Augmenting Sports Videos with VisCommentator

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 1040

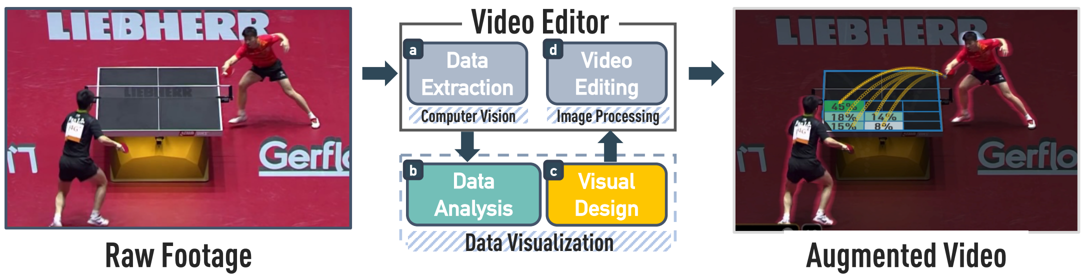

Fig. 1. Left: A raw sports video. Middle: Four steps to create an augmented sport video. Right: An augmented video reveals the probability distribution of the ball placement on the table.

Augmenting sports videos with data visualizations is gaining traction given its ability to communicate insights and explicate player strategies engagingly. However, creating such augmented videos is challenging, as it requires considerable domain expertise in sports analytics and video editing skills. To ease the creation process, we present a design space to characterize augmented sports videos at element- (*what are the constituents*) and clip-levels (*how the constituents are organized*) by systematically reviewing 233 examples collected from TV channels, teams, and leagues. The design space provides guidance for selecting data insights and visualizations for various purposes. Informed by the design space, we present VisCommentator that facilitates the creation of augmented videos for table tennis with data insights and visualizations recommendations. A user study with seven domain experts confirms the usefulness and effectiveness of the system. Another study with sports fans found the resulting videos informative and engaging.

Additional Key Words and Phrases: Video-based Visualization, Sports Data, Augmented Sports Videos, Intelligent Design Tool

1 INTRODUCTION

Video is a popular form to present and broadcast sports events. In recent years, thanks to the advance of techniques such as Computer Vision (CV) and Image Processing (IP), there has been a growing practice of augmenting sports videos with embedded visualizations [13]. These augmented sports videos combine visualizations with video effects (*e.g.*, slow motion, and camera rotation) to embed the data in the actual scenes (*e.g.*, Figure 1Right). Given the ability to communicate insights and explicate player strategies in an engaging manner, augmented sports videos have been widely used by TV channels [12], sports teams [7], and even fan clubs [30] to engage or inform the audiences.

Despite their popularity, these augmented videos, by nature, are difficult to create. Usually, these videos cannot be created without sports experts. Based on the close collaboration with sports experts who provide data analysis and consultancy services for national sports teams and TV channels, we identify that the typical workflow to create augmented sports videos mainly includes two steps: 1) analyze a sports video to obtain data insights; and 2) edit and augment the video with these data insights. We are also informed that these two steps are usually finished by sports experts—even in some cases (*e.g.*, for TV channels) the final video can be created by video editors, sports experts still need to prototype a video or sketch to communicate the insights. This creation process is particularly challenging for sports experts, as it involves many design decisions and tedious operations.

Various systems and techniques have been proposed to augment sports video over the last years. On the one hand, industrial companies have developed commercial tools, such as Viz Libero [38] and Piero [3], for TV sportscast. Although these tools provide powerful video editing functions, they target at proficient video editors (*e.g.*, who work at TV companies) and have a steep learning curve for sports experts. On the other hand, although an increasing research has been conducted on sports data visualization, visualizing data in sports videos has received relatively less attention from the community [26]. A notable exception comes from Stein et al. [32, 33], who developed a system that automatically extracts and visualizes data from and in *soccer* videos. Nonetheless, except for soccer, there are no publicly available methodologies and tools that help sports experts create augmented sports videos by providing data-driven features (*e.g.*, insights detection, visualizations recommendation, and mapping and binding between data and visuals).

To facilitate the creation of augmented sports videos, we are particularly interested in four questions that should be considered in the two creation steps: 1) In the data analysis step, *what kinds of data can be used to augment a sports video (Q1)?* Among all the available data, *what data should be presented regarding to various narrative purposes (Q2)?*; 2) In the video augmentation step, *what visual effects are used to present these data (Q3)? How to organize these visual effects in temporal order with respect to the raw video (Q4)?* From a tool-design perspective, these questions can also be reorganized into *element-level* (Q1, Q3), which are about the building blocks for augmenting sports videos, and *clip-level* (Q2, Q4), which are about the the organization of these building blocks. Without understanding these questions, the visualizations embedded in the fast-moving sports video will easily overwhelm the audience, let alone to engage them.

In this work, we focus on ball sports and aim to facilitate the creation of augmented videos for sports experts. To understand the design practices of augmented sports videos, we first systematically reviewed a corpus of 233 examples collected from reputable sources (*e.g.*, TV channels, leagues, and teams). Our analysis resulted in a design space that characterizes augmented sports videos at element- and clip-levels, which answer the aforementioned four questions. The four design dimensions (*i.e.*, Data Type, Visual Type, Data Level, and Narrative Order) of the design space and their combination frequency provide guidance for selecting data insights and visual effects for various purposes (*i.e.*, engagement or education). Informed by the design space, we design and implement VisCommentator, a proof-of-concept system that facilitates the creation of augmented table tennis videos. VisCommentator takes a raw table tennis video as the input and features two recommendation engines to semi-automatically suggest data insights and visual effects based on the narrative purpose and narrative order selected by the user. Integrated with existing video editors, VisCommentator allows sports experts to fast prototype augmented table tennis videos. A user study with seven sports experts confirms the utility and usability of the system. A group of 23 sports fans found the augmented videos created by the experts informative and engaging. We further report the experts' feedback gathered from the post-study interviews, which implies future improvements and opportunities. The corpus, created videos, and other materials can be found in https://viscommentator.github.io (clean and safe).

2 RELATED WORK

2.1 Video-based Sports Visualization

Due to its advantages of presenting data in actual scenes, video-based sports visualization has been widely used to ease experts' analysis [33] and engage the audiences [20]. Based on the presentation method, video-based visualizations can be divided into two types, *i.e.*, overlaid and embedded. This work focuses on embedded visualizations in sports videos.

Perin et al. [26] comprehensively surveyed the visualization research on sports data, indicating that only a few works can be considered as video-based visualizations. Among these few works, a representative example were developed by Stein et al. [33] for soccer. Their system takes a raw footage as the input and automatically visualizes some tactical information as graphical marks in the video. Later, Stein et al. [32] extended their work by proposing a conceptual framework that semi-automatically select proper information to be presented at a given moment. Recently, Fischer et al. [13] found that video-based soccer visualization systems from industry are actually ahead of most of the academic research. For example, Piero [3] and Viz Libero [38] are both developed for sportscasting and provide a set of powerful functions to edit and annotate a sports video. Besides, CourtVision [10], developed by Second Spectrum [31] for basketball, is another industry product that automatically tracks players positions and embeds status information to engage the audiences. In summary, as detailed by Fischer et al. [13], the research on video-based sports visualization is still in its infancy, while the strong market demand has already spawned very successful commercial systems. Nevertheless, these commercial systems target at proficient video editors and provide comprehensive video editing features, leading to a steep learning curve for sports experts. Moreover, they rarely enable data-driven designs whereas the goal of sports experts is to augment sports videos with data. Given that very little is known about the design practices of augmented sports videos, it remains unclear what data-driven features should be provided to facilitate their creation. In this work, we explore this direction and aim to ease the creation by a systematic study of existing augmented sports videos collected from reputable sources.

2.2 Data Videos for Storytelling

Data video, as one of the seven genres of narrative visualization categorized by Segel and Heer [28], is an active research topic and has attracted interest from researchers. Amini et al. [1] systematically analyzed 50 data videos from reputable sources and summarized the most common visual elements and attention cues in these videos. They further deconstructed each video into four narrative categories by using the visual narrative structure theory introduced by Cohn [11]. Their findings reveal some design patterns in data videos and provide implications for the design of data video authoring tools. Build upon of that, Amini et al. [2] further contributed DataClips, an authoring tool that allows general users to craft data videos with predefined templates. Recently, Thompson et al. [37] and Cao et al. [5] contributed two design spaces related to data videos for developing future animated data graphic authoring tools and characterizing narrative constructs in data videos, respectively.

Although these studies provide insights into data videos design, our scenario is inherently different from theirs and thus yields new challenges. Specifically, these studies use video as a medium to tell the story of data, whereas we focus on augmenting existing videos with data. An existing video imposes extra constraints on narrative orders, data visual forms, *etc.* With these constraints, how to visually narrate data insights in videos remains underexplored. Perhaps the most relevant work is from Stein et al. [32], who annotated the data in soccer videos in a linear way. To fully explore the narrative orders used for presenting data in augmented sports videos, we analyze 233 real-world examples and summarize six narrative orders and their common usage scenarios. These analyzed results provide guidance for designing creation tools of augmented sports videos.

2.3 Intelligent Design Tools

Designing visual data stories usually requires the skills to analyze complex data and map it into proper visualizations, both of which require considerable skills. Therefore, to lower the enter barrier to visual data stories, many researchers develop intelligent creation tools to automate or semi-automate the design process. To ease the visual mapping process, one widely used approach is templates. For example, DataClips [2] and Timeline StoryTeller [4] use templates manually summarized from existing examples to enable semi-automatic creation of data animation and timeline infographics, respectively. On the basis of Timeline StoryTeller, Chen et al. [9] proposed a method to automatically extract extensible templates from existing designs to automate the timeline infographics design. Besides automating the visual mapping process, some tools further facilitate the data analysis process by automatically suggesting data insights. DataShot [42] adopts an auto-insight technique to recommend interesting data insights based on their significance for factsheet generation. DataToon [17], an authoring tool for data comic creation, uses a pattern detection engine to automatically suggest interesting patterns of the input network data.

However, few, if any, tools exist that provide the aforementioned kinds of data-driven support for creating augmented sports videos. The challenges of developing such a tool not only exist in the engineering implementations but also in the integration between the workflows of visualization authoring and video editing. We draw on the line with prior intelligent visualization design tools and design VisCommentator to provide data insights and visualizations recommendations in a common video editing workflow.

3 DESIGN SPACE OF AUGMENTED VIDEOS

To understand the design practices of augmented videos for ball sports, we collected and analyzed 233 videos from TV channels, teams, and leagues, leading to a design space that characterizes augmentations at element- (*what are the constituents*) and clip-levels (*how the constituents are organized*).

3.1 Methodology

Data Sources and Videos. We collected a video corpus of six popular ball sports, including three team sports (*i.e.*, basketball, soccer, and American football) and three turn-taking sports (*i.e.*, tennis, badminton, and table tennis). Specifically, we searched the videos in Google Videos by using keyword combinations such as “breakdown videos + SPORT”, “analysis video + SPORT”, and “AR + SPORT”, where SPORT is one of the six ball sports mentioned above. The searching processing was repeated recursively on each returned site until no more augmented videos were found. To ensure the quality and representative of the videos, we only included videos that are created by official organizations (*e.g.*, TV companies, sports teams) and watched by more than tens of thousands of times. We also purchased some subscriptions (*e.g.*, ESPN+, CCTV VIP) to watch member-only videos during the collection. Three of the authors went through the videos to exclude the problematic ones (*e.g.*, with no or only a few augmentations, not ball sports, or not cover a sports event). In this process, we noticed that most of the videos focused on one sports event (*e.g.*, a goal) and thus were less than 3mins. Some videos were sports events collections and too long (*e.g.*, 45mins). Thus, we sliced these videos into pieces to ensure that each piece is shorter than 3mins and includes at least one sports event. To control the diversity of the videos, we randomly sampled a subset of videos from the raw corpus (which contains more than 1000 videos) following this priority: the number of 1) team sports *vs.* turn-taking sports, 2) different sports types, and

3) different video sources. Finally, our corpus includes 233 videos. Figure 2a and b present the number of videos of different sports types and with different time duration. Figure 3 presents some examples from our corpus.

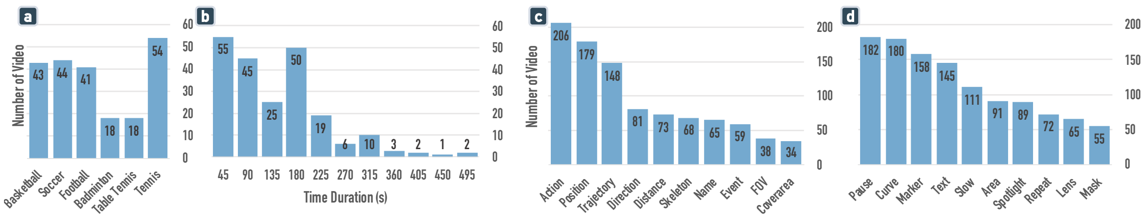

Fig. 2. The number of videos a) of different ball sports and b) with different time duration. The top-10 frequently presented c) data types and d) visual types in augmented sports videos.

Qualitative Analysis. We conducted a qualitative analysis of this corpus. Specifically, we first followed the definition in [1] to segment the augmented sports videos into clips. In our corpus, a video can have more than one sports event, each of which can be segmented into multiple clips. Three of the authors independently reviewed and segmented the videos following this process: each of them segmented $\frac{1}{3}$ videos and verified another $\frac{1}{3}$ videos segmented by others. Gathering the reviews resulted in 871 clips. We then probed into the content of 261 (30%) clips to exam the element-level design. The investigation also referred to prior research [1, 26, 37] and derived the code-sets of two element-level dimensions, Data Type and Visual Type. Next, three of the authors independently conducted open coding on the 261 clips to characterize how these visual elements are organized. Disagreements were resolved through multiple rounds of discussion and iterations with another three authors, one of whom was a senior sports expert and had more than ten years' experience in providing consulting services for the national sports team and sports TV channels. Finally, we came to two clip-level dimensions, namely, Data Level and Narrative Order. The design dimensions were further refined in the following annotation process.

Annotations. We further counted the occurrences of these design dimensions in the corpus. Five postgraduate students were recruited to annotate these videos following our codes. All of them were sports fans. We introduced

Fig. 3. Video examples in the corpus of a) table tennis in linear, b) soccer in flash forward, c) basketball in flashback, d) badminton in time fork, e) tennis in grouped, and f) football in zig-zag.

the details of the design dimensions to the five students, asked them to practice annotation on 15 curated example videos, and started the actual annotation when they were confident enough. After the annotation, cross-validation was further conducted. In total, each student annotated 45 videos and validated 45 videos from two others. Questions and discussions were encouraged throughout the whole process, which helped us further refine the design dimensions. Finally, three of the authors scanned through the annotations and calculated the statistics.

3.2 Element-level Design

The augmentations of an augmented sports videos are presented as visual elements. We characterize these visual elements with Data Type and Visual Type.

Dimension I: Data Type. Sports videos are augmented with different types of data. Perlin et al. [26] identified three types of data used in sports visualizations: tracking (in-game actions and movements), box-score (historical statistics), and meta (sports rules and player information). Given the in-game nature of sports videos, most of the augmented data on a sports video belongs to tracking data. Thus, we simplify the three types into two:

- *Tracking Data* is the data collected or extracted from a specific game, such as the moving trajectories and the actions of players. In recent years, the advances of CV techniques lead to increased tracking data that ranges from low-level physical data to high-level tactical data. This data is always associated with a specific space and time in the video and can be naturally embedded into the video. Thus, most of the data visualized in augmented videos is tracking data.
- *Non-tracking Data* refers to the data not captured from a specific game, including historical data, rules, and player information. Augmented videos usually provide this data as supplemental information to explain or comment on the situations in the game.

Figure 2c shows the top-10 frequently presented data types in our corpus. Most of (9/10) the data is tracking data except for *name*, which belongs to non-tracking data.

Dimension II: Visual Type. The data is presented as different types of visuals. Usually in data videos, data is presented as graphical marks [2], such as bar, pictograph, and map. However, in augmented sports videos, the video content in the raw footage (*e.g.*, players, the court) can also be used to encode data. We categorize these visual representations as Graphical Marks and Video Effects:

- *Graphical Marks* are the visual elements added to the raw footage. Augmented sports videos usually present data as primitive marks, such as dots, lines, and areas, while common data videos [1] contain more complex visualizations, such as donut, pie, scatter plot. We also see some dedicated marks that rarely have been used in common data videos, such as *spotlight*, *skeleton*, and *field of view*. Besides, we only found two types of animation used for the marks in augmented sports videos, namely, *Creation* and *Destruction*, while eight [2] are used in common data videos.
- *Video Effects* are visual effects based on the existing content of the raw footage, such as segmenting and moving a player, showing a slow motion, or rotating the camera. In a certain aspect, these effects are similar to the Attention Cues in [1]. However, in augmented sports videos, these effects are driven or controlled by data, with the goal of not only drawing the viewer's attention but also revealing deeper insights of a sports event (*e.g.*, moving a player to show a what-if situation or rotating the camera to present another point-of-view). These effects are limited by the quality of the raw footage (*e.g.*, frame rate and resolution) and video processing techniques.

Figure 2d shows the top-10 frequently used visual types in our corpus. 7/10 are graphical marks while only 3 belongs to video effects (*pause*, *slow*, and *repeat*).

3.3 Clip-level Design

Narrative Order → / Data Level ↓	Linear	FlashForward	FlashBack	TimeFork	ZigZag	Grouped	
Tactic	191	127	35	30	75	0	52.6%
Event	235	90	21	16	11	9	43.9%
Physics	31	0	0	0	0	0	3.6%
	52.5%	24.9%	6.4%	5.3%	9.9%	1.0%	

Fig. 4. A clip-level design space for augmented sports videos: Data Level (vertical) and Narrative Order (horizontal). The number in cells depict their combination occurrences in our corpus. Darker cells mean more occurrences. The last row and column present the ratio of each option to its dimension.

A more important question is how to organize the data and visual types in a sports video. To cope with this question, we identify two design dimensions, Data Level and Narrative Order, that should be considered for selecting and presenting data in the video.

Dimension III: Data Level. Augmented videos usually present sports data for different purposes. Specifically, some videos present data for *engagement* (*e.g.,showing the jumping height of a player who is dunking*) while others are for *education* (*e.g., highlighting the formation of the team*). We notice that the data for engagement is usually with low semantics (*e.g.*, positions and distances) that can be easily perceived from the raw footage by general audiences, while the one for education is more often with high semantics (*e.g.*, techniques and tactics) that provide extra knowledge to the audiences. Based on this observation, we categorize the data from low to high semantic levels:

Image level includes the frames of a raw footage. The data at this level has the largest quantity and will be used as the input of a system. A video without any augmentations presents the data at this level.

Physics level includes the physical data of objects detected from the Image level data, such as the postures of a player, the positions of the ball, and empty areas of the court. This level of data can be naturally understood from the video by audiences without any prior knowledge. An augmented video that only presents the data at this level is mainly **for engagement purposes**.

Event level contains the data that can only be interpreted from the video with domain specific knowledge. Typical examples include the player's techniques, the formation of the team, and the status of the ball. These kinds of data may be familiar to experienced fans, but provide extra knowledge to novices. Therefore, when presenting event data, an augmented video is considered to be in the middle **between for engagement and for education**.

Tactic level presents the reasoning results of a sports video, explaining why a team wins or loses a score. A conclusion is usually drawn by experts who analyze the data from other levels, indicating the key event and physics data that lead to the result. Hence, an augmented video with the conclusion data is mainly **for education purposes** since it provides the most additional knowledge for audiences.

Figure 4 shows the statistics of clips with different Data Level. Overall, most of the clips present Tactic level (52.6%) and Event level(43.9%) data, while only 3.6% clips present Physics level data. These statistics indicate that most of the augmented videos target at sports fans and provide expert analysis. Data Level is not meant to be a taxonomy of sports data but a design dimension that needs to be considered in creating an augmented video. For example, if a video designer wants to engage the novice audience, s/he should select and present the Physics level data.

Dimension IV: Narrative Order. A sports video usually presents the sports events in linear and rarely employs non-linear narrative structures (*e.g.*, Flashback). However, we notice that the data in these videos are not always presented in chronological order. For example, to explain the tactic of a player, it is common to pause the video and foreshadow the trajectories of the next several movements of the player. Considering the presenting order of data and its actual chronological order, we borrow the idea of *narrative order* [22] to depict how the data is presented in these augmented videos:

Linear presents data in chronological order (Figure 3a), which is the most common way to present data in sports videos. Pausing is used in Linear when there is too much data to show in one moment.

FlashForward foreshadows the data that will happen later than what is being told. FlashForward is frequently used for Tactic level data. When using FlashForward, the video is usually paused. For example, Figure 3b shows the positions of one player in the next several seconds.

FlashBack presents the data that took place earlier than what is happening. FlashBack can be used for both engagement (*e.g.*, to emphasize the achievements of a player) and education (*e.g.*, explain the causality). For example, Figure 3c highlights a player's previous position to reveal how fast a player ran.

TimeFork shows data that never happens in the game and is primary used to present a what-if analysis, explaining the results of different choices of players (Figure 3d). A typical pattern in TimeFork is to show the hypothetical data first, reject these options, and visualize the actual data at the end.

ZigZag plays the video in reverse for a short period and then forwards this part again (Figure 3f). Basic usage of ZigZag is to highlight some key events, such as crossover and spin, for both education and engagement purposes.

Grouped presents multiple sports scenes by using picture-in-picture or multiple windows (Figure 3e). These events are grouped based on some criteria (*e.g.*, correlation) and can be played in parallel or series.

Figure 4 provides the statistic of these Narrative Order in the video corpus. In general, Linear (52.5%) and FlashForward (24.9%) are the two most frequently used order, Grouped (1%) is the least frequently used one, and the remaining three—FlashBack, ZigZag, and TimeFork—all have around 5% to 10% cases.

3.4 Patterns at Clip-level Design

Based on the annotations of our corpus, we identify some common design patterns at clip-level:

+ *Data + Linear. Linear* is the most commonly used (52.5% clips) order to augment sports videos in all data levels. Presenting data in linear is natural since the underlying sports videos are played in linear. Thus, an augmented video creation tool can use the Linear order as the default choice for all data levels. An interesting finding is that Physics level is seldomly narrated in orders other than Linear. After discussing with our experts, we consider this is primary because: 1) the physical data can easily be understood by the audiences, and 2) when augmenting a video with physical data to engage the audiences, using linear without pausing can **avoid interfering the audiences' game-watching experience**.

+ *Event level + FlashForward.* We also see that many clips show Event level data using FlashForward. Through watching the videos in our corpus, we notice that using FlashForward allows to **preview and explain complex data** in the following events (*e.g.*, in the play-by-play breakdown) **without affecting the experience when actually watching it**.

+ *Physics level + FlashForward.* An interesting pattern is that compared to Physics level, Tactic level is less frequently presented in Linear but more in FlashForward. After digging into the clips in our corpus, we consider this is mainly due to another advantage of FlashForward—enhancing the connection between the current and the following events to **emphasize the causality between them**.

Tactic level + ZigZag. ZigZag can be used to **augment one event from different angles** since it can repeat a specific event multiple times. For example, a video can highlight a player in an event and then use ZigZag to replay the same event but with another player highlighted. Besides, ZigZag can be seen as **a composition of two "single" narrative orders**. For instance, some examples use Linear to show certain data, play the video, and then use ZigZag to roll back and show another data using FlashForward. Given these characteristics, ZigZag has been frequently used to present tactical data.

4 VISCOMMENTATOR

Based on the design space, we design and implement VisCommentator, a proof-of-concept that facilitates the creation of augmented sports videos for table tennis.

4.1 Design Goals

We conceive the design goals of VisCommentator based on our collaboration with domain experts, our design space, and prior research [17, 29, 42] on intelligent data design tools.

G1. Integrated with General Video Editing Tools. Augmented sports videos are produced by augmenting raw sports videos with data. Besides the data augmentations, there are many editing on the video (*e.g.*, slice a clip) can and should be finished by general video editing tools. Therefore, instead of being a separated authoring tool, the system to create augmented videos should be able to be integrated into a general video editing tool.

G2. Recommend Data Insights for Different Purposes. Many intelligent data design tools [29, 42] adopt auto-insights techniques to alleviate the data analysis burden by automatically suggesting interesting data patterns. However, as described in Sec. 3.3, augmented videos present different data for different narrative purposes (*e.g.*, engaging or education). Thus, we intend that the system should recommend data insights based on the user's narrative purposes.

G3. Recommend Visualizations for Different Narrative Orders. Visualization recommendation systems [23, 43] usually recommend visualizations based on the effectiveness of visual channels. However, in our scenarios, the narrative order that may result in different availability of visual effects. For example, when using a Linear order, we cannot move a player in the video out of his position. Consequently, the system should further consider the narrative order in the recommendation.

4.2 Usage Scenario

To illustrate how VisCommentator encompasses the design goals and the key user interfaces (UIs), we introduce the workflow taken by Jon, a hypothetical sports expert, to create an augmented video for a table tennis rally.

Brush to select. Jon is familiar with the UI of VisCommentator as it follows the design of general video editing tools (G1), which usually have a video preview (Figure 5a), a main video timeline to be edited (Figure 5b), and an edit panel in the right-hand side (Figure 5c). The loaded video file appears in the main timeline as a clip. After going through the clip, Jon brushes on the timeline to select the last three turns, which he thinks is the highlight moment of the rally.

Configure to augment. An edit panel titled "How to augment this clip?" is opened in the right (Figure 5c). In the panel, there are two button groups, "Narrative Purpose" and "Narrative Order", with *Education* and *Linear* in default, respectively. The bottom is a "Visual Mapping" list with some pre-selected data-visual mappings. By switching different "Narrative Purpose", Jon notices that the pre-selected data items in the "Visual Mapping" list will change accordingly

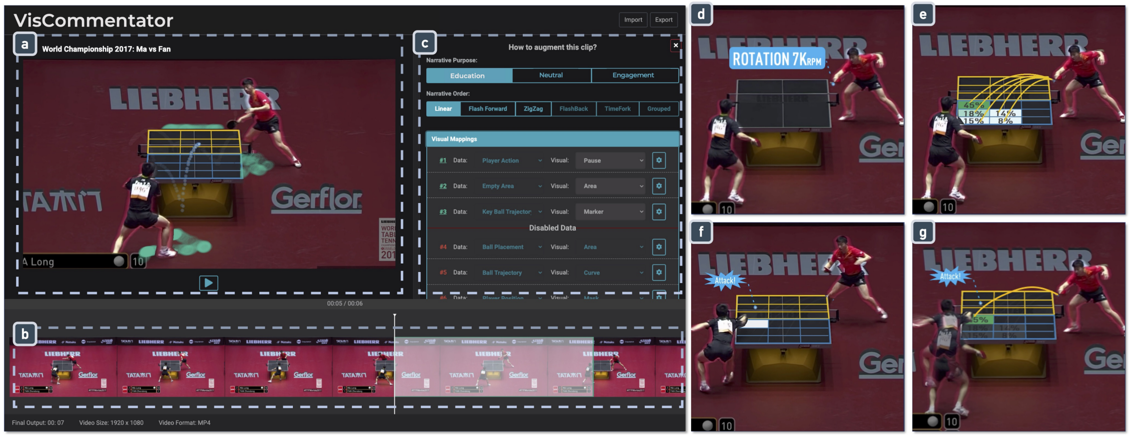

Fig. 5. a)-c) The user interface of the system, including a video preview, a main timeline, and an edit panel. d) The rotation speed of the ball is 7000 rotation per minute. e) The probability distribution of the ball placement on the table. f) The player attacks the ball to win the rally. g) The video uses a FlashForward order to preview the action of the player after showing the potential ball placement.

(G2). By hovering on the *Education* option, Jon is informed that this option will augment the video with tactical data to explain why the player wins the rally. Interested in how this augmentation looks like, Jon clicks the "Generate" button at the bottom of the edit panel. The system then generates and previews a clip augmented by visualizations in the preview window (Figure 5a, G3). The augmented clip reveals that the rotation speed of the ball is so fast (Figure 5d) that the player in red cloth can only return it to a narrow area (Figure 5e, with the highest possibility to the top-left corner of his opponent's court). This obviously gives a chance to his opponent, who performs an active attack (Figure 5f) in the next move and successfully win the game. Although impressed with these augmentations, Jon is curious about the "Narrative Order"and chooses "Flash Forward" to regenerate the augmented clip. The system presents another augmented clip that foreshadows the action of the opponent in the video right after showing the probability distributions of the ball placement (Figure 5g). Jon recognizes that this narrative method explains the causality between the ball placement and the opponent's attack in a stronger manner.

Edit and Fine-tune. However, considering that the target audiences may have little knowledge of table tennis, Jon is afraid that the tactical data is too complex for them. Consequently, Jon chooses *Engagement* for the "Narrative Purpose", which augments the clip with physical data. Instead of annotating the key players' movements, the resulting augmentation highlights the ball trace with slow motion (Figure 5a), which Jon believes will be favored by the audiences. He further selects three more physical data (*i.e.*, "Player Position", "Player Trajectory", and "Ball Placement") in the "Visual Mapping" list to augment the clip. VisCommentator automatically recommends visual effects to present these data, so that Jon only needs to perform minimal editions, *e.g.*, specify the colors. Satisfied with the results, Jon exports the augmented video.

4.3 Data Insights Recommendation

VisCommentator embeds a data insight recommendation system to automatically suggest interesting data insights based on the narrative purpose selected by the user (G2). This is achieved by two main components, namely, data pyramid and auto-insight engine.

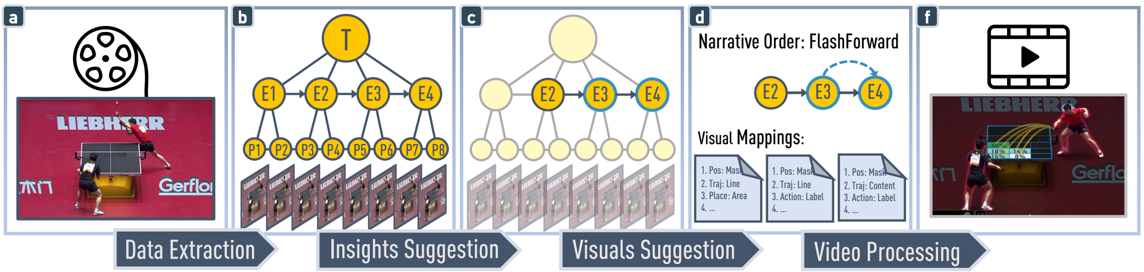

Fig. 6. The pipeline to augment a raw footage. a) The input video. b) The extracted data is organized into a data pyramid. c) After selecting a level of data and brushing a specific period, the system returns a subtree and automatically suggests interesting data insights, *e.g.*, the two events in cyan. d) The system constructs a Directed Acyclic Graph based on the user-selected narrative orders and suggests visual mappings for each data nodes. f) Rendering and generating the augmented video.

Data Organization—Data Pyramid. Data Pyramid is a tree-based structure (Figure 6b) that contains four levels, each corresponding to a data level. Each node in Data Pyramid stores its own data, a pointer to its children, and a pointer to its successor. By this, we can separate the data at different levels while maintaining their chronological orders and relationships (*e.g.*, a ball event *net hit* and its physical data *positions*). A data pyramid can be constructed based on data manually prepared by experts or automatically extracted by CV systems. In this work, we implement a bottom-up approach to extract the data from 50FPS videos of a table tennis rally. Specifically, for 1) the Image level, we slice the raw video into frames; 2) the Physics level, we employ **two CV models** [36, 40] on each frame to segment the objects (*e.g.*, players, ball, and table) and detect their 3D positions, and further derive other physical data such as trajectories, speeds, and empty areas; 3) the Event level, we partition the frames into groups based on time intervals (*i.e.*, 1s in our implementation) and put each group, together with the corresponding physical data, into **another two CV models** [6, 27] to detect sports events, such as player actions, techniques, and ball status; and finally 4) the Tactic level, we use **a pre-trained causal graph** [25] to detect the key events and their key physical attributes that account for the game result (*i.e.*, why one player win). Further technical details can be found in the supplemental materials [1].

Insights Suggestion—Auto-insight Engine. Auto-insight techniques have been increasingly used in visualization design tools [17, 29, 42] to ease the data analysis process. Whenever a user brushes the timeline, our system queries a subtree within the time interval and then detects the insights at a specific data level (Figure 6c) based on the user-selected narrative purpose (*i.e.*, Tactic level for education, Event level for neutral, and Physics level for engagement).

Typical auto-insight systems accept tabular data as the input, enumerate data subspaces (*i.e.*, subtables) as the context, use a specific data attribute in a subspace to divide data items into groups (*i.e.*, as the x-axis), calculate data insights based on a list of predefined types (*e.g.*, extreme, rank, trend, association), and sort and return data insight based on their statistical significance. However, these auto-insight techniques cannot be directly applied to our scenario. We identify two bottlenecks and propose the following adaptations:

- *Consecutive Interval Enumeration*—Typical methods generate subspaces by enumerating the combinations of rows and columns, while in our case, subspaces enumerated from non-consecutive periods of time are meaningless. Thus, we only enumerate consecutive intervals as subspaces.
- *Spatial-temporal Breakdowns*—Usually, a breakdown attribute can be temporal, spatial, or categorical. In our scenarios, however, groups divided by a categorical breakdown cannot be embedded into the video since the groups don't have

[1] https://viscommentator.github.io

spatial or temporal properties that can be mapped to the video. Consequently, we only allow spatial or temporal data attributes to be breakdowns.

Besides these two adaptations, we also revise some significant metrics, *e.g.*, consider the time duration to eliminate accumulated significance. We refer the readers to [29, 42] for other technical details.

4.4 Visualizations Recommendation

Besides data insights, VisCommentator further integrates a visualizations recommendation engine that plans the appearance orders of the augmentations and maps data to different visual effects (G3).

Narrative Order—Directed Acyclic Graph. After obtaining the data insights (Figure 6c), the next task is to define their appearance order in relation to the raw video. In VisCommentator, we model the appearance order of different data by using a Directed Acyclic Graph (DAG). The DAG is constructed based on the Narrative Order selected by the user. For example, in Figure 6c, if the user selects a FlashForward, our system will automatically add a virtual link between node $E3$ and $E4$ (the dashed link in Figure 6d). Then when executing the animation, our system will perform a topological sort to ensure all the prepended nodes of $E3$ are presented before animating to $E4$. Besides, $E4$ will be flagged to avoid repeated visualizations. Current system only implements the common patterns mentioned in Sec. 3.4.

Obviously, not in all cases, a virtual link should be added between two nodes. The addition of links between two nodes depends on their data attributes to show. In our implementation, we use the annotations of our video corpus as prior knowledge to decide whether a link should be added or not. Specifically, before adding virtual links between two nodes (*e.g.*, use FlashForward to show a player's position in $E4$ after showing the placement of a ball in $E3$), we will search the corpus to ensure such pattern exists; otherwise, the link will not be added.

Visual Effects—Maximum Conditional Probability. The next step is to map the data into visuals. There are many techniques for mapping data to visuals, such as rule-based [43], constraint-based [23], decision tree [42], and neural network [14]. Considering our specific domain, we intend to use the collected videos as prior knowledge to accomplish the visual mappings. Specifically, we model the visual mappings using conditional probability distribution: $p = f((d, v) \mid O)$, where d, v, and O is the data, visual, and narrative order, respectively. Intuitively, this model represents the probability distribution of data-visual mappings under different narrative orders. This probability distribution can be estimated based on the occurrence frequency of the combinations of data, visual, and narrative orders in our corpus. Consequently, given d and O, our system will search the v that maximizes p and suggest it to the user. We also use a rule-based method [43] as the default to select effective visual channels to visualize data if there are no mapping records of the data under a narrative order. Our method is simple but effective, since it is built upon the videos collected from reputable sources, and can be extended and optimized in future research.

5 IMPLEMENTATION

VisCommentator is implemented in a browser/server architecture. The browser part, which is built upon HTML + CSS + JavaScript, is responsible for the UI and rendering of the video. We mainly use the HTML Canvas to achieve the rendering of augmented videos. To improve the web-page efficiency, we adopt the OffscreenCanvas [24] technique that leverages the *worker*, a multi-thread like technique in a modern browser, to accelerate the heavy rendering. The server part is implemented based on Node.js + TypeScript. To extract the data from a video, we employ TensorFlow.js [35] that supports running pre-trained deep learning models by using Node.js. The source code will be open-source once the paper is accepted.

6 USER STUDY

We conducted a user study to assess the usability and utility of VisCommentator. The study aimed to evaluate whether sports experts, the target users, can create augmented table tennis videos with our system and observe their creation process to reflect on future improvements. We further invited 23 table tennis fans to qualitatively assess the augmented videos created by the experts.

Participants: We recruited 7 table tennis experts (E1-E7; 3 male; age: 20-30, one didn't disclose) from a university sports science department. All experts were majored in Sports Training with at least 10 years professional experiences in table tennis. All the experts had experiences of using lightweight video editors, *e.g.*, VUE [41] but no experiences on advanced video editing tools, *e.g.*, Adobe Premiere, Adobe After Effects, and Camtasia. The scale of expert participants is consistent with similar sports visualization research [32, 44]. Each participant received a gift card worth $14 at the beginning of the session, independent of their performance.

Task and Materials: The task required the participants to augment two table tennis videos with data insights they discovered. We prepared two table tennis videos, namely, , since they were selected as top 10 rallies in 2019 by ITTF [15, 16]. A digital sheet of the design space and statistics of our corpus and another sheet of all the data extracted from the two videos were provided. We also provided the participants with sketching materials, including blank papers, printed frames of the videos, color markers, and pens.

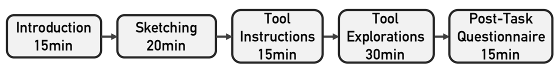

Fig. 7. The overview of the study procedure.

Procedure: The procedure is summarized in Figure 7. The study began with the introduction (20min) of the study purpose, the concept and examples of augmented sports videos, and the design space with 15 curated example videos. We also played an official tutorial video of Vizrt Libero [39], a state-of-the-art commercial tool, to introduce the typical creation process of augmented sports videos. We moved to the sketching phase (20min) when the experts had no more questions of the introduction. The sketching phase aimed to contextualize the task and helped the experts familiar with the raw video. We provided the experts with all the prepared materials and asked them to describe their interpretations of the game and sketch the augmentations on the videos. The experts were encouraged to augment the videos for different target users, including proficient players, expert designers, and sports enthusiasts, and instructed to assume that all the data was available. Questions and discussions were encouraged throughout the whole process.

When the experts were confirmed finishing their sketching, we moved to the next phase and introduced VisCommentator (15min) with a step-by-step example. Experts were then encouraged to explore the system (30min) and create augmentations. For each raw video, we asked the experts to submit one augmented result. The study was concluded by a semi-structured interview with a 7-point Likert scale questionnaire (15min). Each session was run in the lab, using a 24-inch monitor, with a think-aloud protocol, and lasted approximately 1.5 hours.

6.1 Results

Usability: Overall, our system was rated as *easy to learn* ($\mu = 6.00, 95\%CI = [5.04, 6.96]$) and *easy to use* ($\mu = 6.86, 95\%CI = [6.58, 7.14]$) by the experts (Figure 8). As commented by E3, *"the system doesn't have too much complex UI"* and *"I can quickly generate augmentations by clicking several buttons."* Although we had never mentioned the concept of templates, E4 pointed out that *"easy, your system applies templates to the video..."*. When we asked which part is the most

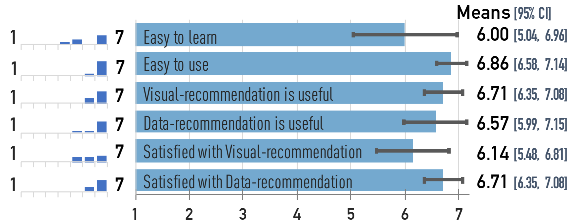

Fig. 8. Experts' feedback on the usability, showing means and 95%CIs. Distributions are shown on the left.

difficult to learn, four of the experts thought was to *"select proper visual effects"*, two (E4 and E5) said the concept of narrative order, and one (E1) believed nothing was too difficult.

Usefulness: The data-driven features—data insights and visuals recommendations—of our system were particularly lauded by the experts and confirmed to be useful ($\mu = 6.57, 95\%CI = [5.99, 7.15]$ and $\mu = 6.71, 95\%CI = [6.35, 7.08]$, respectively). All experts thought that the data recommendation based on narrative purposes was useful, which, commented by E2, was *"just like a meal set"*. E3 detailed that *"it saves your time from selecting data to be presented."* The experts found the visualization recommendation was particularly useful as they didn't want to *"get into the details of visual mappings"*. Besides, E4 responded that *"the recommendation will be better if it can show me the comparison of different narrative orders."*

Satisfaction: The rating also reflected a high user satisfaction for the data ($\mu = 6.71, 95\%CI = [6.35, 7.08]$) and visualization recommendations ($\mu = 6.14, 95\%CI = [5.48, 6.81]$). The experts said the visualization recommendation *"really helps guys like us (sports experts)..."* Comments also implied further improvements of the data insights recommendation: *"The highlighted actions in the last three turns are very reasonable. But the first three turns are less..."* (E2, E3). We explained that this was mainly due to the data extraction models, which can be improved once more data was available.

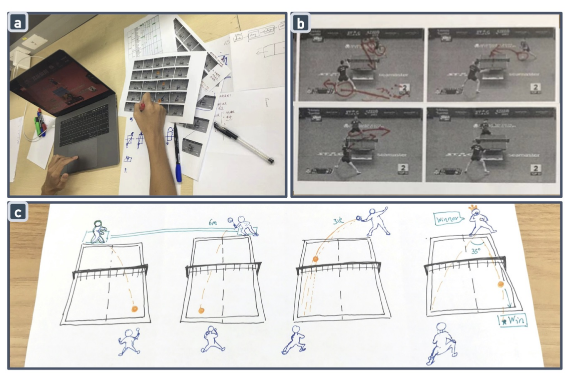

Fig. 9. a) A photo of an expert in the sketching phase. b) Annotations added by an expert on the printed frames of a video. c) A comic strip drawn by an expert, explaining why a player wins the game.

Created Examples. During the sketching phase, we observed that most of the experts were only able to draw simple lines (Figure 9b) to express their ideas. Some experts were even reluctant to sketch on the paper, except E4, who used a comic strip (Figure 9c) to introduce his idea. In the system exploration phase, the experts generated 53 augmented clips in total based on the server logs. These clips covered all the data levels and narrative orders supported by our system, implying that the experts had fully explored the system. As for the 14 final submitted clips, most of them focused (11) on the tactical data of the last 2 or 3 turns of the games since "*usually the last three turns are the most important*" (E1-7). Figure 10 presents two examples from the experts.

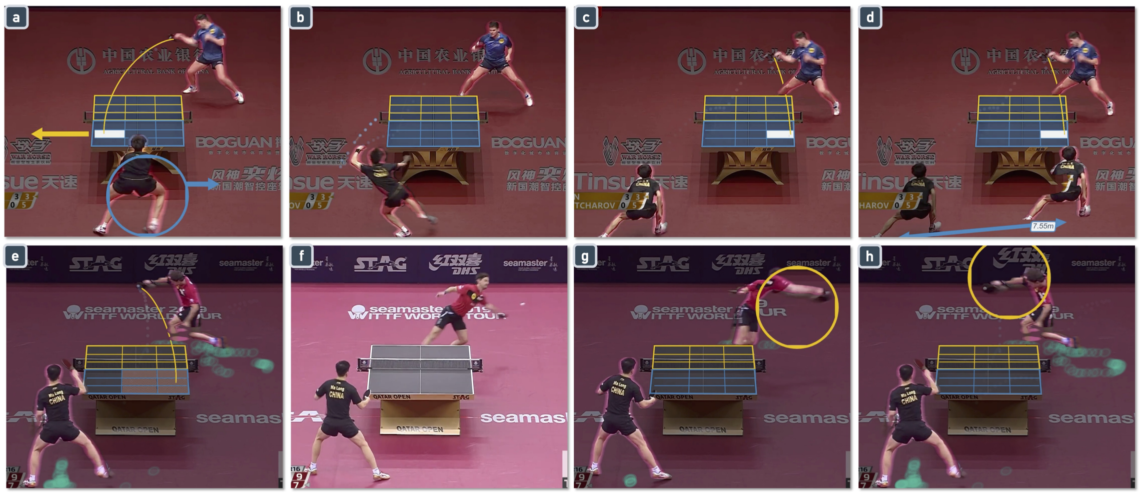

Fig. 10. a)-d) An FlashForward augmented video example created by an expert: a) The augmentation shows that when the player in black plans to move to the right, his opponent attempts to stroke the ball to the left. b) The player in black tries his best to return the ball. c) After he return the ball, his opponent strokes the ball to the right, d) which is too far away for him to return the ball. Thus, he loses this rally. e)-h) An ZigZag example created by another experts: e) The video first shows the Tactical level data that reveals the player in red strokes the ball to an empty area that his opponent cannot cover. f) Then the video rollbacks to the previous turn and replays again. g) In this time, the video highlights that the player in red first uses his left hand. h) However, in the next movement, he switches to use his right hand to stroke the ball, and thus successfully hits the ball to the empty area.

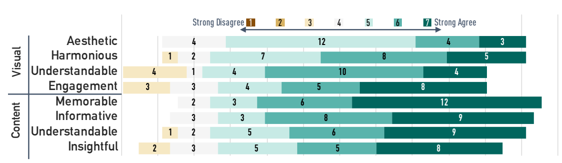

Fig. 11. Overall ratings of the two random selected augmented videos created by the experts, indicating the positive feedback from 23 sports fans.

Audiences feedback. We further conducted an online survey to qualitatively assessed the augmented clips created by the experts from the audience's angle. Given some augmented clips have the similar content, we first deduplicated them and then randomly selected one clip for each video. The two clips, together with their original version and a post-watch survey (drawn from [42]), were posted to a digital forum of a university. We only invited table tennis fans who we considered as the target users of the augmented videos. Figure 11 shows the results collected from 23 participants with a very diverse background (13 majors). Overall, the participants provided positive feedback with respect to both the visual effects data content. Given the augmented videos were created by our proof-of-concept system without further polishing, we consider these feedbacks to indicate the promising of our system.

6.2 Observations and Feedback.

Now we reflect on the interesting observations and feedback, which we believe imply future opportunities.

Design Space for atypical designers. We consider the sports expert in our scenario is a kind of typical designers: on the one hand, the sports experts are not novices, and the content of the augmented videos is usually decided by them; on the other hand, the sports experts are not designers, and lack sufficient skills and knowledge to create a video. Thus, we are interested in the sports experts' feedback on the design space. In the interview, all experts confirmed the usefulness of the design space, especially the Data Level. *"This helps us know our audience...,"* commented by E4, who used to explain table tennis matches with too many tactical details to his friends. As for the Narrative Order, some experts thought that it was *"too fancy"* and believed only using *linear* and *flashforward* were enough. But they also added that *"maybe audiences will like it..."* These responses were consistent with our expectations that the sports experts focus more on the data aspect instead of the visual one. We considered this feedback also indicates that the design of authoring systems for sports experts should provide more automation for the visual part than the data part.

Data analysis—Translate high level findings into specific data. One unique characteristic of augmented sports videos is that a sports video is both the source of data and the canvas for showing the data, which makes the data analysis and communication inseparable in creating augmented sports videos. We did observe that experts paused and analyzed the key period of the videos, took notes of the players and ball, and then augmented the video based on their notes. However, we noticed that the experts' analytic findings were usually very abstract and even difficult to find data equivalent, *e.g.*, *"I can see something in his eyes, he is going to lose"* (E5) and *"his pace is messed up"* (E2) Given the highly abstract nature of these findings, the experts usually needed to translate them into specific data and then further visualize them; or sometimes they just gave up. In such a case, our data insights recommendation features can inspire the experts and helps them discover other interesting patterns. A promising future direction is to explore the way to **translate the experts' high level findings into specific data**. To achieve this, a possible solution is to utilize natural language processing techniques, which have already been employed in some creativity support tools [18, 19].

Data visualization—Select the data, instead of visualizations. Sports experts are more familiar with data instead of visualizations. In the study, we observed that the experts tended to design the augmented videos from a data perspective and considered less on the visuals. For example, the experts frequently said that *"show the ball trajectory/ball placements/player position"*. However, when we asked them *"by using what?"*, most of them responded that *"just use the positions"*, *"not important"* or *"I don't know."* In the interview, we proposed two system designs that allow the experts to 1) *"select a graphical mark to annotate a player's position"* or 2) *"select a player's position attribute and highlight it with system suggested marks"*. All the experts preferred the last one. This also reflected the usefulness of our visualizations recommendation for the experts. However, most commercial tools adopt the first method, which ask the user to **select a mark** to annotate the data rather than **select a data** to be highlighted. Object-oriented UI design [8, 45], which allows the experts to create visual mappings through direct manipulations, can be a potential solution to address this issue.

7 DISCUSSION

Generalization to other ball sports. Our prototype system only supports the augmentation of table tennis videos. However, it can be easily extended to other turn-taking ball sports (*e.g.*, tennis and badminton) once data is available. As for other team ball-sports, they involve more parallel events and thus require a redesign of the data pyramid, which stores the data in a linear manner. Nonetheless, the design space can still provide guidance for designing authoring tools to create augmented sports videos for other ball sports.

From augmenting videos to augmenting reality. As delineated by Stein et al. [33], the study of augmenting videos with visualizations is also related to Immersive Visualizations, especially SportsXR [20], which focuses on augmenting real-world sports activities with data visualizations to support in-situ decision making and engage sports enthusiasts. However, although AR opens up new opportunities for both sports analytics and sports watching experience, it also introduces many challenges that yet to be tackled, such as scene understanding, streaming decision making, and data visualization in 3D real-world canvas. Our research adds to this direction by exploring the ways to present data in videos of the real world.

Communicating data through videos. After decades of research efforts, many systems and methodologies have been proposed to lower the barrier to data visualizations. Theses research systems and methods have been shifted and integrated into commercial softwares, such as Power BI [21] and Tabular [34], to allow data analysts and even general users to communicate data insights. However, communicating data through videos is still a challenging task [37] and has not been fully explored. On the other hand, recent years have witnessed an emerging trend to disseminate information through videos (*e.g.*, YouTube, TikTok). By investigating augmenting sports videos with data, we expect our work can arouse interest and inspire future research in this promising direction.

Limitations. The sample size of our user study is small since the access to sports experts is naturally limited. Nonetheless, the scale of expert participants in the study is consistent with similar sports visualization research [32]. The design space is derived from a corpus of a limited number of videos. A larger-scale study that involves more videos and collective knowledge (*e.g.*, crowdsourcing) is thus suggested. The experts also identified limitations of our system. Most of the limitations were related to system maturity, as our system is not meant to be a fully functional video editor. For example, the audio cannot be edited in our system; some parameters (*e.g.*, playback rate) were fixed to default values. Some limitations indicate promising improvements. For example, the experts suggested that not only the data levels but also the visual effects should also be recommended based on narrative purposes, as *"athletes don't need such fancy visualizations."* (E2)

8 CONCLUSION

This work is motivated by the close collaboration with a group of sports experts who have a strong demand to augment sports videos with data. To ease the creation of augmented sports videos, we first systematically review 233 augmented sports videos collected from reputable sources and derive a design space that characterizes augmented sports videos at element- (*what are the constituents*) and clip-levels (*how the constituents are organized*). Informed by the design space, we design and implement VisCommentator, a prototype system that allows sports experts to augment table tennis videos efficiently. VisCommentator embeds two recommendation engines to suggest data insights and visual effects based on the user-selected narrative purpose (*e.g.*, for education or engagement) and narrative order. A user study with seven sports experts confirmed the usefulness and effectiveness of the system. The study's created videos were found to be informative and engaging by a group of 23 sports fans. We have also discussed and shared the observations and feedback from the study, which suggest future research.

REFERENCES

[1] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. 2015. Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens. In *Proc. of CHI*. ACM, 1459–1468. https://doi.org/10.1145/2702123.2702431

[2] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2017. Authoring Data-Driven Videos with DataClips. *IEEE Trans. Vis. Comput. Graph.* 23, 1 (2017), 501–510. https://doi.org/10.1109/TVCG.2016.2598647

[3] BBC. 2020. Piero. https://www.bbc.co.uk/rd/projects/piero.

[4] Matthew Brehmer, Bongshin Lee, Benjamin Bach, Nathalie Henry Riche, and Tamara Munzner. 2017. Timelines Revisited: A Design Space and Considerations for Expressive Storytelling. *IEEE Trans. Vis. Comput. Graph.* 23, 9 (2017), 2151–2164. https://doi.org/10.1109/TVCG.2016.2614803

[5] Ruochen Cao, Subrata Dey, Andrew Cunningham, James A. Walsh, Ross T. Smith, Joanne E. Zucco, and Bruce H. Thomas. 2020. Examining the use of narrative constructs in data videos. *Vis. Informatics* 4, 1 (2020), 8–22. https://doi.org/10.1016/j.visinf.2019.12.002

[6] CatBoost. 2020. CatBoost. https://catboost.ai.

[7] Celtics. 2020. GE the Breakdown. https://www.nba.com/celtics/video/originals/the-breakdown-vs-lal-030719.

[8] Zhutian Chen, Yijia Su, Yifang Wang, Qianwen Wang, Huamin Qu, and Yingcai Wu. 2020. MARVisT: Authoring Glyph-Based Visualization in Mobile Augmented Reality. *IEEE Trans. Vis. Comput. Graph.* 26, 8 (2020), 2645–2658. https://doi.org/10.1109/TVCG.2019.2892415

[9] Zhutian Chen, Yun Wang, Qianwen Wang, Yong Wang, and Huamin Qu. 2020. Towards Automated Infographic Design: Deep Learning-based Auto-Extraction of Extensible Timeline. *IEEE Trans. Vis. Comput. Graph.* 26, 1 (2020), 917–926. https://doi.org/10.1109/TVCG.2019.2934810

[10] Clippers. 2020. Court Vision. https://www.clipperscourtvision.com/.

[11] Neil Cohn. 2013. Visual Narrative Structure. *Cogn. Sci.* 37, 3 (2013), 413–452. https://doi.org/10.1111/cogs.12016

[12] ESPN. 2020. DETAIL. https://www.espn.com/watch/catalog/f48c68af-f980-4fcb-8b59-2a0db01f50cf/_/country/us.

[13] Maximilian T Fischer, Daniel A Keim, and Manuel Stein. 2019. Video-based Analysis of Soccer Matches. In *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*. 1–9.

[14] Kevin Zeng Hu, Snehalkumar (Neil) S. Gaikwad, Madelon Hulsebos, Michiel A. Bakker, Emanuel Zgraggen, César A. Hidalgo, Tim Kraska, Guoliang Li, Arvind Satyanarayan, and Çagatay Demiralp. 2019. VizNet: Towards A Large-Scale Visualization Learning and Benchmarking Repository. In *Proc. of CHI*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 662. https://doi.org/10.1145/3290605.3300892

[15] ITTF. 2020. Top10 Rally in 2019: Lin vs Ovtcharov. https://www.youtube.com/watch?v=Id2I02ofKd4.

[16] ITTF. 2020. Top10 Rally in 2019: Ma vs Boll. https://www.youtube.com/watch?v=0ff3dAt41pU.

[17] Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J. McGuffin, and Hanspeter Pfister. 2019. DataToon: Drawing Dynamic Network Comics With Pen + Touch Interaction. In *Proc. of CHI*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 105. https://doi.org/10.1145/3290605.3300335

[18] Yea-Seul Kim, Mira Dontcheva, Eytan Adar, and Jessica Hullman. 2019. Vocal Shortcuts for Creative Experts. In *Proc. of CHI*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 332. https://doi.org/10.1145/3290605.3300562

[19] Gierad Laput, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar. 2013. PixelTone: a multimodal interface for image editing. In *Proc. of CHI*, Wendy E. Mackay, Stephen A. Brewster, and Susanne Bødker (Eds.). ACM, 2185–2194. https://doi.org/10.1145/2470654.2481301

[20] Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. 2020. SportsXR–Immersive Analytics in Sports. *arXiv preprint arXiv:2004.08010* (2020).

[21] Microsoft. 2020. PowerBI. https://powerbi.microsoft.com/en-us/.

[22] Nick Montfort. 2007. Ordering Events in Interactive Fiction Narratives. In *Intelligent Narrative Technologies, AAAI Fall Symposium (AAAI Technical Report, Vol. FS-07-05)*, Brian S. Magerko and Mark O. Riedl (Eds.). AAAI Press, 87–94. https://www.aaai.org/Library/Symposia/Fall/2007/fs07-05-016.php

[23] Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. 2019. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Trans. Vis. Comput. Graph.* 25, 1 (2019), 438–448. https://doi.org/10.1109/TVCG.2018.2865240

[24] Mozilla. 2020. OffscreenCanvas. https://developer.mozilla.org/en-US/docs/Web/API/OffscreenCanvas.

[25] Judea Pearl. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, USA.

[26] Charles Perin, Romain Vuillemot, Charles D Stolper, John T Stasko, Jo Wood, and Sheelagh Carpendale. 2018. State of the Art of Sports Data Visualization. In *Comput. Graph. Forum*, Vol. 37. Wiley Online Library, 663–686.

[27] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Network. In *Proc. NIPS.* 91–99.

[28] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Trans. Vis. Comput. Graph.* 16, 6 (2010), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

[29] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. 2021. Calliope: Automatic Visual Data Story Generation from a Spreadsheet. *IEEE Trans. Vis. Comput. Graph.* 1, 1 (2021), 1–1.

[30] ShuttleFlash. 2020. ShuttleFlash. https://www.youtube.com/channel/UCnGHt9C1lHpCcSxwl6GUF0Q.

[31] Second Spectrum. 2020. Second Spectrum. http://secondspectrum.com/.

[32] Manuel Stein, Thorsten Breitkreutz, Johannes Häussler, Daniel Seebacher, Christoph Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel A. Keim, and Halldor Janetzko. 2018. Revealing the Invisible: Visual Analytics and Explanatory Storytelling for Advanced Team Sport Analysis. In *Proc. of BDVA*. IEEE, 1–9. https://doi.org/10.1109/BDVA.2018.8534022

[33] Manuel Stein, Halldor Janetzko, Andreas Lamprecht, Thorsten Breitkreutz, Philipp Zimmermann, Bastian Goldlücke, Tobias Schreck, Gennady Andrienko, Michael Grossniklaus, and Daniel A Keim. 2017. Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Trans. Vis. Comput. Graph.* 24, 1 (2017), 13–22.

[34] Tableau. 2020. Tableau. https://www.tableau.com/.

[35] TensorFlow. 2020. TensorFlow.Js. https://www.tensorflow.org/js.

[36] Tensorflow. 2020. TFjs-BodyPix2. https://github.com/tensorflow/tfjs-models/tree/master/body-pix.

[37] John Thompson, Zhicheng Liu, Wilmot Li, and John T. Stasko. 2020. Understanding the Design Space and Authoring Paradigms for Animated Data Graphics. *Comput. Graph. Forum* 39, 3 (2020), 207–218. https://doi.org/10.1111/cgf.13974

[38] Vizrt. 2020. Viz Libero. https://www.vizrt.com/products/viz-libero.

[39] Vizrt. 2020. Vizrt Libero Tutorial. https://www.vizrt.com/vizrtv/on-demand/getting-started-viz-libero.

[40] Roman Voeikov, Nikolay Falaleev, and Ruslan Baikulov. 2020. TTNet: Real-Time Temporal and Spatial video Analysis of Table Tennis. In *Proc. of CVPR*. IEEE, 3866–3874. https://doi.org/10.1109/CVPRW50498.2020.00450

[41] VUE. 2020. VUE. https://vuevideo.net.

[42] Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, Xiaojuan Ma, and Dongmei Zhang. 2020. DataShot: Automatic Generation of Fact Sheets from Tabular Data. *IEEE Trans. Vis. Comput. Graph.* 26, 1 (2020), 895–905. https://doi.org/10.1109/TVCG.2019.2934398

[43] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock D. Mackinlay, Bill Howe, and Jeffrey Heer. 2016. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Trans. Vis. Comput. Graph.* 22, 1 (2016), 649–658. https://doi.org/10.1109/TVCG.2015.2467191

[44] Yingcai Wu, Ji Lan, Xinhuan Shu, Chenyang Ji, Kejian Zhao, Jiachen Wang, and Hui Zhang. 2017. ittvis: Interactive visualization of table tennis data. *IEEE Trans. Vis. Comput. Graph.* 24, 1 (2017), 709–718.

[45] Haijun Xia, Nathalie Henry Riche, Fanny Chevalier, Bruno Rodrigues De Araújo, and Daniel Wigdor. 2018. DataInk: Direct and Creative Data-Oriented Drawing. In *Proc. of CHI*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, 223. https://doi.org/10.1145/3173574.3173797