SmartShots: Enabling Automatic Generation of Videos with Data Visualizations Embedded

Tan Tang*
Junxiu Tang*
tangtan@zju.edu.cn
tangjunxiu@zju.edu.cn
Zhejiang Lab & Zhejiang University

Jiewen Lai
University of California, Berkeley
jw.lai@berkeley.edu

Lu Ying
Georgia Institute of Technology
lying8@gatech.edu

Peiran Ren
Alibaba Group
renpeiran@gmail.com

Lingyun Yu
Xi'an Jiaotong-Liverpool University
Lingyun.Yu@xjtlu.edu.cn

Yingcai Wu†
Zhejiang Lab & Zhejiang University
ycwu@zju.edu.cn

ABSTRACT

Videos become prevalent for storytellers to inspire viewers' interests. To further enhance narrations, visualizations are integrated into videos to present data-driven insights. However, manually crafting such data-driven videos is difficult and time-consuming. Thus, we present SmartShots, a system that facilitates the automatic integration of in-video visualizations. Specifically, we propose a computational framework that integrates non-verbal video clips, images, a melody, and a data table to create a video with data visualizations embedded. The system automatically translates the multi-media material into shots and then combines the shots into a compelling video. In addition, we develop a set of post-editing interactions to incorporate users' design knowledge and help them re-edit the automatically-generated videos.

CCS CONCEPTS

• **Human-centered computing** → *Visualization systems and tools*;
• **Information systems** → **Multimedia information systems**.

KEYWORDS

visualization, data-driven videos, storytelling

ACM Reference Format:
Tan Tang, Junxiu Tang, Jiewen Lai, Lu Ying, Peiran Ren, Lingyun Yu, and Yingcai Wu. 2020. SmartShots: Enabling Automatic Generation of Videos with Data Visualizations Embedded. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3394171.3414356

*Both authors contributed equally to this research.
†Corresponding author

MM '20, October 12–16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-7988-5/20/10.
https://doi.org/10.1145/3394171.3414356

1 INTRODUCTION

Videos are effective means for attracting audiences due to the engaging and narrative features, which are adopted in various disciplines, such as journalism [9] and education [5]. To achieve a highly engaging narration, *visualization*, as an effective method for communicating insights [8] and telling stories [6, 7], has been employed to enrich videos with data. However, creating videos with data visualizations embedded (**VDE** for short) is not an easy task, which requires considerable efforts to balance embedded data visualizations and video contents. To ease the difficulty of producing VDE, we plan to develop a novel system that facilitates the automatic integration of data visualizations.

To develop such a system, we investigate recent studies [3, 10] that are related to the design of VDE. Amini et al. [2] first introduced data videos and then developed DataClips [3] that can help users combine data-driven motion graphics into a complete data video. However, DataClips is not applicable for producing VDE because it focuses on data visualizations while ignoring both video and audio content. To produce VDE, a designer must consider various design factors like aesthetic layouts, appropriate duration, and effective visualizations. To explore such a complex design space, Tang et al. [10] proposed a series of design guidelines on embedding animated data visualizations into short-form videos. However, how to develop a system that can facilitate the automatic generation of VDE remains an open issue.

In this work, we propose SmartShots, a novel video generation system for producing VDE from automatically assembling multimedia content. SmartShots translates the input data tables, video clips, and images into shots that refer to video pieces and then matches the shots with the input non-verbal music to automate the composition of source materials. Specifically, we determine the duration of video shots to ensure that shot transitions can match music rhythm, which is formulated as a dynamic programming problem. To support the automatic embedding of data visualizations, we employ a novel optimization model that targets optimizing embedded layouts according to the five aesthetic principles, namely, balance, alignment, proximity, readability, and coherence. In addition, we develop a set of post-editing interactions for users to re-design the automatically-generated videos.

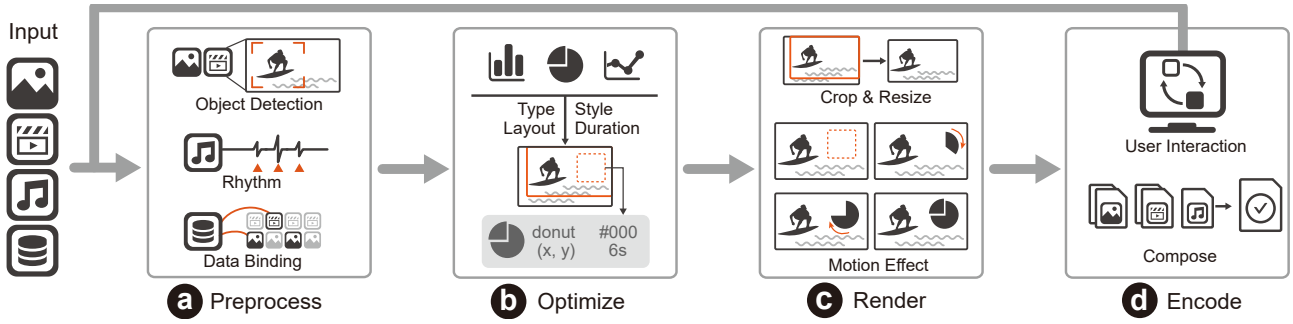

Figure 1: SmartShots pipeline. User interactions are incorporated to make iterative editings of automatically-generated videos.

2 SMARTSHOTS

This section presents SmartShots, an automatic system that facilitates the easy creation of VDE. We implement SmartShots using a client-server architecture which comprises a backend that runs a computational model to generate videos and a web interface that allows users to edit the output videos.

2.1 System Pipeline

We first introduce the video generation pipeline of SmartShots. The inputs of our system include data tables, video clips, images, and non-verbal music, which are processed in the following stages (see Fig. 1): *1) Preprocess* (Fig. 1a): Users associate data tables with the input images or video clips, after that we translate them into data shots that refer to video pieces with data visualizations embedded. Two intelligent modules are employed here to extract the music rhythm and obtain the semantic information of the data shots. *2) Optimize* (Fig. 1b): We develop an optimization module that recommends suitable visualization types, produces appealing layouts, obtains harmonic styles, and determines suitable duration for the data shots. *3) Render* (Fig. 1c): We employ a crop-and-resize scheme to normalize the sizes of data shots and then apply motion effects and render visualizations to these shots. *4) Encode* (Fig. 1d): We output VDE by assembling data shots, composing music, and encoding the video stream according to the users' interactions.

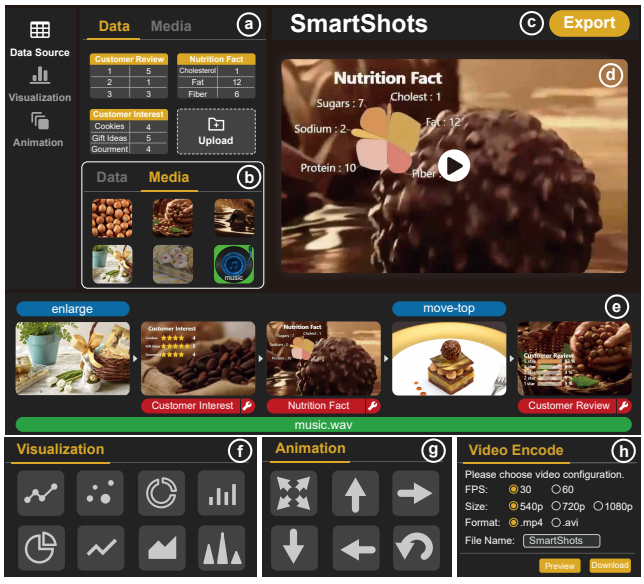

Figure 2: SmartShots interface. Pop-up panels (f-h) for visualization choosing, animation setting and video exporting.

2.2 User Interface

The web interface is built in *TypeScript* [4] and *React* [1]. The *material* interface enables users to upload data tables (Fig. 2a) and other multi-media materials (Fig. 2b). It comprises *visualization* (Fig. 2f) and *animation* (Fig. 2g) panels for users to configure output videos. To support the easy construction of stories, SmartShots enables users to *drag and drop* any entity from the *material* interface to the *timeline* (Fig. 2e). For example, users can pick an alternative visualization, with which they can replace the automatically-chosen visualization according to their preference. SmartShots employs different labels to distinguish the selected entities. The shots surrounded by colorful labels are sequentially presented in the middle of the *timeline* (Fig. 2e): the blue labels indicate motion types, the red labels indicate data tables and their associated visualizations, and the green label indicates the input music. SmartShots also enables users to preview (Fig. 2d) and export videos (Fig. 2c and Fig. 2h).

3 CONCLUSION

In this work, we develop SmartShots that facilitates the automatic generation of videos with data visualizations embedded (**VDE**). VDE extends the form of data videos from motion graphics to videos that combine real scenes and embedded data visualizations. SmartShots can make this new form of videos more accessible and serve numerous amateur users, which effectively reflects the concept of *visualization for mass*. Our work takes the initial step toward an easy creation of VDE, which is full of opportunities with the explosive development of video marketing.

ACKNOWLEDGMENTS

The work was supported by NSFC (61761136020), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1609217), Zhejiang Provincial Natural Science Foundation (LR18F020001). This project is also partially funded by Microsoft Research Asia. Lingyun Yu is supported by XJTLU Research Development Funding RDF-19-02-11.

REFERENCES

[1] 2019. React. Retrieved March 31st, 2019 from https://reactjs.org/.

[2] 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 *Proceedings of ACM Conference on Human Factors in Computing Systems*. 1459–1468.

[3] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2017. Authoring Data-Driven Videos with DataClips. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 501–510.

[4] Gavin Bierman, Martín Abadi, and Mads Torgersen. 2014. Understanding typescript. In *European Conference on Object-Oriented Programming*. 257–281.

[5] Jere Brophy. 2003. *Using Video in Teacher Education.* Emerald Group Publishing Limited.

[6] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The Next Step for Visualization. *IEEE Computer* 46, 5 (2013), 44–50.

[7] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5 (2015), 84–90.

[8] Tamara Munzner. 2014. *Visualization Analysis and Design.* CRC Press.

[9] John Pavlik. 2000. The Impact of Technology on Journalism. *Journalism Studies* 1, 2 (2000), 229–237.

[10] Tan Tang, Junxiu Tang, Jiayi Hong, Lingyun Yu, Peiran Ren, and Yingcai Wu. 2020. Design guidelines for augmenting short-form videos using animated data visualizations. *Journal of Visualization* (2020), 1–14.