Animated Pictorial Maps

Dong-Yi Wu
National Cheng Kung University
Taiwan
cutechubbit@gmail.com

Li-Kuan Ou
National Cheng Kung University
Taiwan
k777k777tw@gmail.com

HuiGuang Huang
National Cheng Kung University
Taiwan
604300468hhg@gmail.com

Yu Cao
The Hong Kong Polytechnic University
Hong Kong
yu-daniel.cao@connect.polyu.hk

Xin-Wei Lin
National Cheng Kung University
Taiwan
p78123029@gs.ncku.edu.tw

Thi-Ngoc-Hanh Le
International University, VNU-HCM
Vietnam
ltnhanh@hcmiu.edu.vn

Sheng-Yi Yao
National Cheng Kung University
Taiwan
nd8081018@gs.ncku.edu.tw

Tong-Yee Lee
National Cheng Kung University
Taiwan
tonylee@mail.ncku.edu.tw

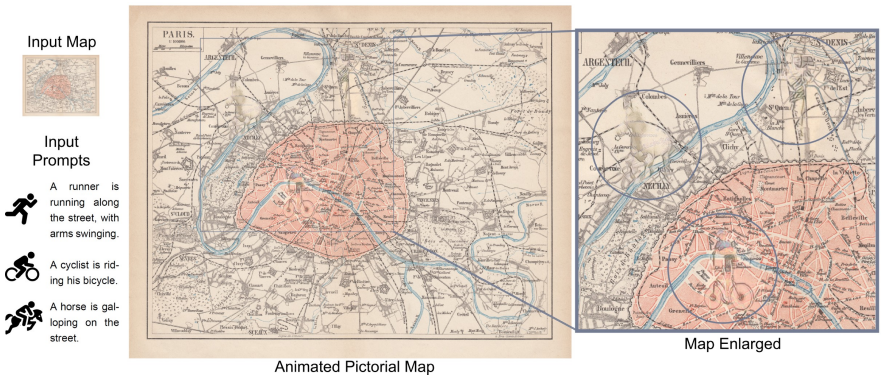

Figure 1: Our method creates a pictorial map showing the 2024 Olympic events and their corresponding venue on a historic map of Paris. We show three sports events: cycling, athletics and equestrian. Our method takes an input map and some content prompts and generates a video sequence matching the style of the map and while preserving the original map content.

ACM Reference Format:
Dong-Yi Wu, Li-Kuan Ou, HuiGuang Huang, Yu Cao, Xin-Wei Lin, Thi-Ngoc-Hanh Le, Sheng-Yi Yao, and Tong-Yee Lee. 2024. Animated Pictorial Maps. In *SIGGRAPH Asia 2024 Posters (SA Posters '24), December 03–06, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3681756.3697896

1 Introduction

Maps have been one of the most fundamental ways we record and communicate spatial information. From fifteenth-century early world maps to Google Maps in our daily lives, maps have been an indispensable tool. Creating a good map involves technical precision of presenting geographic information and artistic design including the choice of color scheme, spacing between map elements, text font, etc. In recent years, AI technology has been used to address various challenges faced in cartography. For example, generative adversarial networks (GAN) have been adopted for map style transfer between contemporary maps and old maps [Li et al. 2021] or U-Net has been trained to add 3D shading to flat 2D maps [Jenny et al. 2021]. However, the most of these works focus on static maps. In this paper, we propose a framework for generating animated maps with text prompts. This work is inspired by a specific type of map called pictorial maps. In contrast to road maps or topographic maps, pictorial maps often depict a given location with illustrations of architecture, people and animals. Extending this idea, we build a framework for animated pictorial maps with style-consistent dynamic objects that blend well with the original map content.

SA Posters '24, December 03–06, 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1138-1/24/12
https://doi.org/10.1145/3681756.3697896

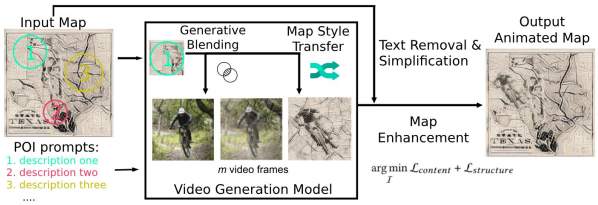

Figure 2: System overview

2 Method

The overview of our proposed method is shown in Figure 2. Our system takes any maps as input and we let users annotate the map with point-of-interest (POI) and the corresponding text prompts, called *POI prompts*. For each given POI prompt, we use our system to generate a corresponding sequence of video frames $\mathcal{V} \in \mathbb{R}^{m \times H \times W \times 3}$ that is style-consistent and blend naturally with the map. Our framework has three main components: Map Generative Blending, Map Style Transfer and Map Enhancement. Map Generative Blending and Map Style Transfer are built on top of a diffusion-based video generation model (we use Text2Video-Zero [Khachatryan et al. 2023] in our experiment). They enable the video generation model to handle the content blending and style transfer while generating the video frames. Furthermore, we observe that readers of maps recognize the map usually by its most salient features, e.g. large rivers or main roads. The generation process might destroy these important features. We therefore enhance the map with its most important features in the Map Enhancement step.

Map Generative Blending. The challenge in this work is how to overlay two pieces of content while minimizing the interference between them. Our strategy is two-fold: First, we selectively show the content at the location when it has more relevance. For example, the midsection of the Eiffel Tower is more relevant compared to its feet and therefore we show the midsection with higher opacity in Figure 3(b) and vice versa. This relevance map can be obtained easily by inspecting the cross-attention layer of the Stable Diffusion model. We transform the attention map to an opacity map for the video content. Secondly, we increase the coherence of the blending by projecting the blended image back to the distribution of the natural image with the help of the denoising diffusion model. That is, we blend the two contents in the slightly noisy latent code and denoise the mixture with the Stable Diffusion denoising model. After the denoising steps, it is interesting to observe that the Eiffel Tower is better embedded into the road networks in Figure 3(b) and the road networks align cleverly with the Eiffel Tower's structure.

Map Style Transfer. During the video generation process, we also transfer the style of the original map when denoising the video frames. To transfer the style, a naive approach would be to swap the similar patches in the latent space between the content and style images. However, the most identifying features of a map often consist of very fine details like hand drawing strokes or paper texture. In this case, the naive fails to transfer the low-level details producing overly-smooth images, as shown in Figure 3(c). Instead, we propose swapping the features from shallower layers of Variational Auto-Encoder (VAE) of stable diffusion. We first decode the latent to pixel

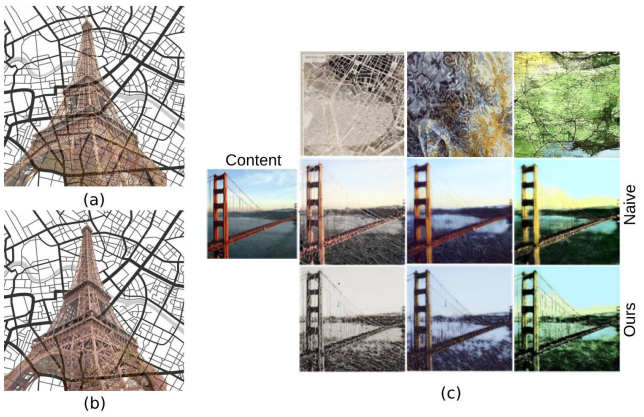

Figure 3: (a) Alpha blending in pixel space (b) Generative blending (c) Style transfer result comparison between the naive approach and ours.

space and color match the content and the style images. Then we encode the content and the style image and style-swapping the latent features from the second and the third DownBlock.

Map Enhancement. We optimize the generated pictorial maps I_c by emphasizing the main feature of the original map. The original map will first be simplified to retain only the essential information and remove noise like high-density roads and contour lines in mountainous areas. We also remove the text from the map. The processed map is denoted as I_s. The optimization is formulated as: $\arg\min_I \mathcal{L}_{content} + \mathcal{L}_{structure}$, where $\mathcal{L}_{content}$ and $\mathcal{L}_{structure}$ measure the VGG19 feature distance between our final maps I and the content map I_c and the processed map I_s, respectively.

3 Results And Conclusion

We test our method on both contemporary and ancient maps with various styles. Figure 1 shows an animated map generated with our method. Compared to the existing map animation software like MapCreator (https://mapcreator.io/), which only supports simple icon movement or simple path animation, our method is able to augment the map with rich content that blends naturally with the map elements. In conclusion, we propose a new form of map creation that makes animated cartography more approachable for non-experts. Currently, the quality and stability of the generated video vary depending on the given prompts, which somewhat limits the creators' imagination. We expect the continually evolving text-to-video model to resolve this issue.

Acknowledgments

This work is supported in part by the National Science and Technology Council under Grants 111-2221-E-006-112-MY3 and 113-2221-E-006-161-MY3, Taiwan.

References

Bernhard Jenny, Magnus Heitzler, Dilpreet Singh, Marianna Farmakis-Serebryakova, Jeffery Chieh Liu, and Lorenz Hurni. 2021. Cartographic Relief Shading with Neural Networks. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1225–1235.

Levon Khachatryan, Andranik Movsisyan, Vahram Tadevosyan, Roberto Henschel, Zhangyang Wang, Shant Navasardyan, and Humphrey Shi. 2023. Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators. In *IEEE International Conference on Computer Vision*. 15908–15918.

Zekun Li, Runyu Guan, Qianmu Yu, Yao-Yi Chiang, and Craig A. Knoblock. 2021. Synthetic Map Generation to Provide Unlimited Training Data for Historical Map Text Detection. In *ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 17–26.