trajectories ($p = .000004$).

Completion time: The effect of analysis type was found, Wilks' Lambda = .34, $F(1, 38) = 75.34$, $p < .001$, with a large effect, $\eta p^2 = .67$. Participants processed the on-camera materials faster ($M = 109.88$, $SD = 45.82$) than the off-camera materials ($M = 186.03$, $SD = 73.62$). There were no significant interaction nor condition effects ($ps > .05$).

3.5.3 Duration of Movement/Action.

Response accuracy: No significant effects were found ($ps > .15$).

Completion time: A significant condition effect was found, $F(1,38) = 21.57$, $p < .00004$, $\eta p^2 = .36$. Further, there was a significant analysis type effect, Wilks' Lambda = .52, $F(1, 38) = 34.81$, $p < .000001$, $\eta p^2 = .48$. Finally, there was a significant interaction effect (Wilks' Lambda = .76, $F(1, 38) = 12.11$, $p = .001$, $\eta p^2 = .24$). Simple main effect analysis confirmed that the only off-camra, participants in Situated condition took significantly longer than those who were in Non-Situated condition, but not with on-camera($p < .00001$, see Figure 7).

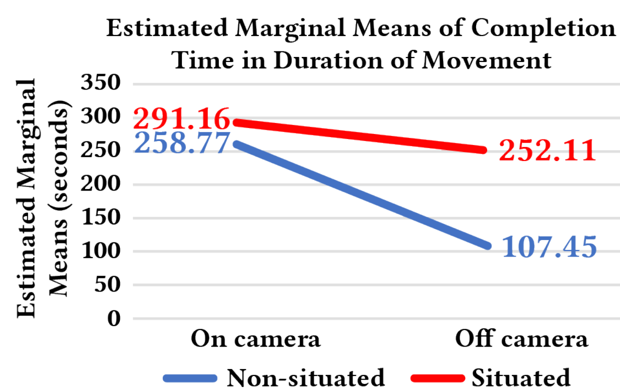

Figure 7: The interaction effect on the completion time in duration of movement/action ($p = .00001$).

3.5.4 Movement Direction.

Response accuracy: No significant effects were found ($ps > .19$).

Completion time: No significant condition effect nor interaction effects were found ($ps > .28$). An analysis type effect emerged, however (Wilks' Lambda = .66, $F(1, 38) = 19.98$, $p = .00007$, $\eta p^2 = .35$.

On average, processing on-camera material required longer time (M = 575.40, SD = 475.15) than processing off-camera materials (M = 366.23, SD = 231.20) regardless the condition.

3.5.5 Absolute Measurements.

Response accuracy: A main effect of condition emerged, $F(1, 38) = 28.61$, $p = .000004$, $\eta p^2 = .43$. The analysis type effect was not found ($p > .05$), however. An interaction effect emerged, Wilks' Lambda = .89, $F(1, 38) = 4.91$, $p = .03$, $\eta p^2 = .11$. Simple main effect analyses yielded that only with In-scene material, participants in Non-Situated condition made larger errors than the participants in the Situated condition did($p = .00001$).

Completion time: A condition effect emerged; $F(1, 38) = 13.95$, $p = .001$, $\eta p^2 = .27$. Participants in the Situated condition spent longer time (M = 181.83; SD = 65.23) than their counterparts did (M = 107.75, SD = 60.06). An effect of analysis type was also found: Wilks' Lambda = .77, $F(1, 38) = 11.64$, $p = .002$, $\eta p^2 = .23$. Participants spent longer time when they were analyzing on-camera materials (M = 118.88, SD = 51.32) than off-camera materials (M = 170.69, SD = 113.16). There was no interaction effect (p =. 06).

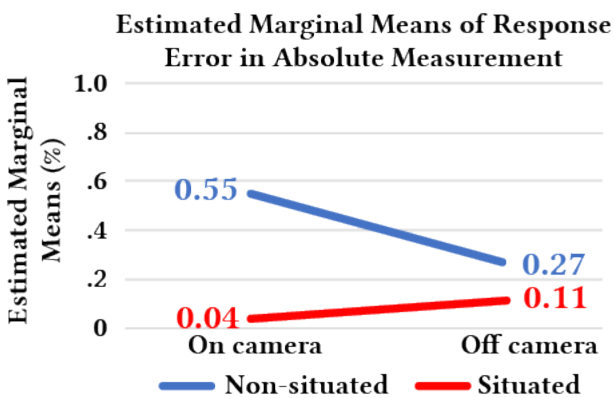

Figure 8: The interaction effect on the response error for absolute measurements: (p = .03).

3.5.6 Overall Confidence.

A question ("How confident are you about the answer you provided above?") assessed the participant's confidence level regarding their own analytical performance using a 7-point Likert scale. This question was provided immediately after each task. For the analysis, the mean of each task confidence was used. A main effect of condition and analysis type were found. When participants were in Situated condition, their confidence was significantly higher; $F(1, 38) = 8.74$, $p = .005$, M = 5.52, SD = .41, than their counterparts' (M = 4.90, SD = .84). Further, analyzing on-camera materials made the participants feel more confident about their analytic performance; $F(1, 38) = 21.38$, $p = .00004$, $\eta p^2 = .36$, M = 5.36, SD = .75, compared to the off-camera materials, M = 5.09, SD = .75. No interaction effect was found.

3.6 Discussion

Significant effects around the response error/accuracy confirmed potential benefits of SA in comparison to the traditional Non-Situated analytic method. Participants' errors were generally larger when they analyzed the videos in an office as opposed to the actual location, as expected. Furthermore, participants' Non-Situated analysis performances were often affected by the type of analysis; the magnitude of errors the participants made varied depending on the type of analysis (on/off-camera) when they performed estimations on projectile trajectories and absolute measurement tasks. When Non-Situated analyses (i.e., common visual analysis) were conducted, participants' errors were greater compared to the Situated analyses, even when the video clip contained the whole incident within the Camera FOV. While the differences we found might appear insignificant, if any decisions related to sentencing are made based on such data for example, the true consequences of these differences could be rather significant. This finding potentially implies the importance of situated analyses.

The accuracy came with a cost, however. Although the situated analytic method improved accuracy, users moving through the scene naturally increased the amount of time it takes to make estimate judgements. This trade-off is further justified when we observed participants' effort to be accurate. Participants in SA used additional techniques to find answers to their questions. For example, some participants made initial overall survey of the scene then performed the task in question. Some other participants used the measurement tools twice to ensure that they provided correct answers. One can argue that in the non-situated settings, a better video quality or multiple camera angles could provide benefits similar to situated video-analysis. However, going over multiple video clips could be inconvenient. The generalizability of this study is limited due to participants characteristics in term of their video analysis experience. Almost none (97.5%) of our participants had experience in video analysis: A future study examining experience effect will be fruitful. Finally, situated video analysis improved the participants' confidence on their judgement; such increased confidence could be important when they are using visual analyses to make decisions, for example. In summary, the results of Study 1 generally indicate sufficient potentials for us to explore SA further, with the caveat that it can take participants longer. In the next step, thus, we head towards an exploration of SA platform, using Mixed Reality (MR).

4 STUDY 2: SITUATED VIDEO DATA VISUALIZATION

In HCI research, elicitation studies and design workshops have been commonly used for various purposes; for example, to explore user gesture interaction designs [41], to gain insight on current needs for exploring data and visualization designs [1], and to present new situated visualizations of data, which is displayed in proximity to the physical referents in the environment [5, 12]. However, as far as we are aware, there is no research exploring the means to design visual analytical tools for situated video analysis. Encouraged by the results of Study 1 which confirmed the potential of SA, we conducted an elicitation study which included sketching and ideation activities for SA. The goal of this study was to capture visualization designs to support situated video data Visualization, potentially to enhance the analytical process in MR environment.

4.1 Participants

The study was advertised on campus bulletin boards at a local university. 12 participants (M = 8, F = 4) were recruited, none of whom partook in the first study. Their age ranged between 20 and 31 (M = 24.50, SD = 4.37). 25% of them reported that English was their first language and 16% preferred not to report their first language. No language issues were exhibited during the study. All participants reported normal or corrected-to-normal vision. Each session was conducted with two participants (i.e., pair). Two people were required so the participants could bounce back and forth their ideas with their partner. For the time constraint, we did not include more than 2 people in a session, however. Participants took turns throughout the session. They received a $20 gift card as compensation for their time.

4.2 Apparatus

A Microsoft Surface Pro 2 and Microsoft HoloLens were used for the following reasons, 1) to familiarize participants with the experiences of video viewing while being mobile and in-situ, 2) to have platforms which incorporate participants' sketches using MR while making them aware that their drawings will *not* be in VR in the study, 3) to help participants understand the mapping of the events they see in the video clip in the actual location. Participants were informed not to consider VR platform to ensure they understand in-situ visualization in which their drawing will be attached into physical environment of the event of interest [16, 17]. Although Microsoft Surface Pro 2 and Microsoft HoloLens differ in terms of input modality and interaction, participants' interaction with HoloLens and Microsoft Surface Pro 2 was limited to only video playing. We consider participants' interaction in both device as a minimum exposure to different interaction techniques to see whether participants will incorporate these interaction into their sketching ideas per scenario as well as contribute toward novelty.

Once again, for the ability to switch from laptop mode to tablet mode and vice-versa, a detachable keyboard was provided to the paired participants to support their mobility. A video camera, the Canon HF-M52, was used to record video scenarios in a similar manner to that of Study 1. All five scenarios were performed in FOV and on one clip (1:24 mins). The video clips' size and format were the same for both the Microsoft Surface Pro 2 and Microsoft HoloLens to counterbalance the large screen size difference between the two devices. A4 sheets of paper (21 cm X 29.7 cm) were provided to participants with a 2D and 3D representation of the space the video was captured. Colored pens for use in sketching activities were also provided to participants, with this technique showing positive results in situated visualizations [5].

4.3 Method

Six workshop sessions were coordinated with two participants each. Each session took between 1.5 and 2 hours (including 10 minutes of interview). At the beginning of each session, a research assistant was introduced as a moderator/note taker. After signing a consent form, they filled out a short demographic questionnaire. The Microsoft Surface 2 and Microsoft HoloLens were introduced, and instruction for watching video on both devices was provided. The two participants, in groups of two, were then taken to a university atrium where the video was captured, thus situated. Participants sat at a table next to each other, and were informed that their help was needed to develop a situated visualization in 3D space of event data from a video scenario. Participants were provided with one scenario at a time, each containing a set of instructions. The order of scenarios was counterbalanced for all groups. First, participants were asked to watch the video using both Microsoft Surface 2 and Microsoft HoloLens, and walk toward the location where the event(s) took place. Second, participants were asked to discuss how they would visually represent an event in the video. Third, participants were asked to sit at a table and sketch their visualization ideas on the provided paper, to represent the events from the video. Participants were asked to produce two 2D sketches to describe their ideas per scenario; they processed one scenario at a time. A post-study interview with participants was conducted to explore participants' experiences further.

4.4 Results

We systematically coded the transcripts, sorted the photos based on ideas, analyzed the use of different form factors in participants' sketches, and created a summary of all findings with relevant quotes from the transcripts. Sixty sketches were generated in total (i.e., 6 pairs x 2 per scenario x 5 scenarios). These sketches were redrawn digitally, copying the original drawings as closely as possible. Each digital sketch was summarized and analyzed in detail. For the analysis, a research assistant watched the video clip participants watched, looked at the participants' drawings, then drafted a short explanation of what each drawing expresses. These processes yielded three major components from the participants' ideas: 1) Information Density Levels, 2) Interactivity, and 3) Event-Narrative. Examples of the emergent sketch themes are provided in Table 1.

4.4.1 Information Density Levels: Participants in our study had mid-level video analysis experience (M = 3.33, SD = 1.92) while only 3 participants scored above the mid-point(i.e., 3.5). Also, participants reported low to no experience with HWD (M = 2.25, SD = 1.60) where 5 participants have no experience and only 3 participants scored over the midpoint. The analysis of participants' drawings revealed two levels of analysis: low- and high-density levels. While low-density drawings did not include details of any event data, they provided an abstract view of the event and a quick sense of the event. Participants were generally inclined to produce sketches that revealed a minimum amount of event data. An example of a low-density drawing in which a character is shown throwing a ball is demonstrated in Figure 9-b. Drawings falling into the category of high-density detail of visualization; more detailed information about the events such as time, location, duration, etc, can be seen in Figure 9-a, 9-c, and 9-f. High-density drawings captured important and relevant event data (see Figure 9). 60.00% of drawings fell into low-density category. The level of analysis (i.e., Low vs. High) varied depending on the video event scenarios (see Section 3.4). For example, the absolute measurements scenario is a simple scenario where data, (e.g., the height of person) can be communicated with a simple visual representation(see Table 1); more overview drawings (10) than detailed drawings (2) were produced by participants. On the other hand, users were prompted to create more high-density

Table 1: The distribution of theme categories of participants' sketches.

Scenario	Information Density Levels		Interactivity		Event-Narrative	
	Low-density	High-density	Interactive	Non-interactive	Narrative	Non-narrative
Projectile Trajectories	4	8	2	10	3	9
Key Changes in the Environment	7	5	3	9	1	11
Movement Direction	8	4	1	11	4	8
Duration of Movement/Action	7	5	2	10	3	9
Absolute Measurements	10	2	1	11	2	10
Total (%)	**36 (60.00%)**	**24 (40.00%)**	**9 (15.00%)**	**51 (85.00%)**	**13 (21.67%)**	**47 (78.33%)**

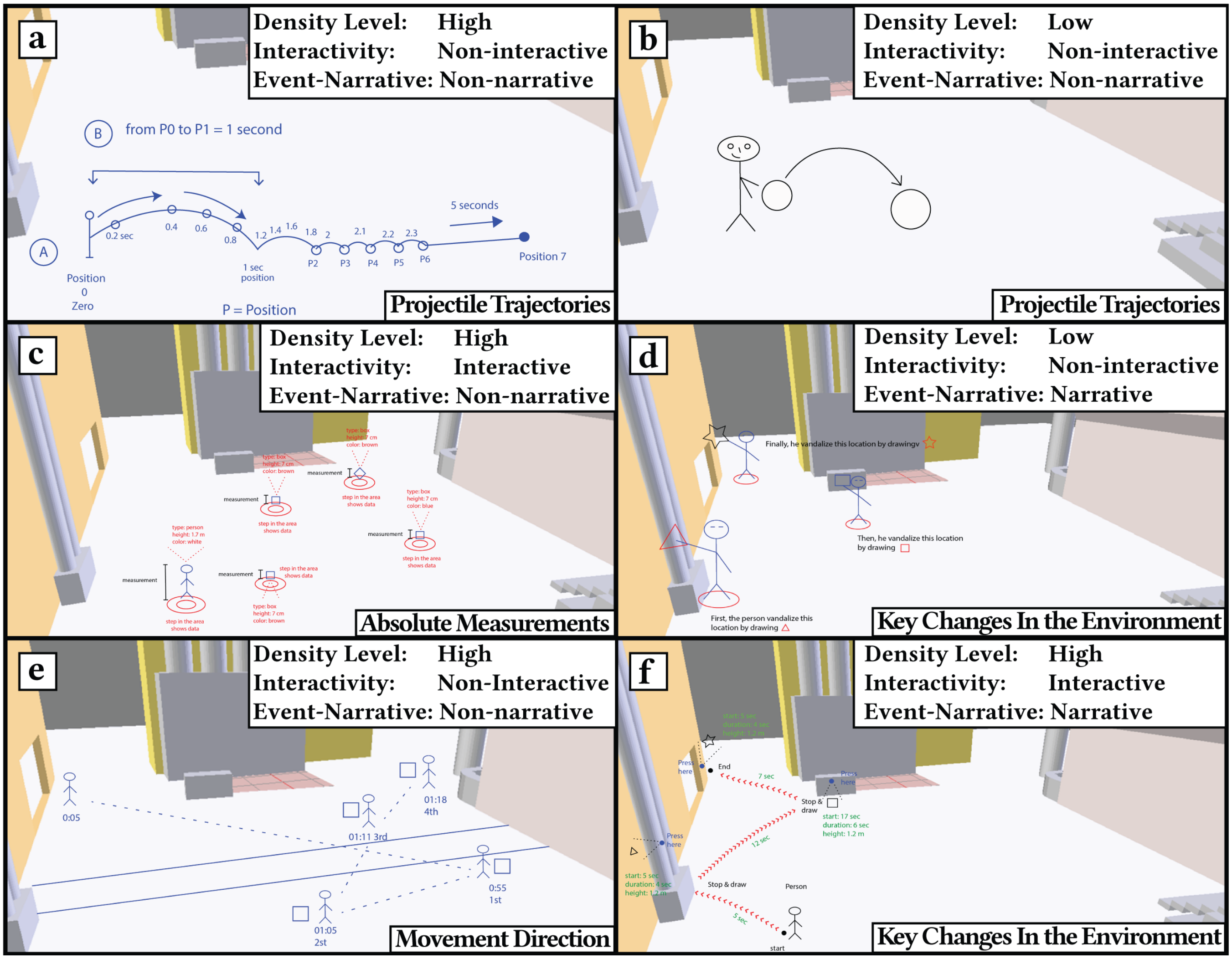

Figure 9: Sample of participants' sketches.

detail drawing than low-density detail drawings for more complex scenarios such as the projectile trajectories scenario.

4.4.2 Interactivity: Only 15% of the sketches contained an interactive component. Various interactive functions were introduced by participants. For example, functions such as clicking on a vandalized wall to reveal more information; the time of occurrence, duration of the act, and height of the actor (see Figure 9-f). The use of physical movement was also proposed in the sketches. In Figure 9-c, a participant indicates the location of an object or an individual standing.

4.4.3 Event-Narrative: Annotations were included by some participants to directly instruct a user to explore (e.g., asking users to click, stand at the location of interest) for further information (Figure 9-c). Another example of annotation usage leveraged chronological order of events in the video to inform users (Figure 9-d). 21.67% of the sketches included event narratives.

4.4.4 Interview: Participants were interviewed after the design session, to capture their thoughts about their experiences and drawings. Analysis of the participants' interviews revealed following four themes.

2D and 3D Data Visualization: All the participants preferred 3D over 2D representations: Rationales for this preference were supporting in-situ video analysis approach, physical mobility between events, and different viewpoints. For example, P1 mentioned that 3D visualization "…helps you to move around different event in the video and can look at them as you are part of the event" [sic]. P2 reported "…video events happened in 3D space and to make sense of the event data it should be visualized in the same 3D space" [sic]. Further, participant P10 stated that "…you can see the object from all viewpoint" [sic]. Other participants suggested that 3D visualization supports a multivariate representation of events where "you can add depth into it you can add more information" [sic] (P3). Thus, all the participants recognized the strength of 3D visualization.

Benefits of Situated Video Visualization: Participants reported several anticipated benefits of the video visualization techniques; 1) *reduction of video browsing time* (P1, P2, P3), 2) providing a better understanding of events (P3, P4, P5, P8, P11), and 3) supporting interactivity which increases engagement with events (P4, P5, P9, P12). Participants felt that visualizing the information could reduce the time and effort of event exploration. For example, P1 mentioned that "in the video you have to watch 1 to 2 mins where in the picture drawing (drawing of the video events) you can see the data and people can look at it in like 10 seconds …video watching is sequential you have to watch all the video." In addition, participants felt that their drawings could *enhance user understanding* of video events. As reported by P3 and P4 "I think for people less trained, our drawing will help them understand the 3D aspect by adding depth into the scenario" [sic] (P3), and "data that we draw has all the necessary information that someone needs to examine events in the video" [sic] (P5).

Situated Video Analysis Platform: Participants reported their analysis platform preference (i.e., Microsoft Surface 2 vs. Microsoft HoloLens) if they were to conduct future analysis of 3D drawing of a video. 41.67% of them preferred tablet use, whereas 58.33% preferred the Head Worn Display (HWD); $X^2(1, N = 12) = .33, p = .56$. Despite a stated appreciation for HWD, device weight, size, complex interactions, and social acceptability were the main reasons and limitations participants cited for preferring the tablet. Interestingly, however, none of these comments refer to the efficacy of HWD. Freehand interactions, mobility, and a better sense of immersion between the virtual and physical environment were among the rationale provided by participants who felt positive about the HWD. For example, P1 felt that using HWD will help her to focus on the task, stating "…(you) are not distracted and you can focus on the objects" [sic]. Participants P9, P11, and P12 mentioned that tablets do not support full immersion with a digital world. For example, P11 stated "…I was seeing the same video using the tablet, it was very hard for me to generate the sense of the location, time, and direction." [sic]. Participants P5, P7, and P8 expressed the ease of mobility in a 3D scene using HWD. For example, P8 stated "…It gives the ability to move easily and have your hand free." (P8). Thus, they recognized the potential of HWD.

Situated Video Visualization Challenges: One participant felt that the visualization of the extracted video data could capture important and relevant data, such that it could replace traditional videos, stating that "The drawing we come up with will make it easy for video analyzers to understand and make sense of what happened even if they did not see the video" [sic] (P6). On the other hand, several participants expressed their concerns about possible errors made in the process of transferring and encoding extracted data from a video to visualization. For example, P4 mentioned that "…the hindrances of transforming video events is that if designers made mistakes or wrongly transform the data."

4.5 Discussion

Participants' responses revealed interesting visualization themes, insight, and challenges. The information density levels in visualization, which was found in our participants' sketches, is a common finding in information visualization [36]. The high-density visualization is considered as a first step in visual investigation and exploration techniques, "Overview first, zoom, filter, and then focus details-on demand" [36]. A situated visual investigation and exploration require perception and action [42]. Participants expressed interaction in their drawings to support both embodied cognition [42] and embodied interaction [11]. For example, hand gesture (i.e., clicking) and physical body movement (i.e., standing at a certain location) are in accord with the exploration activity of video analysis. Further investigation is required to explore different interaction techniques that are suitable for immersive analytic (i.e., situated video visualization) [13]. Furthermore, narrative visualization incorporates information, communication, and exploration visualization to convey a story [23]. Some of the participants maintain a narrative of the events when they sketched video clips of the scenarios. Textual annotation is a design tactic used to leverage the information presented, to direct user attention, stress the chronological order of events, or show transitions in an event [23]. When extracting data from the video, it is important to use tools that ensure the validity and accuracy of the data. Based on the observation of the studies, a shortcoming of the situated video analysis technique would be the physical effort required by users when there are in the place of the event. However, situated video analysis techniques could be beneficial for different application domain. For example, during a sport training session (e.g., a soccer player visualizing kicked ball trajectories using situated video analysis, they will have better understanding during training regarding how to replicate such a kick).

5 DESIGN TAKE-AWAYS

Two studies investigated situated video analysis and visualization sketches in a rather holistic manner. The results of the first study repeatedly indicated the potential benefit of situated video analysis. Furthermore, results from the second study revealed meaningful themes and design considerations for future prototypes. The following take-aways are presented for future consideration by designers of situated Video Analytic interfaces:

- For situated video analysis, future analytical program should incorporate a low- and a high-level detail visualization of events, to provide the capability to interact with the event data, and narrative of the event.
- Visualizations for situated video analysis should include the original video footage as a tool for validation or reference.
- The use of annotations in situated video analysis visualization supports event narrative via grabbing users' attention, clarifying chronological order of events, and indicating event transitions. Also, textual annotation could be used as a way to capture and exchange user insight and conclusions.
- Situated video analysis visualization could incorporate multiple levels of contextual information, to support multivariate types of analyses (e.g., including fine details relating to, but not limited to, the variables of time, duration, object classification, objection location, object direction, object velocity, event summary, etc).
- Situated visualization should incorporate embodied cognition and embodied interaction.

6 FUTURE DIRECTION

Our results suggest that situated video analysis can improve users' performance in common video analytic tasks. Although traditional video analyses are normally conducted in non-situated settings, rapid advances in mixed reality systems has created valuable new opportunities. In this vein, we explore the design of novel user interfaces to support situated video analysis. Figure 1 shows a situated video visualization of projectile trajectory made by one group.A necessary step is to follow up with experts to confirm what themes and feature that would be useful in practice. We are interested in the implementation of situated video visualization sketches; however, we may look to also explore their effect on non-situated analysis as well. Situated video visualization could be useful in non-situated setting as well especially when surveillance areas are not accessible for further analysis. Situated data visualization interaction, to some extent, adopts traditional non-situated controls and interaction metaphors (e.g., using buttons or sliders, to interact with data). Future considerations for the extension/continuation of this research project will include the study of interaction methods which use physical body movements to perform data exploration and interaction functions, and ultimately, the development of the analytic tool for situated analyses.

7 CONCLUSION

This paper offers an early look at video analysis activities and performance in both a situated and a non-situated setting. From our first study, we observed that compared to the performance of participants in a traditional non-situated setting, users' performance improved in the situated settings with heightened accuracy and confidence in their judgement tasks. The situated tasks naturally took participants longer as it involved physical inspection of the environment. This paper further attempts to provide practical requirements and take-aways for situated video analysis designs, including information density levels, interaction, and event-narrative which should be considered. In future work, we plan to explore the design of interfaces for video analysis.

REFERENCES

[1] Fereshteh Amini, Khalad Hasan, Andrea Bunt, and Pourang Irani. 2017. Data Representations for In-situ Exploration of Health and Fitness Data. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Barcelona, Spain) *(PervasiveHealth '17)*. ACM, New York, NY, USA, 163–172.

[2] Ryuichi Ayase, Terumasa Higashi, Satoru Takayama, Setsuko Sagawa, and Nobuyuki Ashida. 2008. A method for supporting at-home fitness exercise guidance and at-home nursing care for the elders, video-based simple measurement system. In *HealthCom 2008 - 10th International Conference on e-health Networking, Applications and Services*. 182–186.

[3] Chiraz BenAbdelkader and Yaser Yacoob. 2008. Statistical body height estimation from a single image. In *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*. 1–7.

[4] R. Borgo, M. Chen, B. Daubney, E. Grundy, G. Heidemann, B. Höferlin, M. Höferlin, H. Leitte, D. Weiskopf, and X. Xie. 2012. State of the Art Report on Video-Based Graphics and Video Visualization. *Computer Graphics Forum* (2012).

[5] Nathalie Bressa, Kendra Wannamaker, Henrik Korsgaard, Wesley Willett, and Jo Vermeulen. 2019. Sketching and Ideation Activities for Situated Visualization Design. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. ACM, New York, NY, USA, 173–185.

[6] Doug Brown and Wolfgang Christian. 2011. Simulating What You See. In *MPTL 16 and HSCI 2011*. Ljubljana, Slovenia.

[7] Madhav Chamle, K. G. Gunale, and K. K. Warhade. 2016. Automated unusual event detection in video surveillance. In *2016 International Conference on Inventive Computation Technologies (ICICT)*, Vol. 2. 1–4.

[8] M. Chithra and R. Arunkumar. 2015. Person identification from video by similarity between feature set. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. 1–6.

[9] Antonio Criminisi, Andrew Zisserman, Luc J Van Gool, Simon K Bramble, and David Compton. 1999. New approach to obtain height measurements from video. In *Investigation and Forensic Science Technologies*, Vol. 3576. International Society for Optics and Photonics, 227–239.

[10] Carr Daniel B., Denis White, and MacEachren Alan M. 2005. Conditioned Choropleth Maps and Hypothesis Generation. *Annals of the Association of American Geographers* 95, 1 (2005), 32–53. arXiv:https://doi.org/10.1111/j.1467-8306.2005.00449.x

[11] P. Dourish. 2004. *Where the Action is: The Foundations of Embodied Interaction*. The MIT Press, Cambridge, Massachusetts.

[12] Alix Ducros, Clemens N. Klokmose, and Aurélien Tabard. 2019. Situated Sketching and Enactment for Pervasive Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS '19)*. Association for Computing Machinery, New York, NY, USA, 217–228.

[13] Tim Dwyer, Kim Marriott, Tobias Isenberg, Karsten Klein, Nathalie Riche, Falk Schreiber, Wolfgang Stuerzlinger, and Bruce H. Thomas. 2018. *Immersive Analytics: An Introduction*. Springer International Publishing, Cham, 1–23.

[14] Neven A. M. Elsayed, Ross T. Smith, Kim Marriott, and Bruce H. Thomas. 2016. Blended UI Controls for Situated Analytics. In *2016 Big Data Visual Analytics (BDVA)*. 1–8.

[15] Thomas Bruce H Elsayed, Naven A. M., Ross T. Smith, Kim Marriott, and Julia Piantadosi. 2015. Using augmented reality to support situated analytics. In *2015 IEEE Virtual Reality (VR)*. 175–176.

[16] Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction* (Tokyo, Japan) *(SUI '16)*. ACM, New York, NY, USA, 99–102.

[17] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. 2014. Ethereal planes: a design framework for 2D information space in 3D mixed reality environments. In *SUI '14*.

[18] Dan B. Goldman, Brian Curless, David Salesin, and Steven M. Seitz. 2006. Schematic Storyboarding for Video Visualization and Editing. In *ACM SIGGRAPH 2006 Papers* (Boston, Massachusetts) *(SIGGRAPH '06)*. ACM, New York, NY, USA, 862–871.

[19] Khalad Hasan, Tovi Grossman, and Pourang Irani. 2011. Comet and Target Ghost: Techniques for Selecting Moving Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 839–848.

[20] Markus Höferlin, Benjamin Höferlin, Gunther Heidemann, and Daniel Weiskopf. 2013. Interactive Schematic Summaries for Faceted Exploration of Surveillance

Video. *IEEE Transactions on Multimedia* 15, 4 (June 2013), 908–920.
[21] Markus Höferlin, Benjamin Höferlin, and Daniel Weiskopf. 2009. Video Visual Analytics of Tracked Moving Objects. *Proceedings of 3rd Workshop on Behaviour Monitoring and Interpretation* 541, 59–64.
[22] Markus Höferlin, Benjamin Höferlin, Daniel Weiskopf, and Gunther Heidemann. 2011. Interactive Schematic Summaries for Exploration of Surveillance Video. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval* (Trento, Italy) *(ICMR '11)*. ACM, New York, NY, USA, Article 9, 8 pages.
[23] Jessica Hullman and Nick Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec 2011), 2231–2240.
[24] Zhong Ji, Yuting Su, Rongrong Qian, and Jintao Ma. 2010. Surveillance video summarization based on moving object detection and trajectory extraction. In *2010 2nd International Conference on Signal Processing Systems*, Vol. 2. V2–250–V2–253.
[25] Don Kimber, Tony Dunnigan, Andreas Girgensohn, Frank Shipman, Thea Turner, and Tao T. Yang. 2007. Trailblazing: Video Playback Control by Direct Object Manipulation. In *2007 IEEE International Conference on Multimedia and Expo*. 1015–1018.
[26] Kong Chien Lai, Yeow Peng Chang, Kin Hoe Cheong, and Siak Wang Khor. 2010. Detection and classification of object movement - an application for video surveillance system. In *2010 2nd International Conference on Computer Engineering and Technology*, Vol. 3. V3–17–V3–21.
[27] Y. Li, Y. Zhang, J. Lu, R. Lim, and J. Wang. 2010. Video Analysis and Trajectory Based Video Annotation System. In *2010 Asia-Pacific Conference on Wearable Computing Systems*. 307–310.
[28] Yong-Jin Liu, Cuixia Ma, Guozhen Zhao, Xiaolan Fu, Hongan Wang, Guozhong Dai, and Lexing Xie. 2016. An Interactive SpiralTape Video Summarization. *IEEE Transactions on Multimedia* 18, 7 (July 2016), 1269–1282.
[29] Shao-Ping Lu, Song-Hai Zhang, Jin Wei, Shi-Min S. Hu, and Ralph R. Martin. 2013. Timeline Editing of Objects in Video. *IEEE Transactions on Visualization and Computer Graphics* 19, 7, 1218–1227.
[30] Ahmed H. Meghdadi and P. Pourang Irani. 2013. Interactive Exploration of Surveillance Video through Action Shot Summarization and Trajectory Visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec 2013), 2119–2128.
[31] Brendan Tran Morris and Mohan Manubhai Trivedi. 2008. A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* 18, 8 (Aug 2008), 1114–1127.
[32] Chong-Wah Ngo and Feng Wang. 2009. *Video Summarization*. Springer US, Boston, MA, 3320–3324. https://doi.org/10.1007/978-0-387-39940-9_1026
[33] Cuong Nguyen, Yuzhen Niu, and Feng Liu. 2012. Video Summagator: An Interface for Video Summarization and Navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 647–650.
[34] Y. Pritch, A. Rav-Acha, and S. Peleg. 2008. Nonchronological Video Synopsis and Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 11 (Nov 2008), 1971–1984.
[35] Zhen Qin and Christian R. Shelton. 2017. Event Detection in Continuous Video: An Inference in Point Process Approach. *IEEE Transactions on Image Processing* 26, 12 (Dec 2017), 5680–5691.
[36] Ben Shneiderman and Catherine Plaisant. 2004. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.
[37] Manuel Stein, Thorsten Breitkreutz, Johannes Haussler, Daniel Seebacher, Tobias Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel Keim, and Halldor Janetzko. 2018. Revealing the Invisible: Visual Analytics and Explanatory Storytelling for Advanced Team Sport Analysis. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*. 1–9.
[38] Anthony Tang, Saul Greenberg, and Sidney Fels. 2008. Exploring Video Streams Using Slit-tear Visualizations. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Napoli, Italy) *(AVI '08)*. ACM, New York, NY, USA, 191–198.
[39] Bruce H. Thomas, Gregory F. Welch, Pierre Dragicevic, Niklas Elmqvist, Pourang Irani, Yvonne Jansen, Dieter Schmalstieg, Aurélien Tabard, ElSayed Neven A. M., Ross T. Smith, and Willett Wesley. 2018. *Situated Analytics*. Springer International Publishing, Cham, 1–23.
[40] J. J. Thomas and K. A. Cook. 2006. A visual analytics agenda. *IEEE Computer Graphics and Applications* 26, 1 (Jan 2006), 10–13.
[41] Jagoda Walny, Bongshin Lee, Paul Johns, Nathalie Henry Riche, and Sheelagh Carpendale. 2012. Understanding Pen and Touch Interaction for Data Exploration on Interactive Whiteboards. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec 2012), 2779–2788.
[42] M. Wilson. 2002. Six views of embodied cognition. Psychonomic. *Psychonomic Bulletin & Review* 9 (2002), 625–636.