VisTellAR: Embedding Data Visualization to Short-form Videos Using Mobile Augmented Reality

Wai Tong, Kento Shigyo, Lin-Ping Yuan, Mingming Fan, Ting-Chuen Pong, Huamin Qu, *Member, IEEE,* and Meng Xia

Abstract—With the rise of short-form video platforms and the increasing availability of data, we see the potential for people to share short-form videos embedded with data in situ (e.g., daily steps when running) to increase the credibility and expressiveness of their stories. However, creating and sharing such videos in situ is challenging since it involves multiple steps and skills (e.g., data visualization creation and video editing), especially for amateurs. By conducting a formative study (N=10) using three design probes, we collected the motivations and design requirements. We then built VisTellAR, a mobile AR authoring tool, to help amateur video creators embed data visualizations in short-form videos in situ. A two-day user study shows that participants (N=12) successfully created various videos with data visualizations in situ and they confirmed the ease of use and learning. AR pre-stage authoring was useful to assist people in setting up data visualizations in reality with more designs in camera movements and interaction with gestures and physical objects to storytelling.

Index Terms—Personal data, augmented reality, data visualization, storytelling, short-form video

✦

1 INTRODUCTION

SHORT-FORM video platforms (*e.g.*, TikTok and Snapchat) have been increasingly popular for people to share their moments [1], [2]. At the same time, with the rise of self-tracking apps [3], [4], people are more accessible to personal data [5], [6], [7]. We have surveyed the top 10 health and fitness apps in both Google Play Store[1] and Apple Store[2] and found that 11 out of 17 apps include data visualization and numbers. We can imagine that people include these data in their short videos, especially with data visualization, to express their stories and engage the audience. For example, people may share a video of themselves running while mentioning the distance they have run so far in the week. However, the current way to show data is only verbal explanation or textural script, which is not scalable. Thus, people embed data visualization (*e.g.*, figures and charts) into videos, which are often referred to as data videos [8]. These data visualizations in videos have been shown to increase the story's credibility and facilitate the communication of

insights in data [9], [10], [11]. Moreover, embedding data near objects [12], in reality, could help convey emotion [13] and enhance different abilities of viewers, such as attention, numeracy, and comparison [14].

Nonetheless, data videos are often created by data designers or domain experts for news media, visualization blogs, and online video platforms in prior works [8], [10], [15], [16], [17], [18]. These videos focused more on presenting complex visualizations and producing professional data insights. In contrast, personal visualization has different requirements for the data videos, such as using simple and standard charts [5], [6]. Moreover, the attributes of short-form videos bring unique requirements and challenges in embedding data visualization, especially for amateur video creators with little expertise in creating data visualization. First, it is challenging to align data visualization with the real-world content in video frames [19]. Second, the creation process should be simple enough for video creators to create, edit, and share the video at the moment.

Previous workflow (*i.e.*, the previous approach in Figure 1) to embed data visualization is via post-editing techniques after the video is shot. Users need to prepare a video first, the visualization then, and finally embed the visualization into the video. Moreover, most existing commercial video post-editing tools (*e.g.*, Adobe After Effects) and research prototypes for embedding data visualizations in videos [20], [21] are desktop-based. This workflow and the existing tools limit creators' ability to create, edit, and share data videos in situ immediately. First, this workflow does not allow video creators to adjust the use of data visualization (*e.g.*, type, position, size) on-site before video taking to integrate the visualization into the scene seamlessly by avoiding noisy background and insufficient space. Second, these tools hinder users from sharing their stories

- *W. Tong are with the School of Performance, Visualization and Fine Art, Texas A&M University, College Station, Texas, United States.*
  E-mail: wtong@tamu.edu
- *K. Shigyo, L. Yuan, T. Pong, and H. Qu are with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.*
  E-mail: {kshigyo, lyuanaa, tcpong, huamin}@ust.hk
- *M. Fan is with Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China.*
  E-mail: mingmingfan@ust.hk
- *M. Xia is with the Department of Computer Science and Engineering, Texas A&M University, College Station, Texas, United States.*
  E-mail: mengxia@tamu.edu

Manuscript received April 19, 2005; revised August 26, 2015.

1. https://play.google.com/store/apps/category/HEALTH_AND_FITNESS?hl=en&gl=US&pli=1

2. https://apps.apple.com/us/charts/iphone/health-fitness-apps/6013

on time since they need to switch between video-taking and post-editing tools even at different locations.

With the advancements in mobile augmented reality (AR), which allows detecting and overlaying virtual objects in-situ [22], we see an opportunity to leverage such technology to assist video creators with creating data visualizations and embedding them onto the short-form video in-situ [23], [24]. However, these works targeted one scene rather than the video creation process. They were limited in providing specific components (*e.g.*, visualization sequence control) [25] and an integral and effective pipeline for adding visualizations in short-form videos. Thus, we want to explore how AR helps embed data visualization in creating short-form videos.

To understand the motivations, challenges, and design requirements of constructing short-form videos with data visualizations using mobile AR, we conducted a formative study and a design probe activity with 10 amateur video creators. Based on the results and the literature review, we derived four main design requirements: integrate scene understanding to blend data visualization with the physical reality, enable in-situ visualization configuration, allow easy dataset importing and editing, and support visualization fine-tuning and narrative sequence control in the post-editing stage, and designed VisTellAR, a mobile AR authoring tool for amateur video creators to embed data visualizations in short-form videos to express themselves and share the moments in a data-supported way. Users can link VisTellAR with mobile health data (*e.g.*, Google Fit) or other online datasets (*e.g.*, Weather API), create and configure data visualizations in-situ before video-taking based on tracking and object detection, and post-edit the visualizations after video-taking, as illustrated in Figure 1.

To study how people use VisTellAR to add data visualization to short-form videos and evaluate their user experience, we conducted a two-day take-home user study with 12 video creators. We collected and analyzed participants' log data and videos produced, as well as the rating and feedback, to gain a deeper understanding of their video creation experiences with VisTellAR. The result showed that all participants could use VisTellAR to create an average of 14.8 seconds of short-form videos with data visualization in an average of 5.58 minutes. We demonstrated the effectiveness of our tools by providing examples created by the participants. We then analyzed the usage patterns to understand how participants used AR authoring to add data visualization to videos. In addition, questionnaires and interview results showed that VisTellAR is useful and efficient. Based on the result, we found that pre-stage AR authoring was useful to assist amateur video creators in setting up data visualizations in reality. Moreover, mobile AR authoring brought more designs with camera movements, gestures, and physical objects to storytelling with data visualizations.

2 RELATED WORK

2.1 Embedding Data Visualization to Videos

Embedding data visualizations to videos for storytelling, also known as data videos, is getting more popular due to the expressiveness brought by their combination of diversified visual and auditory stimuli [8], [15]. In data videos,

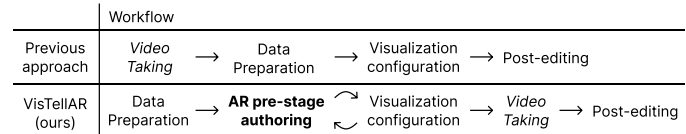|	Workflow			
Previous approach	*Video Taking* →	Data Preparation →	Visualization configuration →	Post-editing
VisTellAR (ours)	Data Preparation →	**AR pre-stage authoring** →	Visualization configuration →	*Video Taking* → Post-editing

Fig. 1. This figure illustrates the pipeline of VisTellAR and the comparison to the existing approach. VisTellAR provides an AR pre-stage authoring and allows visualization configuration before video-taking, while the existing approach [20], [21] relies on post-editing techniques, which require an existing video as input.

the role of visualization could be different. For example, some are used as figures to help convey a part of the story, and others can drive telling a data story [8]. Previous research analyzed videos collected with specific keywords related to data videos (*e.g.*, "data video" and "animated infographic") from news media, visualization blogs, online video platforms (*e.g.*, YouTube.com and Vimeo.com), and the like [8], [10], [15], [16], [17], [18]. Most of these cases use data visualizations to drive data storytelling, presented as data-driven motion graphics.

Recently, short-form video has become a popular medium for archiving the moment and expressing oneself [2] due to the rapid growth of short-form video platforms, for instance, TikTok, Snapchat, and Instagram. Telling stories with data moves from professional interests to personal contexts [5], [26] and from experts to ordinary people [7]. With the increased integration of sensors into mobile devices (*e.g.*, smartphones and smartwatches), self-tracking apps are more accessible in the market, leading to the movement of the "quantified self". Data visualization could allow people to express themselves with data insights [6]. For example, one can visualize their tracked happiness, highlight the current emotion, and share it with friends. Moreover, by embedding data visualization to real objects, data visualization could convey emotion [13], [27] and further enhance different abilities, such as recall and numeracy [12], [14]. Thus, we see the opportunities of embedding data visualization in short-form videos to enhance the creativity and expressiveness of the story. However, embedding data visualizations in short-form videos is not easy. Unlike previous data videos with data-driven motion graphics, short-form videos are usually connected to a real-world context. Moreover, different from carefully editing the video after shooting it, people want to share their data videos at that moment and location [2].

Furthermore, embedding data visualization in short-form videos might pose more challenges to amateur video creators, especially for those who do not have experience in creating data visualization. Limited data visualization creation experience might prevent them from using data visualization to express themselves in short-form videos. As a result, in this work, we explore the design and propose an in-situ authoring tool that helps non-experts create a short-form video with their data to express themselves.

2.2 Video Post-editing Techniques

To embed data visualizations to short-form videos, powerful post-editing tools, such as Adobe After Effects [28], are used by professional designers [8], [15], [20]. Designers can first take or collect video clips and then import them

to post-editing tools to manually add data visualization to the clips. Besides simply overlaying charts and figures on the video footage, designers could add 2D visuals in 3D space following the camera movements by using advanced camera and motion tracking techniques, such as Blender motion tracking. However, manually editing frame-by-frame is time-consuming and tedious. More importantly, embedding data visualization requires solving the overlap between newly added data visualization and the objects in the original scene of the video [29]. Researchers recently tried to adopt advanced techniques to lower the difficulties of embedding data visualization into videos. For example, Tang *et al.* [20] created Smartshots for video creators to add data visualization automatically to a suitable position and time in a given video using a layout optimization-based algorithm. Chen *et al.* [21] contributed VisCommentator to ease the creation of adding visual marks on table tennis videos by utilizing machine learning-based methods.

However, most of these post-editing tools are desktop-based, requiring a recorded video as input. Users also need to switch back and forth between video-taking apps and video editing tools. These steps hinder creators from embedding data visualizations and sharing their video stories in situ. Thus, we explore how to combine video taking and editing for people to create data videos promptly in situ.

2.3 Authoring Data Visualization in Mobile AR

Besides post-editing, people may author the data visualizations before the video-taking process, for example, using tools like DXR [30], IATK [31], VRIA [32], RagRug [33], and FlowImmersive[3] to pre-configure the data visualizations in AR applications. However, pre-configuring data visualizations needed to be performed using Unity or Web Technology on the desktop before being deployed to the AR devices. These tools do not directly support users in creating a visualization in the AR environment. By supporting direct authoring in AR environments, users can 1) adjust the background based on the visualization created, 2) avoid frequent switching between the desktop application and camera for adjustment, and 3) create and share the result promptly without the need need to switch devices and possibly locations.

Therefore, authoring in the mobile AR environment is getting more attention and has been used in different applications, such as robotics [34], prototyping [35], [36], and storytelling [37]. By using different sensors and algorithms, machines can understand the scene and allow us to "interact" with reality, such as by placing visual objects on the table [38] and calculating the distance between two objects [24]. Data visualization researchers also actively explore the possibility of mobile AR authoring. Chen *et al.* [23] built MARVisT for the non-experts to create custom glyphs for personal data visualization using mobile AR. However, these works only focus on data visualization creation, which lacks visualization sequence control for storytelling with data visualization [25]. To this end, we focus on using mobile AR authoring to allow users to not only create data visualization but also compose data video in-situ.

3. https://flowimmersive.com

3 FORMATIVE STUDY

To understand the motivations and requirements for amateur video creators to add data visualizations in short-form videos, we conducted a formative study.

3.1 Participants

We recruited 10 participants (three females, aged between 19 and 31) from the local university (six participants) and Reddit (four participants) through word of mouth and advertisement respectively. The Reddit subforum where we posted our recruitment ad is r/SampleSize, in which research participants are usually recruited. Participants are all amateur video creators and reported an average of 3.25 years of short-form video platform usage experience. The distribution of the number of videos created is 1-10 (4), 11-20 (2), 21-30 (0), and >30 (4). For data visualization creation, the distribution of participants' experience is None (2), Novice (2), Knowledgeable (1), Proficient (3), and Expert (2). All of them have no experience with mobile AR authoring. Overall, participants have experience making short-form videos, diverse experience in data visualization creation, and no experience in mobile AR authoring. Participants received an Amazon.com eGift card of $20 each as compensation for the 60-minute study.

3.2 Procedure

The formative study contains two parts: (part 1) a semi-structured interview for understanding motivation and (part 2) a design probe activity for getting initial feedback on the design of a mobile AR authoring tool. All sessions are done through Zoom.

Part 1: Semi-structured Interview. The interview covered questions about participants' scenarios and procedures for creating short-form digital stories, where they archive the moments and express themselves, and the need for adding data or facts to short-form videos. Then, we demonstrated one short-form video with data visualization created by the authors and obtained feedback and the challenges of adding data visualizations to short-form videos. Responses were recorded using written notes, audio, and video.

Part 2: Design Probe. We adopted the design probe activity from [39] to elicit participants' feedback on the potential workflow for authoring short-form videos with data visualizations using mobile AR. A design probe is employed as a means to collect user requirements and as feedback on a particular application in an iterative design cycle [40]. Four authors discussed and designed three different design probes for creating short-form videos with data visualizations using mobile AR. Figure 2 shows the three design probes. All the authors have experience in short-video creation, AR authoring, and data visualization.

Specifically, a video creation pipeline has three stages: the pre-stage (before taking the video), the video-taking stage, and the post-stage (after taking the video). The three design probes emphasized the AR authoring of data visualizations at each different stage in the pipeline. Additionally, in the pipelines, we provided different common interactions (*e.g.*, tap, sketch) that were used to create 3D content on the mobile AR applications [24], [36], [38], [41], [42]. We also
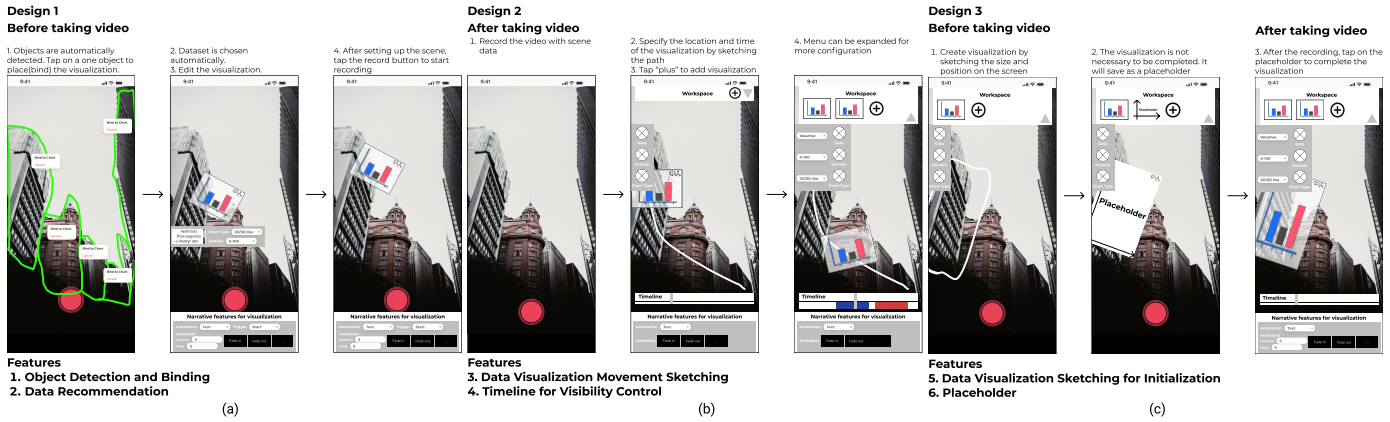
Fig. 2. This figure shows the probes we provided in the formative study. The three probes are (a) pre-stage focused approach with *object detection & binding* and *data recommendation* features, (b) post-stage focused approach with *data visualization movement sketching* and *timeline for visibility control* features, and (c) hybrid approach with *data visualization sketching for initialization* and *placeholder* features. A clearer and larger version can be found in the supplementary material.

provided different levels of automation (*e.g.*, with/without automatic scene understanding) in the visualization creation process to inspire features that balance the speed and quality for achieving the "moments". The steps and features (F1-F6) of each pipeline are described below:

- The first pipeline (Figure 2(a)) emphasizes AR authoring and preparing data visualizations before taking the video. The objects in the scene are first automatically detected (F1). Next, creators can tap on one object to place the visualization, where the dataset is also automatically chosen based on the semantic meaning of the object (F2). Then, creators can fine-tune the visualizations using AR authoring. Finally, creators can tap the record button to start the video recording. This **pre-staged focus** approach is inspired by Apple Reality Composer and SceneAR [37], where they support in-situ editing of AR content.

- The second pipeline (Figure 2(b)) allows creators to add data visualizations after taking the video. Creators first take the video where the information about the 3D scene is recorded. After taking the video, creators can specify the location and time of the visualization in the video by sketching the path in the AR authoring (F3). Finally, creators can edit the visibility time of each visualization (F4). The **post-staged focus** approach is inspired by Pronto [38] and Rapido [36]. Their techniques support the 3D reconstruction of the scene after taking the video.

- The third pipeline (Figure 2(c)) supports a **hybrid** approach by asking participants to add a placeholder for data visualization before taking a video and edit the details after taking the video. Specifically, creators could create visualizations by sketching the size and position on the screen (F5). However, it is not necessary to complete the visualization. The placeholder mechanism (F6) is inspired by PapARVis Designer [43], where a placeholder for the dataset was introduced to allow users to fill in the data later. Similarly, creators may not have clear ideas initially, but need to record that moment first. Creators can edit the data visualizations after taking the video.

After introducing all pipelines and features to participants, we asked them to describe how they would create short-form videos with data visualization and comment on the pros and cons of each feature.

3.3 Findings

Two authors analyzed the feedback using affinity diagramming [44]. Here we report the findings of the motivation for adding data visualization and initial feedback on the probes.

Motivation of Adding Data Visualization. In terms of motivation, participants mentioned that adding data visualizations to short-form videos can 1) emphasize the message they want to convey with more explanation and elaboration (5), 2) increase the credibility of the story they share (2), 3) make the video cool, eye-catching, professional, and intuitive (2), and 4) reduce the amount of voiceover (2).

In terms of scenarios, participants thought that adding data visualizations could 1) reveal additional information (*e.g.*, angles when dancing, the nutrition or amounts of food when cooking); 2) visualize big data (*e.g.*, the stock price over the years); 3) share self-tracking data (*e.g.*, daily steps, calories consumed during a workout, travel distance). These scenarios, covering both work and casual moments, span everyday life, illustrating a wide range of needs for adding data visualizations to the video to assist storytelling.

Feedback on the Design Probes. All participants were new to mobile AR authoring and were optimistic about using mobile AR authoring to add data visualization to their videos. Each of the three workflows was preferred by some participants. Based on the responses, five participants preferred to have data visualization authoring at the pre-stage because it can help them prepare a better camera angle for storytelling. For example, P3 imagined that "*when I am finding a camera angle, I wish I could preview the effect of adding a bar chart [in the target position] from different angles. It may help me to find a better shooting angle.*" Other five participants preferred post-editing because they were more familiar with fine-tuning the video in the post-editing stage. Moreover, some suggested providing the AR authoring in the pipeline at both stages. For example, P2 commented, "*I think there are both needs for the pre- and post-stages. The pre-stage assists me in understanding the scene and taking a good video [with the visualizations], while the post-stage helps me better modify the video to emphasize the message I want to share.*"

Summary. Here is the summary of participants' feedback on the features proposed across the three pipelines.

- **F1**: Object Detection and Binding. All participants thought it helped place the visualization and make the creation process efficient. Moreover, they wanted the visualization to move smoothly, implying the need for the object tracking function.
- **F2**: Data Recommendation. Four participants commented that dataset recommendation was good when the dataset was not prepared. However, five participants were concerned that "*it is tedious to change the dataset if the system recommendation is incorrect,*" since the dataset and intention could not always be inferred by the scene.
- **F3**: Data Visualization Movement Sketching. Four participants supported having this feature, while others worried that it would be inconvenient to use fingers to specify the 3D path in the AR environment due to the limited size of the mobile screen. However, they mentioned wanting to edit the visualization in the 3D scene.
- **F4**: Timeline for Visibility Control. All participants agreed that controlling the sequence and duration of the visualizations was important. Additionally, they hoped that the feature could be easy to use.
- **F5**: Data Visualization Sketching for Initialization. Five participants raised several issues about this feature. For example, the mobile phone screen size is limited, and it is tedious to sketch using fingers. Participants preferred adding visualization by tapping.
- **F6**: Placeholder. Six participants voted for this feature, as it could help them maintain quality and capture the moment. Specifically, they could find a better angle and customize the details later.

3.4 Design Requirements

Based on the findings from the formative study, we have distilled four design requirements (DR1-4) and adopted three features, *i.e.*, F1, F4, and F6.

DR1: Integrate scene understanding for better blending data visualization and reality. Scene understanding is important for AR to locate virtual objects [45]. Based on the probe study, participants were highly optimistic about the object detection and binding feature (F1). Since this feature is convenient and helpful for content creation, it can provide some insight for participants in placing visualizations in reality. Moreover, they hope the visualization can be auto-sizing instead of sketching, which requires detecting the features and context of the environment.

DR2: Enable AR (in-situ) visualization configuration. To better align the visualization creation and the 3D real environment, AR authoring is necessary [46]. When describing the placeholder feature (F6), participants found the AR editing practice interesting. Participants stated that having an AR configuration stage is similar to preview features. They would like to place the visualization before shooting the video, which could help them find a better shooting angle. Moreover, participants mentioned that AR authoring could serve as low-fi prototyping.

DR3: Allow quick dataset importing and visualization editing. One of the important aspects of short-form video creation is fast creation because people would like to make short-form videos to share particular moments [2]. Therefore, participants should be able to access datasets and create charts quickly. In addition, since most short-form video creators do not have much experience in data visualization construction, the tool should expose users to a simple interface instead of showing low-level details, such as bar width and data transformation [47].

DR4: Support visualization fine-tuning and time control in the post-editing stage. Most participants are familiar with post-editing in current practice. Participants mentioned that the timeline (F4) is important for visibility control during post-editing. Moreover, due to the fast pre-stage authoring to achieve certain moments, people need to fine-tune the visualization and timing of the appearance of the visualization afterward. Thus, we need to support the post-editing stage for visualization fine-tuning and visibility control.

4 VISTELLAR

Informed by the design requirements, we built VisTellAR for amateur video creators to create short-form videos with data visualization. Different from the previous workflow (subsection 2.2) that requires users to add data visualizations to an existing video using post-editing techniques [20], [21], VisTellAR allows users to preview and perform real-time data video composition with reality as background. The overall workflow and comparison to the existing approach are shown in Figure 1.

4.1 Pipeline and Implementation

Data Preparation. To support the quick use of personal datasets (DR3), we use Google Fit API and Weather API to provide users' health data and weather data according to users' current IP addresses, respectively. We fetch the data and store it in the local device every time users enter the authoring screen. To increase the diversity of data, we also provide a subset of offline sample datasets from DXR [30].

AR Authoring. To support AR Visualization Configuration (DR2), we utilize Unity to maintain the 3D scene and Unity ARFoundation to utilize ARCore or ARKit to maintain the relative coordinates of the 3D scene and reality using device tracking. To allow users to anchor the visualizations on physical objects and environment (DR1), we used ARFoundation's plane detection. Moreover, we used MediaPipe BoxTracking [48] for real-time object detection to allow users to tap and bind single or multiple visualizations to the environment by referencing physical objects. We then estimate the final 3D coordinates of the detected object using ARFoundation's feature point detection. Users can further update the data visualization with the 3D editing widget.

Visualization Configuration. After specifying the location of the visualization, users can adjust the visualization configuration, *i.e.*, data, mark, axis, and behavior. To simplify the visualization configuration (DR3), we utilized DXR [30]. It adopts a high-level visualization grammar similar to Vega-Lite [49], which supports easy and rapid visualization creation and allows the visualization to be updated based on changes in the visualization configuration widget (Figure 5) at runtime. To reduce the complex configuration, we provide a partial specification to the users [47] and support the top three common types of charts, *i.e.*, line chart, bar chart,
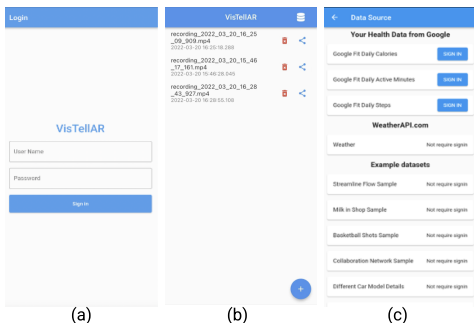
Fig. 3. This figure shows the application interface of VisTellAR, including the login screen, review & share screen, and dataset management screen.
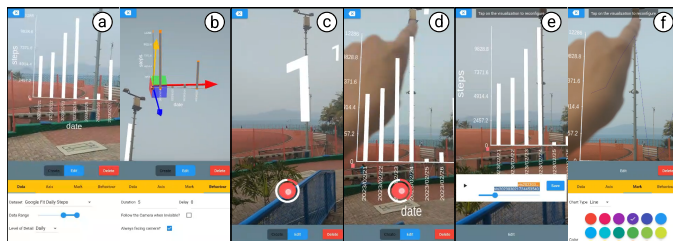

Fig. 4. This figure illustrates an authoring process. (a-b) VisTellAR detects planes and objects for users to anchor visualizations in reality. Users can edit the data, mark, axis, and behavior. (c-d) During video-taking, users can voice over, perform hand gestures, and see a count-down that notifies them when the visualization will be shown. (e-f) After taking the video, a timeline is shown to indicate when visualizations take place in the video. Users can reconfigure visualizations if needed.

scatterplot, for personal data representation [6]. To support our needs, we have extended DXR [30] to support line charts (the original version only supports bar charts and scatterplots) and deployment on mobile devices.

Data Visualization Post Editing. To support post-editing directly to the data visualization (DR4), we saved the position of the device and visualizations and the raw video frame at every timestamp. Adapting from Pronto [38], these records allow users to reconfigure data visualizations in the recorded 3D environment of the video.

4.2　User Interface Design

VisTellAR consists of two primary interfaces, including the *Application Interface* and the *AR Authoring Interface*.

Application Interface contains three screens: *Login Screen* (Figure 3(a)) allows users to log in with a specified account, which is also used for indexing the log data. *Review & Share Screen* (Figure 3(b)) lists the recorded videos for users to review and share. The button on the top-right corner navigates users to the *Dataset Management Screen*. *Dataset Management Screen* (Figure 3(c)) provides a list of real-time personal and static datasets (DR3), which can be used for authoring data visualizations in the video. Users can tap on the list to review the dataset and click the login button to authorize the tool to retrieve the personal dataset from online resources (*i.e.*, Google Fit).

AR Authoring Interface contains *Pre-stage Authoring Screen* (Figure 4(a, b)), *Video-taking Screen* (Figure 4(c, d)), and *Post-stage Authoring Screen* (Figure 4(e, f)).

Pre-stage AR Authoring Screen (Figure 4(a, b)) allows users to add and edit visualizations before taking the video
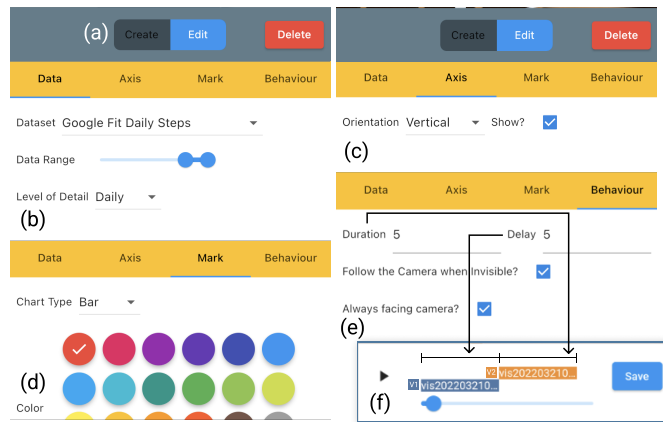

Fig. 5. This figure shows the configuration panels of the visualization in VisTellAR, including data, axis, mark, and behavior.

(DR2). Specifically, VisTellAR first detects objects and planes in the camera (DR1). It visualizes objects detected using yellowish green bounding boxes and planes detected using dotted planes. Users can then tap on one of these objects and the planes detected to embed a visualization into the 3D environment. A visualization configuration widget will then appear at the bottom, as shown in Figure 4(a, b). The widget consists of four tabs: *Data*, *Axis*, *Mark*, and *Behavior* (Figure 5). In the *Data* tab (Figure 5(b)), users can change the dataset, sort the dataset, and specify the range and level of data detail for temporal data. In the *Axis* tab (Figure 5(c)), users can change the orientation and visibility of both axes. In the *Mark* tab (Figure 5(d)), users can modify the type of visual mark and the color of the marks. Last but not least, in the *Behavior* tab (Figure 5(e)), users can change the behaviors of the AR visualizations during video taking: *i.e.*, visibility (when the visualization should be visible), location (whether the visualization should follow the camera movement after leaving the camera view), and orientation (whether the visualization should always look at the camera). The visibility control has two variables: *duration* and *delay*. With these two variables, users can show visualizations sequentially or simultaneously. For example, as shown in (Figure 5(f)), the yellow bar (V2) has the *delay* of 5, which means that the visualization (V2) appears after five seconds when the video starts; and *duration* of 5, which means that visualization (V2) appears for five seconds. With this setting, audiences can see V2 appearing after V1.

Video-taking Screen (Figure 4(c,d)) After setting up the scene with data visualization using *Pre-stage Authoring Screen*, users can start taking a short-form video for up to 30 seconds by pressing and holding the circle and entering *Video-taking Screen*. To inform users about when the visualization appears, we display a number counting down at the position where the visualization is going to appear. Users can also voice over and perform actions to the data visualization during video shooting.

Data Visualization Post-editing Screen (Figure 5(e, f)) After taking the video, VisTellAR provides a timeline for users to fine-tune the video (DR4). The duration of all visualizations is visualized on the timeline widget to assist users in locating the moments involving the visualizations. When multiple visualizations are added, there are multiple bars

to represent each duration correspondingly. For example, as shown in Figure 5(f), there are two visualizations added in this video, and we can see their visible time by the blue and yellow bars at the top of the timeline. Users can tap on any visualization to update its configurations. The exact configuration widget as in the *Pre-stage AR Authoring Screen* is shown, where users can modify the dataset, axis, mark, and behaviors again. However, users cannot adjust the position, rotation, and scale of the data visualization.

4.3 Usage Scenario

We introduce Mary, a student interested in sharing personal moments using short-form videos. Mary actively shared her exercise records about her diet plan on social media. One day, Mary skipped the exercise because of the high temperature. After exercising today, she decided to explain to her friend on social media why she skipped sharing her exercise records one day. To make the story more vivid and convincing, she decided to combine the story with her daily steps data and weather data. Mary opened the VisTellAR and went to the *Dataset Management Screen* (Figure 3(c)) to check if the data was available. After confirmation, she entered the *Pre-stage AR Authoring Screen* and scanned the environment. After scanning, objects and planes were detected and visualized and Mary could tap on the plane to initialize a visualization as shown in Figure 4(a). By using the AR authoring interface (Figure 5), Mary created a bar chart to show her step data. After that, Mary initialized another visualization to show the weather data. Since weather data was related to the sky and provided a cleaner background, she entered "Edit" mode, tapped on the visualization, and dragged the green axis to move the visualization upward to the sky (Figure 4(b)). Then, to allow the visualization to appear at different moments, Mary tapped the Behavior tab and set the delay of the second visualization to "15" because she wanted to introduce the step data first. During video taking (Figure 4(c)), Mary introduced the background before the visualization popped up. When the visualization was presented, she pointed to the chart to emphasize the big difference in her step data in the last few days (Figure 4(d)). She explained that the reason for skipping recent exercises was the bad weather, using the second visualization as a reference. After taking the video, Mary reviewed it carefully (Figure 4(e)) and thought that a line chart would be a better fit and the visualization's color might be better if it was in purple color that contrast to the background. She then checked out the frame when the visualization first appeared and tapped on the visualization. Mary updated the chart type to a line chart and the color to purple in the Mark tab (Figure 4(f)). After fine-tuning, Mary saved the video and immediately shared it with her friends in the playground. The whole process took around five minutes.

5 TAKE-HOME EXPLORATORY STUDY

To examine the effectiveness and usability of VisTellAR, we conducted a two-day take-home exploratory study inspired by the study evaluating a mobile AR authoring tool in [37]. The reason for adopting the take-home study is to expose participants to their living environment [50], an essential property for short-form video taking.

	Day 1	Day 2
Morning	**Onboarding** 1. Introduce the background 2. Tutorial 3. Trial of the app **Video Taking** (Avg. 3 videos)	**Video Taking** (Avg. 3 videos)
Night	**Short Interview** 1. Report on the video created 2. Gather interesting findings	**Debriefing Interview** 1. Report on the video created 2. Gather interesting findings 3. Report on the overall usage experience

Fig. 6. The figure shows the process of the two-day take-home study. Participants need to create at least six videos in total in two days.

5.1 Participants

We recruited 12 participants (5 females, 7 males, aged 21 - 29) who did not participate in the formative study and had experience in short-form video creation through word of mouth. They are college students from three countries with diverse backgrounds in video creation, visualization creation, and mobile AR usage experience. They have an average of 2.25 years of using the short-form video platform (mainly TikTok and Instagram Reels). The self-rated expertise on short-form video creation is rated as follows: None (0), Novice (4), Knowledgeable (5), Proficient (3), and Expert (0). For the authoring experience of data visualization, the distribution is None (4), Novice (0), Knowledgeable (6), Proficient (1), and Expert (1). They mainly used Microsoft Excel to create data visualization. For the mobile AR usage experience, the distribution is as follows: None (2), Novice (5), Knowledgeable (3), Proficient (1), and Expert (1). Pokemon Go is the mobile AR app they reported most to be used. All participants' Android (2) or Apple (10) mobile phone models are checked to be capable of using AR functions. Each participant received a $40 Amazon.com eGift card as compensation for their time after completing the study.

5.2 Procedures

The two-day take-home study consists of the onboarding session, the main study, and the debriefing session (Figure 6). All sessions were done through Zoom.

Onboarding Session. We invited each participant to join an onboarding session for approximately 45 minutes on the morning of the first day. In the beginning, we provided participants with a consent form and a demographic questionnaire to obtain their consensus and background information. We then distributed the prototype via Testflight for iOS users and Firebase App Distribution for Android users. After that, we gave a tutorial using presentation slides to demonstrate all features of the tool. At the end of the onboarding session, we asked participants to create a video to ensure that they understood the materials and were able to use the tool. Participants were free to ask any questions and seek help during this session.

Main study. After the onboarding session, participants were given two days to use VisTellAR to create six short-form videos with data visualizations. To boost their creativity and explore a broader range of usages of VisTellAR, we

encouraged participants to try different datasets given in the app to create short-form videos in different places (*e.g.*, at home or outside) and at different times (*i.e.*, morning, afternoon and evening) during the onboarding session. We currently provided participants with 1) their own health data through Google Fit API, authorized through a Google account, 2) one-week historical weather data based on their IP address provided by WeatherAPI.com, and 3) sample datasets from DXR [30] (*i.e.*, Streamline Flow, Milk in the Shop, Basketball Shots, Collaboration Network, Different Car Model Details, Bookshelf Keywords). To not limit the participants from the given datasets, we allowed them to imagine that the sample data were related to themselves or think about any other dataset if no dataset given was suitable to their stories. To keep participants' memories fresh, we organized a short interview on the first night. We first asked them to fill in questionnaires about the basic information (*e.g.*, what the story is about) for each video they created on the first day, followed by three five-scale Likert scale questions on the effectiveness of the visualization and usefulness of the VisTellAR. We then asked for interesting cases and issues that participants encountered.

Debriefing Session. On the night of the last day of the study, we first asked them to complete questionnaires about the created video, the same as on the first day. We then conducted a semi-structured interview to ask questions on the usefulness and effectiveness of the tool based on their two-day experience. Questions include but are not limited to, the likes and dislikes of the tool, their patterns of behavior when using the tool, more usage scenarios with VisTellAR, and stories they want to create beyond the two-day study. Also, we asked for the reasons for their particular behavior patterns (*e.g.*, does not speak during video taking). Lastly, participants were asked to complete a questionnaire on effectiveness and usability with ten 5-point Likert scale questions, adapted from the USE questionnaire [51]).

5.3 Data Collection

The data we collected during the user study are listed below: 1) short-form videos they created, 2) questionnaires about each video and VisTellAR, 3) interview recording and notes, and 4) event data (*e.g.*, the start time and end time of each stage of the authoring, the buttons clicked) in VisTellAR during the two-day study.

6 RESULTS

We first describe the effectiveness (*i.e.*, whether participants could use VisTellAR to quickly add visualizations for storytelling) using the general log data, and questionnaire responses, together with a gallery of selected screenshots of videos created by participants. We then report on the usability of VisTellAR using both questionnaire and interview responses. We finally provide usage patterns by analyzing detailed video content, event logs, and interview responses.

6.1 Effectiveness

All 12 participants successfully created 75 short-form videos with data visualizations in total using VisTellAR. The average duration of the videos created is 14.8 seconds (SD=6.4

seconds). Based on 5-scale Likert responses to the questionnaire, as shown in Figure 8, participants were able to use embed data visualizations (M=4.28, SD=0.548) to help express the story (M=4.31, SD=0.541) in the short-form video. Participants also agreed that VisTellAR was effective for them to add visualizations for storytelling. For example, P11 was surprised that "*before taking a video, I doubted the usefulness [of adding data visualization]. However, after taking a video [with visualization], it is useful. It can greatly blend and complement my content [in the video]. The dataset provided is quite close to my daily life, and the content of my video is also close to daily life. So, the visualization can fit into my video.*" In particular, it helped participants add data visualizations to the video (Figure 8, Q1, median: 5), and use these visualizations to express their stories (Figure 8, Q2, median: 5) in a short period of time (Figure 8, Q3, median: 4). Based on the user logs, the average time for authoring is 5.58 minutes. Two participants were neutral, and one slightly disagreed that the duration of authoring was short compared to adding AR face filters, which take less than seconds.

In addition, in terms of satisfaction, participants agreed that they could create the videos with data visualization as they wanted (Figure 8, Q4, median: 4), and the tool was fun to use (Figure 8, Q6, median: 4.5). For example, P11 agreed, "*it is convenient because I don't need to export to a computer and the visualization settings are enough for me to create the visualization I want,*" and P1 commented, "*it is fun and intuitive. It can show information while taking the video.*" Participants would also recommend the tool to friends (Figure 8, Q5, median: 4). On the other hand, some participants were not satisfied with the current state of VisTellAR due to limited visualization control (*e.g.*, *cannot change the margin between bars*) and limited personal datasets. For example, P5 commented, "*If we can have more customized data close to my everyday life apart from daily steps, it is better, and I can create more attractive stories.*"

Gallery. Figure 7 presented screenshots of five interesting short-form videos of participants (picked by the authors and consented by participants) using different datasets at different locations. In (a), the participant was pointing toward the current warm temperature to explain why wearing the jacket behind. In (b), the participant was running through a visualization that showed the daily calories consumed. In (c), the participant was standing in front of a bike-sharing station, indicating that the correlation between the weather and the bike rental data explains the decrease in the bike rental rate due to the decrease in the temperature. The participant in (d) used turning the page of a book as the trigger by setting delays and turning the page of a book while the delay counts down is counting down to 0 to show the visualization of keywords and their counts in the book in the video. Finally, in (e), the participant used the sample data about the streamline flow to show the wind field around the air conditioner. The videos can be found in the supplementary materials.

6.2 Usability

For usability, participants strongly agreed that VisTellAR was easy to learn (Figure 8, Q9, median=5) and remember (Figure 8, Q10, median=5). For example, P3 commented,
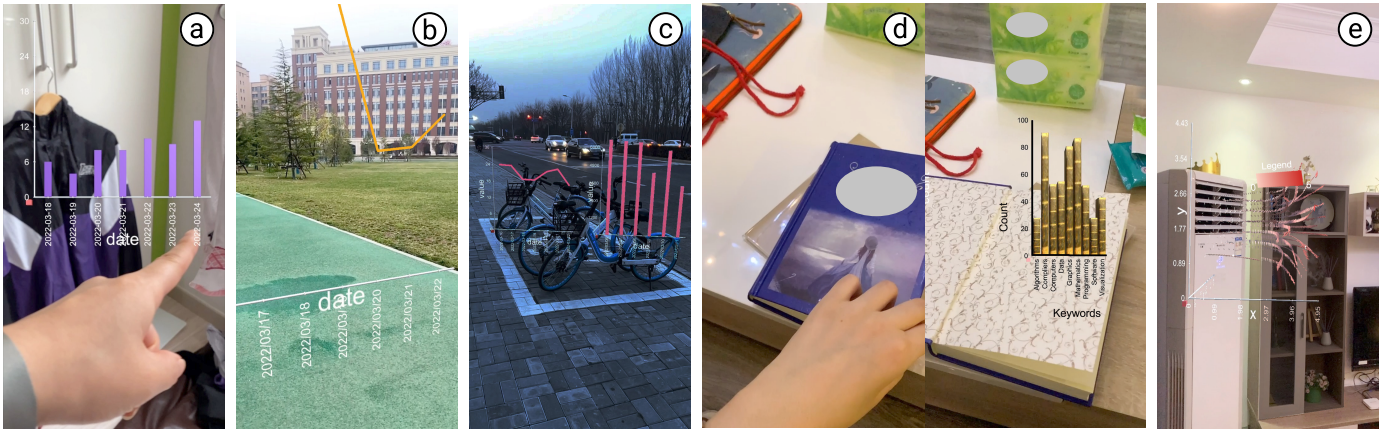
Fig. 7. This figure shows screenshots of five interesting short-form videos created by the participants: (a) pointing toward the current warm temperature to explain why wearing the windbreaker; (b) running through a visualization which showed the daily calories consumed; (c) showing the correlation of the weather and the bike rental data to explain the decrease in bike renting because of the decrease in the temperature; (d) turning a page to demonstrate the keywords and their counts in the book; and (e) visualizing the streamline wind data from the air conditioner.
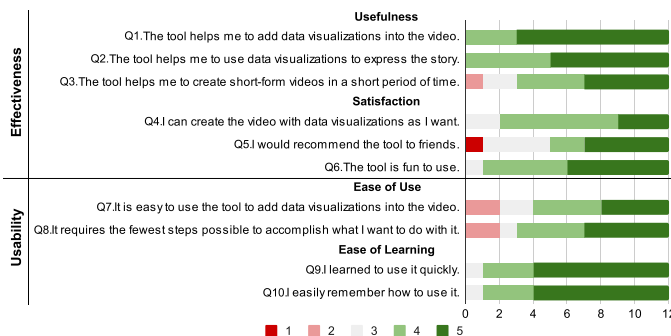

Fig. 8. This figure shows the questionnaire responses on the effectiveness and usability of VisTellAR. (1:strongly disagree; 5:strongly agree)

"*The method of freely placing charts in the shooting stage is very fresh, which undoubtedly lowers the threshold compared to post-editing.*" Most participants agreed that VisTellAR was easy to use (Figure 8, Q7, median=4) and requires fewest steps (Figure 8, Q8, median=4). However, some participants mentioned that 3D manipulation widgets were complicated to use. For example, P1 stated, "*the rotation axis is hard to control. I accidentally rotated with two axes and got lost.*" P9 added that "*It can be used, but it is not as simple as manipulating in two-dimensional space on a computer.*" Moreover, P4, P5, and P6 commented they encountered difficulty configuring charts in AR mode. P5 suggested that "*it would be better to provide maybe graph mode or something that allows me to edit graphs [such as] size and angle or chart types without any scenery.*" To involve multiple charts in a user-friendly way, P6 suggested "*to take multiple clips in different places with different visualizations [and merge them].*"

6.3 Usage Patterns

To further gain insights into how participants expressed their stories with data visualizations using VisTellAR, we investigated the usage patterns of VisTellAR. Two authors coded the videos on the following aspects: the dataset used, with or without the voice-over, the object the visualization was bound to, whether the data was semantically related to the object bound, whether the data visualization supported the story, whether it involved camera movement, whether

participants interacted with data visualization, whether there was the time control for the visualization duration, and whether the video involved multiple charts. We also analyzed the event data (*e.g.*, the start and end times of each authoring stage and the buttons tapped).

AR Pre-stage Authoring and Post-editing. Participants spent more time in AR pre-stage authoring than post-editing while using VisTellAR. They spent an average of 3 minutes 30 seconds (SD=3 minutes 43 seconds) for AR pre-stage authoring and an average of 1 minute (SD=1 minute 21 seconds) for post-editing. The split point between the pre-stage authoring and post-stage authoring is the first time the participant clicks the "record" button. Participants were satisfied with the pre-staged authoring, which is unusual compared to the traditional video-taking and editing process. AR pre-stage authoring allows users to preview and edit the scene, which is more intuitive and effective. For example, P8 mentioned, "*setting before taking the video is more natural ... I prefer to have everything ready before taking the video.*" and P9 explained, "*using the advantages of object detection in AR pre-stage authoring help me to create the video without post-editing.*" Participants also found post-editing necessary. For example, P5 said, "*although I set the duration and delay beforehand, sometimes my speech speed didn't meet the expected duration and delay time, so it was very helpful that I could edit them after I made the videos.*"

Relationships between Short-form Videos and Data Visualizations. We further analyzed the relationships between the video content and the data visualizations used. Specifically, we were interested in the *dataset* used, the *number of data visualizations used*, and *visualization task*

Datasets. In most cases (65/75 videos), they were semantically related to the context of the videos. We observed that Google Fit (29/75 videos) and Weather Data (20/75 videos) were used most frequently due to the familiarity of the data. However, while sample datasets, such as bookshelf (5/75 videos) and milk (7/75 videos), are also related to daily life, they depend highly on individual location and habits. In those cases where the datasets provided were not related to their locations or habits, some participants chose to use the visualization data agnostically (13/75 videos), meaning that

they only used the graphical shape of the visualization without relying on a concrete dataset or data schema [52]. During interviews, participants suggested adding more datasets, such as environmental data (*e.g.*, the current distribution of people), personal objects data (*e.g.*, the number of clothes), Covid 19 pandemic data, and social data (*e.g.*, the number of chats with friends on social media apps).

Number of data visualizations used. Participants used one (62/75 videos) or multiple charts (13/75 videos) to express their stories. For example, participants reveal the correlation between the decreasing trend of temperature and bike rental rate by showing two decreasing charts (Figure 7(c)). Note that in this video, the creator used his daily steps dataset to simulate the decreasing trend of bike rental rate. P3 further commented, "*I tried to use multiple charts and found that two charts and three charts can better connect the stories, and charts can express the video content more abundantly so that the video viewers have a more intuitive understanding.*"

Visualization Tasks. The results reveal that participants could communicate information using data visualizations, given that only half of the participants were familiar with data visualizations. For example, they described the data distribution (41/75 videos) and the trend (16/75 videos). In addition, some participants could even use multiple graphs to present complicated analytics tasks such as comparison (4/75 videos) and correlation (2/75 videos) between two datasets in a short-form video, as shown in Figure 7(c).

Object Binding. Most visualizations (66/75 videos) were bound or anchored to physical objects, suggesting that the object detection feature was useful. P10 commented, "*it is convenient to anchor the data to the object.*" The relationship could be directly related (38/66 videos) or indirectly related (28/66 videos). Directly related means that visualizations are semantically related to target objects, for example, placing bike rental data next to some bikes (Figure 7(c)), putting visualizations with book data next to the book (Figure 7(d)) and putting air flow data next to the air conditioner (Figure 7(e)). Indirectly related means putting visualizations in the environment or objects semantically indirectly related to the data. For instance, putting visualization about the weather next to the jacket (Figure 7(a)), putting calorie data on the playground (Figure 7(b)), and referring to the weather data next to the bikes (Figure 7(c)).

Time Control. 24/75 videos added the time control of the visualizations and delayed the appearance of the visualizations. The delay before the visualization appears was mainly added for delivering the background description, allowing showing interaction (*e.g.*, turning a page of the book), or avoiding overlapping with the previous visualization. For example, P11 attempted to open a book and show the visualization afterward (Figure 7(d)). Moreover, with the countdown feature, P8 commented, "*that's intuitive to use. It [countdown feature] helps my storytelling. That wasn't hard to control the delay for me [in the pre-stage authoring] ...*". However, it was not true for every participant. P8 continued, "*... but it will be better to tap to make the visualization visible*". P6 echoed, "*it is difficult to think about the timing*" and suggested displaying the visualization by predefined gestures.

Interaction with Data Visualization. We observed four types of interaction with data visualization in short-form videos. The first is *describing verbally*. Most participants (8/12) chose to tell stories (53/75 videos) and refer to data visualization to increase credibility and add more information. The second is *gazing* only. Four participants have created videos that only put the data visualization on-screen without talking. P8 stated that she seldom talks while taking short-form videos because "*it is hard to describe the data visualization in noisy places.*" P12 added that she did not speak while filming the scene. P4 and P12 continued that a well-designed data visualization should be self-explanatory. All four participants asked for the features of adding text and annotations for both the visualization and video in the short-form videos. The third is *camera rotation and movements*. Participants used camera movement during short-form video taking with data visualization (45/75 videos). For example, P8 described the short-form video she made "*I took the video from bottom to up to show that I had walked so many yesterday. I am so tired now and lying on the bed.*" Instead of standing still to take the videos, participants even involved camera movements (*e.g.*, walking and moving towards visualizations) in their videos (18/75 videos). For example, P1 emphasized, "*I want to take the video with similar dramatic effect in the movies.*" In one of P1's videos, he walked outside of a building and a visualization of the weather. He then zoomed into the visualization and was surprised by the low temperature outside. Furthermore, P9 ran through a visualization showing calories while taking the video, as shown in Figure 7(b). Lastly, two participants tried to interact with the data visualization with *different gestures*, although it is not intentionally supported. They used the delay function to produce similar effects. P11 attempted to open a book and show the visualization (Figure 7(d)), and P8 tried to *open her hand* and make the visualization pop up, *kick* to make the visualization disappear, and also *point* towards a bar (Figure 7(a)).

7 DISCUSSION

AR pre-staged authoring and data visualization post-editing are both necessary. Based on the common video editing framework, we introduced AR pre-staged authoring. For the pre-staged authoring, participants were positive about adding such stages to their video-taking workflow. Moreover, they appreciated the preview and object alignment features that allow users to specify the scene before video-taking. While the AR pre-staged authoring moved the scene creation process beforehand, post-editing is also necessary. For the post-editing, participants mentioned that they could fine-tune the details of the video. Moreover, supporting more features in the data visualization post-editing features could help balance the time in the pre-staged because it enables a larger fault tolerance for setting a perfect scene in the pre-staged authoring.

Visualization Usage in Short-form Videos. Short-form videos favor personal visualizations. Participants were excited to use personal data [5] in their short-form videos. They have used mostly Google Fits data and Weather data. They also suggested more datasets related to personal objects, social data, and events they are aware of (*e.g.*, Covid 19) rather than complicated and academic-focused datasets. As a result, most participants were satisfied with the given

simple visualizations and used only one chart per video. The first reason was that the visualization tasks were simple. Over half of the videos mainly used data visualization to describe the data distribution. This result echoes past work on personal visualization, suggesting using simple and standard visualizations during presentation [6].

Moreover, participants mainly used one chart because it fits the duration (at most 30 seconds in our case) and the small screen size of mobile devices. On the other hand, multiple charts could better link different story pieces with visual patterns as demonstrated in Figure 7(c). While single visualization was favored in short-form videos, supporting multiple visualizations in longer-duration videos might also be important because of specific scenarios, such as educational videos, and more complex tasks, such as comparison.

Mobile AR authoring enables camera movement and novel interactions for data visualization presentation. The study found that over half of the videos were taken with camera rotation and movement. For example, a participant moved around to present a 3D visualization better (Figure 7(e)). Enabling the camera allows better 3D visualization presentations, natural data visualization blending, and engaging presentations. In AR, users can also perform hand gestures and manipulate physical objects to interact with the data visualization. In the study, two participants pointed towards the specific bar and turned a page to show a visualization, respectively (Figure 7(a&d)). Involving gestures in data visualization makes the presentation more personal and natural [53]. Moreover, manipulating physical objects for presenting data visualization could engage video takers.

Provide AR guidance for embedded data visualization design during video shooting. We observed that AR guidance during video shooting could be potentially helpful to the video-taking process. People mostly like the countdown features and request more during the video-taking process. Moreover, providing guidance could reduce errors caused during the video taking and the time spent on post-editing. We see an opportunity to provide real-time guidance to help data visualization novices take short-form videos following the data visualization guidelines.

Limitation and Future Work. There are limitations despite the encouraging results of the user study. For the prototype, since our focus is to embed data visualization into the short-form video as the first step, the current support in narration is limited. For example, some participants wanted to add text and annotations rather than voice-over for narration. Thus, more techniques in narrative visualization [10] could be implemented and added to AR authoring. Moreover, since we mainly want to understand how people use AR features and include data visualization in the short-form videos, we did not optimize the user interface and the user control (*e.g.*, using a slider widget to control the delay and duration). In addition, the number of participants in our second user study was small. In future work, we aim to involve more participants, especially expert content creators, to gather more professional insights on the editing process. Lastly, our analysis was primarily focused on the authoring experience based on the videos collected from the take-home studies. Conducting further analysis from the audience's perspective would help to confirm the findings and

assess the quality of the videos. Despite these limitations, our results are still helpful as the first work to understand participants' behaviors using mobile AR authoring to add data visualization to the short-form videos.

In the future, it is interesting to further investigate the use of camera movement for presenting data visualization as it helps present 3D visualization and increase engagement. Moreover, simplifying the pre-staged authoring process in AR remains challenging. Future work could investigate the multimodel input, such as RealityTalk [54], for data visualization creation in real time. Furthermore, we plan to integrate VisTellAR with advanced self-tracking apps, such as OmniTrack [3] and situated personal tracker [55]. As such, users can create more diverse personal data visualization with various flexible data sources, such as sleep quality and self-defined metrics, related to themselves for short-form stories. Lastly, adding AR on-the-fly guidance during video-taking could also be the next step.

8 CONCLUSION

We presented VisTellAR, a mobile AR authoring tool that helps amateur video creators to add data visualizations to enrich their short-form videos. VisTellAR was built on four design requirements from a formative study with video creators (N=10) and existing literature. VisTellAR introduced a flexible pipeline, using AR to provide both pre-staged authoring and post-editing to help users create, configure, and embed visualizations in the video. Through a two-day take-home exploratory study with another set of participants (N=12), we found that VisTellAR is effective and fun. Participants successfully created short-form videos using VisTellAR and confirmed the tool's usefulness and efficiency. Based on the participants' feedback and videos, AR pre-stage authoring lowers the barrier for ordinary people to create and use data visualizations, especially in storytelling. Moreover, the AR features bring more design opportunities to engage the audience with camera movement and interaction with gestures and physical objects to data visualizations. This work could act as a stepping stone to inspire more investigations of embedding data visualization in short-form video creations and beyond.

ACKNOWLEDGMENTS

REFERENCES

[1] "Tiktok says 1 billion people use the app each month," https://www.cnbc.com/2021/09/27/tiktok-reaches-1-billion-monthly-users.html, accessed: 2022-01-09.

[2] B. Omar and W. Dequan, "Watch, share or create: The influence of personality traits and user motivation on tiktok mobile video usage," 2020.

[3] Y.-H. Kim, J. H. Jeon, B. Lee, E. K. Choe, and J. Seo, "Omnitrack: A flexible self-tracking approach leveraging semi-automated tracking," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1–28, 2017.

[4] D. A. Epstein, C. Caldeira, M. C. Figueiredo, X. Lu, L. M. Silva, L. Williams, J. H. Lee, Q. Li, S. Ahuja, Q. Chen, P. Dowlatyari, C. Hilby, S. Sultana, E. V. Eikey, and Y. Chen, "Mapping and taking stock of the personal informatics literature," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, dec 2020. [Online]. Available: https://doi.org/10.1145/3432231

[5] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury, "Personal visualization and personal visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 3, pp. 420–433, 2014.

[6] E. K. Choe, B. Lee *et al.*, "Characterizing visualization insights from quantified selfers' personal data presentations," *IEEE computer graphics and applications*, vol. 35, no. 4, pp. 28–37, 2015.

[7] B. Lee, E. K. Choe, P. Isenberg, K. Marriott, and J. Stasko, "Reaching broader audiences with data visualization," *IEEE Computer Graphics and Applications*, vol. 40, no. 2, pp. 82–90, 2020.

[8] F. Amini, N. Henry Riche, B. Lee, C. Hurter, and P. Irani, "Understanding data videos: Looking at narrative visualization through the cinematography lens," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 1459–1468.

[9] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale, "More than telling a story: Transforming data into visually shared stories," *IEEE computer graphics and applications*, vol. 35, no. 5, pp. 84–90, 2015.

[10] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1139–1148, 2010.

[11] R. Kosara and J. Mackinlay, "Storytelling: The next step for visualization," *Computer*, vol. 46, no. 5, pp. 44–50, 2013.

[12] W. Willett, Y. Jansen, and P. Dragicevic, "Embedded data representations," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 461–470, 2016.

[13] B. Lee, D. Brown, B. Lee, C. Hurter, S. Drucker, and T. Dwyer, "Data visceralization: Enabling deeper understanding of data using virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1095–1105, 2020.

[14] W. Willett, B. A. Aseniero, S. Carpendale, P. Dragicevic, Y. Jansen, L. Oehlberg, and P. Isenberg, "Perception! immersion! empowerment! superpowers as inspiration for visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 22–32, 2021.

[15] F. Amini, N. H. Riche, B. Lee, A. Monroy-Hernandez, and P. Irani, "Authoring data-driven videos with dataclips," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 501–510, 2016.

[16] Y. Shi, X. Lan, J. Li, Z. Li, and N. Cao, "Communicating with motion: A design space for animated visual narratives in data videos," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–13.

[17] X. Lan, X. Xu, and N. Cao, "Understanding narrative linearity for telling expressive time-oriented stories," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–13.

[18] L. Yang, X. Xu, X. Lan, Z. Liu, S. Guo, Y. Shi, H. Qu, and N. Cao, "A design space for applying the freytag's pyramid structure to data stories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 922–932, 2021.

[19] B. Ens, B. Bach, M. Cordeil, U. Engelke, M. Serrano, W. Willett, A. Prouzeau, C. Anthes, W. Büschel, C. Dunne *et al.*, "Grand challenges in immersive analytics," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–17.

[20] T. Tang, J. Tang, J. Lai, L. Ying, Y. Wu, L. Yu, and P. Ren, "Smartshots: An optimization approach for generating videos with data visualizations embedded," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 12, no. 1, pp. 1–21, 2022.

[21] Z. Chen, S. Ye, X. Chu, H. Xia, H. Zhang, H. Qu, and Y. Wu, "Augmenting sports videos with viscommentator," *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[22] G. A. Lee, C. Nelles, M. Billinghurst, and G. J. Kim, "Immersive authoring of tangible augmented reality applications," in *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 2004, pp. 172–181.

[23] Z. Chen, Y. Su, Y. Wang, Q. Wang, H. Qu, and Y. Wu, "Marvist: Authoring glyph-based visualization in mobile augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 8, pp. 2645–2658, 2019.

[24] R. Suzuki, R. H. Kazi, L.-Y. Wei, S. DiVerdi, W. Li, and D. Leithinger, "Realitysketch: Embedding responsive graphics and visualizations in ar through dynamic sketching," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 166–181.

[25] A. Satyanarayan and J. Heer, "Authoring narrative visualizations with ellipsis," in *Computer Graphics Forum*, vol. 33, no. 3. Wiley Online Library, 2014, pp. 361–370.

[26] L. Danzico, "Telling stories using data: An interview with jonathan harris," vol. 1, pp. 04–11, 2008.

[27] Q. Zhu, L. Yuan, Z. Xu, L. Yang, M. Xia, Z. Wang, H.-N. Liang, and X. Ma, "From reader to experiencer: Design and evaluation of a VR data story for promoting the situation awareness of public health threats," *International Journal of Human-Computer Studies*, vol. 181, p. 103137, 2024.

[28] "Adobe After Effects," https://www.adobe.com/products/aftereffects.html, 2022, accessed: 2022-03-29.

[29] T. Tang, J. Tang, J. Hong, L. Yu, P. Ren, and Y. Wu, "Design guidelines for augmenting short-form videos using animated data visualizations," *Journal of Visualization*, vol. 23, no. 4, pp. 707–720, 2020.

[30] R. Sicat, J. Li, J. Choi, M. Cordeil, W.-K. Jeong, B. Bach, and H. Pfister, "Dxr: A toolkit for building immersive data visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 715–725, 2018.

[31] M. Cordeil, A. Cunningham, B. Bach, C. Hurter, B. H. Thomas, K. Marriott, and T. Dwyer, "Iatk: An immersive analytics toolkit," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 200–209.

[32] P. W. Butcher, N. W. John, and P. D. Ritsos, "Vria: A web-based framework for creating immersive analytics experiences," *IEEE Transactions on visualization and computer graphics*, vol. 27, no. 7, pp. 3213–3225, 2020.

[33] P. Fleck, A. S. Calepso, S. Hubenschmid, M. Sedlmair, and D. Schmalstieg, "Ragrug: A toolkit for situated analytics," *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[34] Y. Cao, Z. Xu, F. Li, W. Zhong, K. Huo, and K. Ramani, "V. ra: An in-situ visual authoring system for robot-iot task planning with augmented reality," in *Proceedings of the 2019 on Designing Interactive Systems Conference*, 2019, pp. 1059–1070.

[35] M. Nebeling, J. Nebeling, A. Yu, and R. Rumble, "Protoar: Rapid physical-digital prototyping of mobile augmented reality applications," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[36] G. Leiva, J. E. Grønbæk, C. N. Klokmose, C. Nguyen, R. H. Kazi, and P. Asente, "Rapido: Prototyping interactive ar experiences through programming by demonstration," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 626–637.

[37] M. Chen, A. Monroy-Hernández, and M. Sra, "Scenear: Scene-based micro narratives for sharing and remixing in augmented reality," *arXiv preprint arXiv:2108.12661*, 2021.

[38] G. Leiva, C. Nguyen, R. H. Kazi, and P. Asente, "Pronto: Rapid augmented reality video prototyping using sketches and enaction," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[39] K. Hara, C. Chan, and J. E. Froehlich, "The design of assistive location-based technologies for people with ambulatory disabilities: A formative study," in *Proceedings of the CHI conference on human factors in computing systems*, 2016, pp. 1757–1768.

[40] K. Boehner, J. Vertesi, P. Sengers, and P. Dourish, "How hci interprets the probes," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, pp. 1077–1086.

[41] N. Saquib, R. H. Kazi, L.-Y. Wei, and W. Li, "Interactive body-driven graphics for augmented video performance," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[42] K. C. Kwan and H. Fu, "Mobi3dsketch: 3d sketching in mobile ar," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–11.

[43] Z. Chen, W. Tong, Q. Wang, B. Bach, and H. Qu, "Augmenting static visualizations with paparvis designer," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[44] R. Hartson and P. S. Pyla, *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier, 2012.

[45] M. Billinghurst, "Grand challenges for augmented reality," *Frontiers in Virtual Reality*, vol. 2, p. 12, 2021.

[46] M. Satkowski, W. Luo, and R. Dachselt, "Towards in-situ authoring of ar visualizations with mobile devices," in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2021, pp. 324–325.

[47] L. Grammel, M. Tory, and M.-A. Storey, "How information visualization novices construct visualizations," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 943–952, 2010.

[48] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[49] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-lite: A grammar of interactive graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 341–350, 2016.

[50] D. Ledo, S. Houben, J. Vermeulen, N. Marquardt, L. Oehlberg, and S. Greenberg, "Evaluation strategies for hci toolkit research," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–17.

[51] A. M. Lund, "Measuring usability with the use questionnaire12," *Usability interface*, vol. 8, no. 2, pp. 3–6, 2001.

[52] T. Tsandilas, "Structgraphics: Flexible visualization design through data-agnostic and reusable graphical structures," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 315–325, 2020.

[53] B. D. Hall, L. Bartram, and M. Brehmer, "Augmented chironomia for presenting data to remote audiences," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022, pp. 1–14.

[54] J. Liao, A. Karim, S. S. Jadon, R. H. Kazi, and R. Suzuki, "Realitytalk: Real-time speech-driven augmented presentation for ar live storytelling," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022, pp. 1–12.

[55] N. Bressa, J. Vermeulen, and W. Willett, "Data every day: Designing and living with personal situated visualizations," in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–18.

Mingming Fan received the Ph.D. degree in computer science in 2019 from the University of Toronto in Toronto, ON, Canada. He is currently an assistant professor at The Hong Kong University of Science and Technology (Guangzhou) and The Hong Kong University of Science and Technology, where he directs the APEX (Accessible & Pervasive User Experience) group. His research interests include human-computer interaction, computational user experience (UX), aging and accessibility, and information visualization.
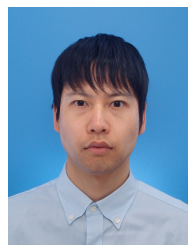
Ting-Chuen Pong is a professor of the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology (HKUST). He received his Ph.D. in Computer Science from Virginia Polytechnic Institute and State University in 1984. His research interests include computer vision, multimedia computing and IT in Education. For more information, please visit http://www.cse.ust.hk/faculty/tcpong/.

Wai Tong is currently an instructional assistant professor in the School of Performance, Visualization and Fine Art at Texas A&M University. He obtained his Ph.D. and B.Eng. degrees in Computer Science and Engineering from the Hong Kong University of Science and Technology in 2023 and 2018, respectively. His research interests include data visualization, human-computer interaction, and immersive technology. For more information, please visit https://wtong2017.github.io/.

Huamin Qu is a professor in the Department of Computer Science and Engineering (CSE) at the Hong Kong University of Science and Technology (HKUST) and also the director of the interdisciplinary program office (IPO) of HKUST. He obtained a BS in Mathematics from Xi'an Jiaotong University, China, an MS and a PhD in Computer Science from the Stony Brook University. His main research interests are in visualization and human-computer interaction, with focuses on urban informatics, social network analysis, E-learning, text visualization, and explainable artificial intelligence (XAI). For more information, please visit http://huamin.org/.

Kento Shigyo is currently a Ph. D. student in the Interdisciplinary Programs Office at the Hong Kong University of Science and Technology (HKUST). He obtained his Master's degree in Computers and Communication from Kobe University. His research interests include Data Visualization, Human-Computer Interaction, and Immersive Technology.

Meng Xia is an assistant professor at Texas A&M University. She obtained a Ph.D. degree from the Hong Kong University of Science and Technology and worked as a Postdoc at Carnegie Mellon University and Korea Advanced Institute of Science and Technology, respectively. Her research interests mainly focus on Human-AI Interaction, Data Visualization, and Education Technology. More details can be found at https://www.xiameng.org.

Lin-Ping Yuan is currently a Ph.D. candidate in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology (HKUST). She obtained her B.Eng. degree in Software Engineering from Xi'an Jiaotong University, China in 2019. Her research interests include Augmented Reality, Virtual Reality, Human-Computer Interaction, and Data Visualization. More details can be found at https://yuanlinping.top.