Realistic Surgical Simulation from Monocular Videos Supplementary Materials

A Overview

The contents of this appendix include:

1. Attachment Descriptions (Sec. B).
2. Details of Material Point Method (Sec. C).
3. More Experimental Results (Sec. D).
4. More Implementation Details (Sec. E).

B Attachment Descriptions

We strongly suggest reviewing our attached **HTML page**, which contains the following materials:

1. **More Visual Results:** The video page includes demo videos showing different surgical operations across various real surgical scenes. We also provide the complete videos referenced in Fig. 3 of the paper, alongside quantitative comparisons between four methods (Pcd, Mesh, Baseline, SurgiSim) and ablation results without Video Guide.
2. **User Study Page:** The original interface used in our user study, containing 9 sets of videos for qualitative comparison and Video Guide ablation analysis.

Please click on 'index.html' and select 'Videos' to view all attachments.

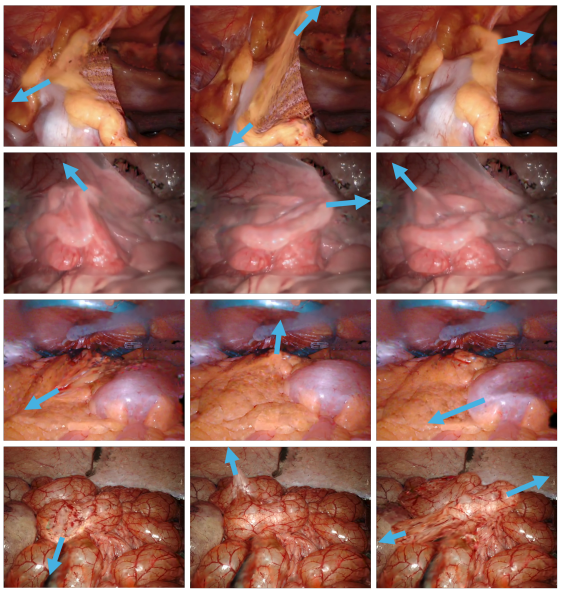

Figure 1: Visualization of more simulation cases on different scenes using SurgiSim.

Metric	SurgiSim	EndoNeRF	EndoSurf	LerPlane	4D-GS	EndoGaussian
PSNR↑	35.490	27.077	34.795	34.643	22.832	36.31
SSIM↑	0.966	0.900	0.945	0.922	0.827	0.971
LPIPS↓	0.056	0.107	0.119	0.072	0.368	0.050

Table 1: Quantitative comparison of reconstruction quality on the EndoNeRF dataset. While our method is not specifically designed for novel-timestamp synthesis, it achieves competitive performance across all metrics.

C Details of Material Point Method

The full MPM methods include particle-to-grid (P2G) and grid-to-particle (G2P) to transfer properties between these particles and an Eulerian grid. Following Stomakhin et al. (2); Xie et al. (5), we use C^1 continuous B-spline kernels for two-way transfer. The mass and momentum are transferred from particles to grid nodes:

$$m_i = \sum_p m_p \, w_{ip}, \tag{1}$$

$$m_i \mathbf{v}_i = \sum_p m_p \left(\mathbf{v}_p + \mathbf{C}_p (\mathbf{x}_i - \mathbf{x}_p) \right) w_{ip}, \tag{2}$$

where m_i is the mass at grid node i, $m_p \mathbf{v}_p$ are the mass and velocity of particle p and $\mathbf{v}_i$ is the velocity at grid node i. $\mathbf{x}_i$ and $\mathbf{x}_p$ are the positions of grid node i and particle p, respectively, w_{ip} is the B-spline weighting function between particle p and grid node i and $\mathbf{C}_p$ is the affine velocity matrix of particle p, capturing local velocity gradients.

After grid velocities are updated, particle velocities and affine matrices are interpolated from the grid:

$$\mathbf{v}_p^{n+1} = \sum_i w_{ip} \mathbf{v}_i^{n+1}, \tag{3}$$

$$\mathbf{C}_p^{n+1} = \frac{12}{\Delta x^2 (b+1)} \sum_i w_{ip} \mathbf{v}_i^{n+1} (\mathbf{x}_i - \mathbf{x}_p)^\top, \tag{4}$$

where $\mathbf{v}_p^{n+1}$ is the updated velocity of particle p, $\mathbf{v}_i^{n+1}$ is the updated velocity at grid node i, $\mathbf{C}_p^{n+1}$ is the updated affine velocity matrix for particle p, Δx is the grid spacing and The term $(\mathbf{x}_i - \mathbf{x}_p)^T$ represents the transpose of the position difference vector.

D More Experimental Results

D.1 Visualization of Surgical Simulation

Fig. 1 presents additional simulation cases, performing different operations on each of the different surgical scenes. The results demonstrate SurgiSim's advanced capability to adapt to diverse new scenes and perform various kinds of operations, including severe pulling and cutting. See the videos for Fig. 1 and more demos in index.html.

D.2 Reconstruction Quality Analysis

We evaluate SurgiSim's reconstruction capabilities against state-of-the-art dynamic tissue reconstruction methods on the EndoNeRF dataset (3). For a fair comparison following the protocol in (6), we utilize the original dataset masks and stereo depth information rather than our SAM-refined masks and estimated monocular depth.

Tab. 1 presents quantitative results comparing SurgiSim with EndoNeRF (3), EndoSurf (8), LerPlane (7), 4D-GS (4), and EndoGaussian (1). While these metrics primarily evaluate novel-timestamp synthesis capabilities—which is not the primary objective of SurgiSim—our method still achieves compelling results, consistently ranking second best across all metrics. This strong reconstruction performance, achieved as a byproduct of our focus on creating realistic surgical simulation environments, demonstrates SurgiSim's capability to accurately model tissue deformations across video frames, proving its powerful ability to extract high-quality canonical scenes from dynamic inputs.

E More Implementation Details

E.1 Training and Performance

The training is per scene. For each input video, it takes less than 4 minutes to build a canonical scene from the input video with Surface Thickening (Sec. 3.2) done. For physical parameter estimation, because we train in a rolling manner (Sec. 3.3), the time complexity concerning the length of the operation is $O(n^2)$. The mean optimization time is 10 minutes, but the time can reach 20 minutes for long sequences. When performing novel simulations after optimization, the simulation speed can reach 7 fps by setting the simulation step duration 10 times longer for fewer steps.

All experiments were conducted on a machine equipped with a Core i7-13700K CPU and a single NVIDIA RTX 4090 GPU, running Ubuntu 24.04. The code will be made public to promote the virtual surgery.

E.2 Details on User Study

To evaluate the fidelity of our simulations, we conducted a user study involving 68 participants including both surgeons and laypersons. These participants were categorized into two distinct groups: the Surgeons group, consisting of 44 board-certificated surgeons, and the Ordinary group, comprising 24 laypersons with no medical background.

The study was structured into two parts. In the first part, participants were tasked with selecting one most realistic simulation from four options, results from four different methods. In the second part, they were required to choose one superior simulation between the two presented results. Each participant was exposed to nine sets of simulation results, with the order of the sets randomized to prevent order bias. Before analysis, we executed a basic data-cleaning process to remove any invalid or outlier responses.

We provide the page used for our user study in the supplement materials. To view the page, just click on the 'index.html' and select 'User Study'.

E.3 Future Work

In the simulation environment setup, our method struggled to handle the invisible portions on the sides of the tissues, leading to imperfections in the texture generated by our thickening approach. Similarly, the texture on the exposed areas after cutting still lacked sufficient realism and required manual correction. In the future, we hope to use diffusion or the large reconstruction model to fix these textures. Our method does not yet support more complicated operations like topological inversions of structure. In the future, we hope to build a geometry-aware MPM method based on 3D scene understanding techniques.

References

[1] Yifan Liu, Chenxin Li, Chen Yang, and Yixuan Yuan. Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. arXiv preprint arXiv:2401.12561, 2024.

[2] Alexey Stomakhin, Craig Schroeder, Lawrence Chai, Joseph Teran, and Andrew Selle. A material point method for snow simulation. ACM Transactions on Graphics (TOG), 32(4):1–10, 2013.

[3] Yuehao Wang, Yonghao Long, Siu Hin Fan, and Qi Dou. Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In International conference on medical image computing and computer-assisted intervention, pp. 431–441. Springer, 2022.

[4] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20310–20320, 2024.

[5] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4389–4398, 2024.

[6] Mengya Xu, Ziqi Guo, An Wang, Long Bai, and Hongliang Ren. A review of 3d reconstruction techniques for deformable tissues in robotic surgery. arXiv preprint arXiv:2408.04426, 2024.

[7] Chen Yang, Kailing Wang, Yuehao Wang, Xiaokang Yang, and Wei Shen. Neural lerplane representations for fast 4d reconstruction of deformable tissues. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 46–56. Springer, 2023.

[8] Ruyi Zha, Xuelian Cheng, Hongdong Li, Mehrtash Harandi, and Zongyuan Ge. Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In International conference on medical image computing and computer-assisted intervention, pp. 13–23. Springer, 2023.