Video editing with Gaussian splatting

Advisor: Javier Vazquez-Corral & Danna Xue

Insitution: Computer Vision Center, Universitat Autònoma de Barcelona

Background:

Video editing plays a crucial role in today's creative industries, from film-making and advertising to social media content creation. It enables creators to bring their visions to life, enhancing storytelling through visual effects, scene adjustments, and content personalization. However, most current video editing techniques face several challenges: they require extensive manual intervention, lack consistent control over edits, and are time-consuming, limiting their scalability for real-world applications.

Previous works [1,2] try to conduct video editing based on the implicit neural representation model, the model learns a 2d texture map which can be mapped by the model into different frames, so that the editing can be simply conducted on this texture map and produce consistent editing in the whole sequence. However, the video captures the objects and the scene in a 3d world, and a 2d texture map produces some unrealistic visual effects when representing the videos. Moreover, the implicit neural representation for representing a video needs to be trained for around 20 minutes and even longer, which is not efficient enough for interactive editing.

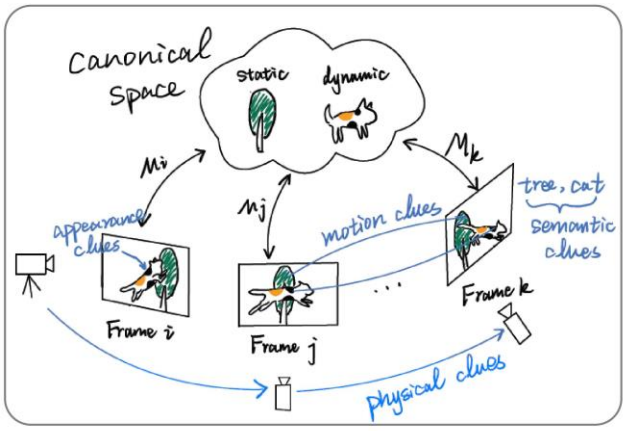
Gaussian Splatting (GS) [3] is a state-of-the-art technique for novel view synthesis—the task of generating photorealistic images of a scene from viewpoints not seen in the original input photographs. It has become a sensation for its ability to achieve unprecedented rendering speeds and high visual quality, often rendering at 100+ frames per second on modern GPUs. 3DGS/2DGS represents a significant shift from prior approaches like Neural Radiance Fields (NeRFs) by using an explicit, intuitive scene representation instead of a large, implicit neural network.

Recent works extended GS to image and video rendering [4,5,6], which can be a promising way of image compression and representation. However, there are still some problems in representing the long videos [5] with large motion and conduct artifacts-free editing on the canonical Gaussian representation.

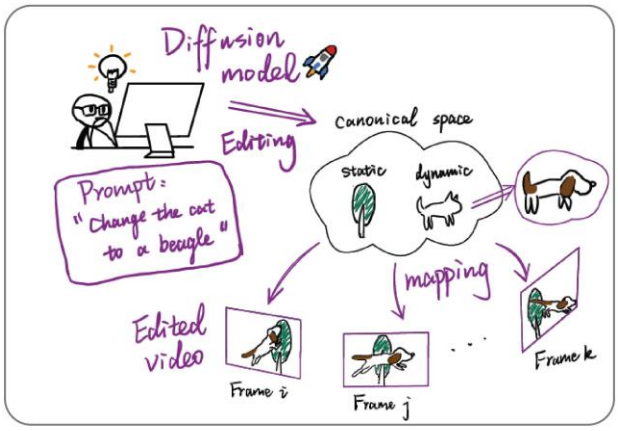
Target:

1. Build a compact and efficient representation for video with 2D/3D GS.
2. Build a framework for consistent video editing on GS. Edit in the canonical space build by GS, while keeping the consistency in the whole sequence.

3. Use palette-based color adjustment method [7] or text-based image generation model for editing.



Step 1: Build the video representation



Step 2: Conduct video editing

What will you learn from the project?

1. Recent techniques in 2D/3D vision.
2. Knowledge about color editing.
3. Latest studies about video representation and editing with Gaussian splatting.

Reference:

[1] Kasten, Yoni, Dolev Ofri, Oliver Wang, and Tali Dekel. "Layered neural atlases for consistent video editing." ACM Transactions on Graphics (TOG) 40, no. 6 (2021): 1-12.

[2] Pilligua, Maria, Danna Xue, and Javier Vazquez-Corral. "HyperNVD: Accelerating Neural Video Decomposition via Hypernetworks." In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 22933-22942. 2025.

[3] Kerbl, Bernhard, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. "3D Gaussian splatting for real-time radiance field rendering." ACM Trans. Graph. 42, no. 4 (2023): 139-1.

[4] Zhang, Xinjie, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. "Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting." In European Conference on Computer Vision, pp. 327-345. Cham: Springer Nature Switzerland, 2024.

[5] Liu, Mufan, Qi Yang, Miaoran Zhao, He Huang, Le Yang, Zhu Li, and Yiling Xu. "D2gv: Deformable 2d gaussian splatting for video representation in 400fps." arXiv preprint arXiv:2503.05600 (2025).

[6] Wang, Longan, Yuang Shi, and Wei Tsang Ooi. "GSVC: Efficient Video Representation and Compression Through 2D Gaussian Splatting." In Proceedings of the 35th Workshop on Network and Operating System Support for Digital Audio and Video, pp. 15-21. 2025.

[7] Ren, Chenqu, Haolei Qiu, Yeheng Shao, Zherui Qiu, and Kaiwen Song. "PaletteGaussian: 3D Photorealistic Color Editing with Gaussian Splatting." In 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 1206-1215. IEEE, 2024.