EndoPlanar: Deformable Planar-based Gaussian Splatting for Surgical Scene Reconstruction

Thatphum Paonim[1 [0009−0004−1337−3655]], Chayapon Sasnarukkit[1 [0009−0006−8099−6873]], Natawut Nupairoj[1 [00000−0002−7272−6797]], and Peerapon Vateekul[1 [0000−0001−9718−3592]]

Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
peerapon.v@chula.ac.th

Abstract. Accurate reconstruction of deformable soft tissues from endoscopic stereo videos is essential to improve surgical navigation and automation in robot-assisted image-guided procedures. While recent Gaussian splatting techniques achieve real-time rendering with impressive results on endoscopic datasets, conventional 3D Gaussian splatting methods suffer from volumetric biases, leading to inaccuracies in 3D geometry and depth estimation. To overcome these limitations, we propose EndoPlanar, a novel deformable planar-based Gaussian splatting approach. By flattening volumetric Gaussians to a 2D plane, our method enables unbiased depth computation and normal map estimation, which are difficult to achieve with traditional ellipsoidal Gaussians. Furthermore, we introduce a regularization strategy for smooth planar-derived normal maps to refine surface quality. Additionally, we enhance model initialization using Gaussian mixture-based background segmentation, improving the representation of unseen objects and accelerating convergence. We evaluate EndoPlanar on two standard benchmarks, EndoNeRF and StereoMIS, demonstrating promising performance by outperforming all baselines in reconstruction quality with PSNR of 34.51 dB while maintaining real-time inference speeds of 307.5 FPS.

Keywords: 3D reconstruction· Endoscopic reconstruction· Deformable scene · Planar-based Gaussian Splatting.

1 Introduction

Robotic-Assisted Minimally Invasive Surgery (RAMIS) relies on stereo endoscopes to capture dynamic tissue scenes, crucial for spatial awareness, navigation, and other downstream applications [3, 11, 12, 18]. However, complex endoscopic environments pose substantial challenges for traditional 3D reconstruction methods, especially under limited fields of view and occlusions caused by instruments.

Recent scene reconstruction approaches leverage neural radiance field (NeRF) [10], combining ray marching and MLPs for implicit scene representation. For endoscopy applications, EndoNeRF [14] pioneered NeRF-based endoscopic reconstruction, while EndoSurf [17] improved surface quality through integration

of the Surface Density Field. Despite advancements, these methods face significant computational demands. Hexplane [1] partially addresses this by encoding spatio-temporal features at sample points, reducing training time for dynamic scenes, although ray marching remains computationally intensive.

To address this limitation, 3D Gaussian Splatting (3DGS) [7] emerged as a solution with faster training and rendering speeds by representing scenes as Gaussian point clouds and employing rasterization. GaussianFlow [15] and 4DGS [8] extended this approach to dynamic scenes, while Endo4DGS [6] and Endo-Gaussian [9] adapted it for deformable soft tissues using Hexplane representations with tiny MLPs. Deform3DGS [16] further enhanced computational efficiency through learnable linear combinations. However, Gaussian-based methods still face geometric inaccuracies—their ellipsoidal nature can lead to suboptimal depth estimation and complicated surface normal calculation. To address these limitations, recent strategies such as PGSR [2] and 2DGS [5] propose flattening the representation and using planar equations for more precise depth mapping.

Building upon these insights, we propose EndoPlanar, a real-time, deformable planar-based Gaussian splatting framework for reconstructing soft tissues from endoscopic stereo videos. Evaluations on the EndoNeRF and StereoMIS benchmarks demonstrate that EndoPlanar achieves state-of-the-art reconstruction fidelity at real-time speeds (307.5 FPS). The key contributions of our work include:

- Initialization with MOGF: Mixture of Gaussians Fusion strategy combines MOG2 background subtraction with foreground masking to initialize point clouds, enhancing convergence speed and representation of unseen objects.
- Planar-Based Deformable Gaussians: Flatten 3D Gaussians into 2D plane representations to eliminate volumetric bias, improving the accuracy of rendered depth maps and enabling unbiased surface normal estimation.
- Edge-Aware Normal Regularization: Regularization scheme compares rendered normals with local plane normals derived from depth maps, weighting with depth gradients to preserve sharp surface boundaries.

2 Method

In this section, we first describe the preliminary methods employed in this work, 3D Gaussian Splatting and the Flexible Deformation Model (FDM) (Section 2.1). Next, we introduce our model, EndoPlanar, as illustrated in Fig. 1. We first apply Mixture of Gaussians Fusion (MOGF) across all frames to detect motion and fill occlusions, producing a canonical point cloud (Section 2.2), which is then converted into an initial set of Gaussians' attributes, representing the reconstructed 3D scenes. Next, we deform these Gaussians from time 0 to time t using FDM, followed by unbiased depth rendering to obtain a color image, a depth map, and a normal map (Section 2.3). Finally, the Gaussians' attributes are optimized according to our training objective (Section 2.4).

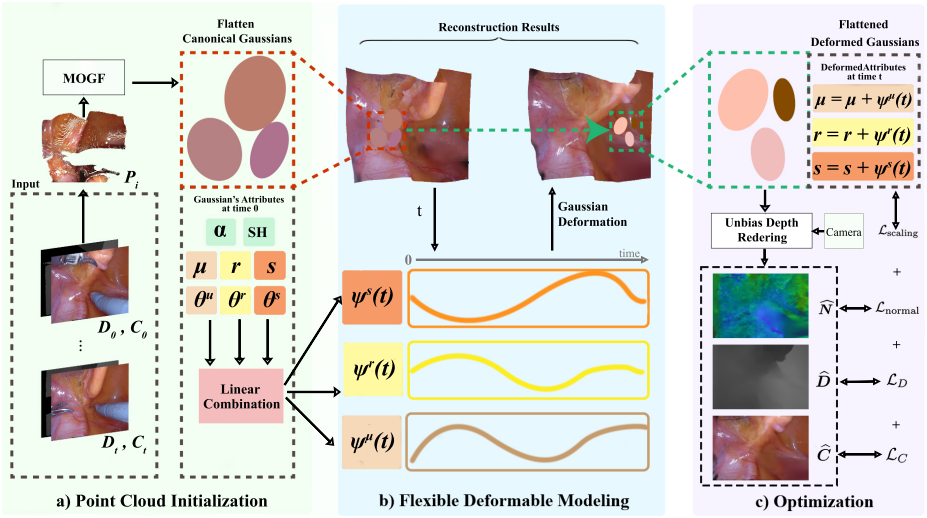

Fig. 1. Overview of the EndoPlanar pipeline. The approach consists of three main stages: (a) Point Cloud initialization, (b) Flexible Deformation Modeling, and (c) Unbiased Depth Rendering and Optimization.

2.1 Preliminaries

3D Gaussian Splatting [7]: It explicitly represents a 3D static scene as a set of 3D Gaussians $\{G_i\}_{i=1}^n$. Each Gaussian contains five sets of learnable rendering-related attributes: a center $\boldsymbol{\mu}_i \in \mathbb{R}^3$, a rotation matrix $\boldsymbol{R}_i \in \mathbb{R}^{3\times3}$, a scaling matrix $\boldsymbol{S}_i \in \mathbb{R}^{3\times3}$, an opacity o_i, and spherical harmonic coefficients (SH) for view-dependent color. The color map $\widehat{\boldsymbol{C}}$ in Fig. 1 can be calculate using α-blending:

$$\widehat{\boldsymbol{C}} = \sum_{i=1}^{n}\Big(c_i\,\alpha_i\prod_{j=1}^{i-1}(1-\alpha_j)\Big), \tag{1}$$

where c_i is the color derived from the SH coefficients, and α_i is the opacity contribution computed by evaluating the 2D Gaussian $\boldsymbol{\Sigma}_i'$ scaled by o_i.

Deformable Gaussian Splatting [16]: To extend Gaussian Splatting into 4D settings, we adopt the Flexible Deformation Modeling (FDM) scheme, as shown in Fig. 1. This approach explicitly represents the temporal deformation fields of Gaussians' attributes. Specifically, for each Gaussian, its center $\boldsymbol{\mu}_i$, rotation matrix $\boldsymbol{R}_i$, and scaling matrix $\boldsymbol{S}_i$ are associated with a deformation field modeled as a linear combination of Gaussian basis functions. Each basis function $b(t;\theta,\sigma)$ is defined as $b(t;\theta,\sigma) = \exp\big(-(t-\theta)^2/(2\,\sigma^2)\big)$, where θ and σ are learnable parameters that control the center and width of the function. Given a learnable amplitude ω_j for each of B basis functions, we form the final deformation field

$\psi(t)$ and add the result back to the base Gaussians at time 0, thus obtaining the deformed Gaussians at time t. Formally, $\psi(t)$ is defined as

$$\psi(t) = \sum_{j=1}^{B} \omega_j \, b\big(t; \theta_j, \sigma_j\big). \tag{2}$$

2.2 Point Cloud Initialization

We introduce Mixture of Gaussians Fusion (MOGF), an extension of Motion-Aware Point Fusion (MAPF) [16]. First, we replace threshold-based motion detection with a statistical background-subtraction approach using Mixture of Gaussians (MOG2) [19] by updating the distribution over all frames except the first one. We then apply the learned statistical background model to the first frame in order to detect subtle motions. Second, we retain the occlusion filling workflow from MAPF to reconstruct areas occluded by surgical instruments. Starting with a canonical point cloud from the initial frame, we project the combined result of these two components into 3D:

$$\boldsymbol{P_i} = \boldsymbol{K}^{-1}, \boldsymbol{D_i}, \big(\boldsymbol{I_i} \odot \boldsymbol{M_i}\big), \tag{3}$$

where $\boldsymbol{K}$, $\boldsymbol{P_i}$, $\boldsymbol{I_i}$, $\boldsymbol{D_i}$, and $\boldsymbol{M_i}$ represent the intrinsic camera matrix, 3D point cloud, 2D pixel coordinates, depth map, and valid mask of the i-th frame, respectively, with $\odot$ denoting element-wise multiplication. Finally, we merge these projections with the initial cloud to form the final canonical point cloud.

2.3 Deformable Planar-Based Gaussian Splatting

3D Gaussians represent ellipsoidal shapes, making direct estimation of geometric attributes challenging. To address this, [2, 5] proposed flattening each 3D Gaussian into a 2D plane, enabling unbiased estimation of normals and depth. However, their method is designed for static scenes, whereas endoscopic scenes involve soft tissues undergoing deformation over time. Therefore, we extend their method to a 4D setting to handle these time-varying deformations.

Flattening Deformable 3D Gaussians. Each Gaussian is defined by $\boldsymbol{\Sigma}_i = \boldsymbol{R}_i \, \boldsymbol{S}_i \, \boldsymbol{S}_i^\top \, \boldsymbol{R}_i^\top$, where $\boldsymbol{R}_i$ is an orthonormal basis and $\boldsymbol{S}_i = \mathrm{diag}(s_1, s_2, s_3)$. To enforce progressive flattening over time, we minimize the smallest scale factor:

$$\mathcal{L}_{\text{scaling}}(t) = \big\| \min\big(s_1(t), s_2(t), s_3(t)\big) \big\|, \quad s_i(t) = \exp\big(s_{i,0} + \psi_i(t)\big). \tag{4}$$

Here, $s_{i,0}$ are the base scale factors at $t = 0$, and $\psi_{s_i}(t)$ represents the deformable field that modifies the scale factor over time. We use exponential scaling to ensure positivity and initialize $s_{3,0}$ to a large negative value (-10^9) to make $s_3(t)$ close to zero for numerical stability and fast convergence.

Unbiased Depth Rendering. After using FDM to deform the Gaussians to time t, we render normals and depth using their deformed attributes. For Gaussian i, we select the axis with the smallest scale factor as its normal vector, $\boldsymbol{n}_{i,t}$, and flip it if necessary to ensure that it always faces away from the camera. The plane distance is $d_{i,t} = \big(\boldsymbol{R}_{wc}(\boldsymbol{\mu}_{i,t} - \boldsymbol{T}_c)\big) \cdot \big(\boldsymbol{R}_{wc}\,\boldsymbol{n}_{i,t}\big)$, where $\boldsymbol{\mu}_{i,t}$ is the deformed Gaussian center at time t, $\boldsymbol{T}_c$ is the camera center in the world coordinates and $\boldsymbol{R}_{wc}$ is the rotation from the world to the camera coordinates. We then obtain the final normal map $\widehat{\boldsymbol{N}}$ and plane distance map $\boldsymbol{D}_\perp$ through α-blending:

$$\widehat{\boldsymbol{N}} = \sum_{i=1}^{N} \boldsymbol{R}_{wc}\,\boldsymbol{n}_{i,t}\,\alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \quad \boldsymbol{D}_\perp = \sum_{i=1}^{N} d_{i,t}\,\alpha_i \prod_{j=1}^{i-1}(1-\alpha_j). \tag{5}$$

For a pinhole camera with intrinsic matrix $\boldsymbol{K}$, each pixel $\boldsymbol{p} = [u, v]$ corresponds to a ray $\boldsymbol{r}(\boldsymbol{p}) = \boldsymbol{K}^{-1}[u, v, 1]^\mathsf{T}$. The final depth map is then computed by intersecting these rays with the blended plane. Formally,

$$\widehat{\boldsymbol{D}}(\boldsymbol{p}) = \frac{\boldsymbol{D}_\perp(\boldsymbol{p})}{\widehat{\boldsymbol{N}}(\boldsymbol{p}) \cdot \boldsymbol{r}(\boldsymbol{p})}. \tag{6}$$

2.4 Optimization

Normal Surface Regularization: To refine surface normals, especially in regions lacking direct supervision, we introduce a regularization term that enforces consistency between the rendered normals $\widehat{\boldsymbol{N}}(\boldsymbol{p})$ with a local plane normal $\boldsymbol{N}_d(\boldsymbol{p})$ derived from the depth map. For each pixel $\boldsymbol{p}$, we sample four neighboring pixels, project their depths into 3D points $\{\boldsymbol{P}_j\}_{j=0}^{3}$, and compute:

$$\boldsymbol{N}_d(\boldsymbol{p}) \;=\; \frac{\big(\boldsymbol{P}_1 - \boldsymbol{P}_0\big) \,\times\, \big(\boldsymbol{P}_3 - \boldsymbol{P}_2\big)}{\big\|\big(\boldsymbol{P}_1 - \boldsymbol{P}_0\big) \,\times\, \big(\boldsymbol{P}_3 - \boldsymbol{P}_2\big)\big\|}. \tag{7}$$

Because color edges in endoscopic images may not reflect true surface boundaries, we employ an edge-aware weight based on the ground-truth depth gradient. Let $\boldsymbol{D}$ be the ground-truth depth map. Our loss is:

$$\mathcal{L}_{\text{normal}} \;=\; \frac{1}{|W|} \sum_{\boldsymbol{p} \in W} \Big(1 - \|\boldsymbol{\nabla} \boldsymbol{D}(\boldsymbol{p})\|\Big)^2 \big\|\boldsymbol{N}_d(\boldsymbol{p}) - \widehat{\boldsymbol{N}}(\boldsymbol{p})\big\|_1, \tag{8}$$

where W is the set of valid pixels. This weighting reduces the penalty in regions with large depth gradients, thus preventing over-smoothing of sharp surface boundaries.

Training: Given a tissue mask $\boldsymbol{M}$, we supervise the reconstructed scene using ground-truth color images $\boldsymbol{C}$ and stereo depth maps $\boldsymbol{D}$, in addition to our normal regularization. Concretely, we render both color $\widehat{\boldsymbol{C}}$ and depth $\widehat{\boldsymbol{D}}$ from the current reconstruction at each training iteration, then impose the following losses:

$$\mathcal{L}_C \;=\; \big\|\,\boldsymbol{M} \odot \big(\widehat{\boldsymbol{C}} - \boldsymbol{C}\big)\big\| \quad , \quad \mathcal{L}_D \;=\; \big\|\,\boldsymbol{M} \odot \big(\widehat{\boldsymbol{D}}^{-1} - \boldsymbol{D}^{-1}\big)\big\|. \tag{9}$$

Given λ_C, λ_D, λ_{normal}, and λ_{scaling} balancing weights, our training objective is

$$\mathcal{L} \;=\; \lambda_C\,\mathcal{L}_C \;+\; \lambda_D\,\mathcal{L}_D \;+\; \lambda_{\text{normal}}\,\mathcal{L}_{\text{normal}} \;+\; \lambda_{\text{scaling}}\,\mathcal{L}_{\text{scaling}}, \tag{10}$$

3 Experiment

3.1 Experimental Setup

We evaluated our method on EndoNeRF [14] and StereoMIS [4]. For StereoMIS, we sampled content-driven segments (180-300 frames) from videos P2_6, P2_7, and P3, ensuring each segment captured a complete action. We then applied [13] to generate pseudo ground-truth depth. Each scene's frames were split 7:1 for training and testing. We measured rendering quality using PSNR, SSIM, and LPIPS, and evaluated surface quality with depth RMSE. We also recorded training time and rendering speed. Our model uses 17 learnable Gaussian basis functions for FDM and trains for 3,000 iterations with an initial learning rate of 1.6×10^{-3}. As the depth loss $\mathcal{L}_D$ was initially much smaller than the photometric loss $\mathcal{L}_C$, we tuned λ_D on selected EndoNeRF video, increased λ_D until the depth metrics ceased to improve, then applied λ_D uniformly on all datasets. We also use large λ_{scaling} to ensure properly flattening of Gaussians. The final weights for $\lambda_C, \lambda_D, \lambda_{\text{normal}}, \lambda_{\text{scaling}}$ are $1, 5, 0.0085$, and 100, respectively. All experiments were implemented in PyTorch on an NVIDIA RTX A5000 GPU.

Table 1. Quantitative results across subsets for each dataset. Our model, EndoPlanar, is listed in the 'Ours' row. Bold values highlight the winner for each metric.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓	RMSE↓	Train(s)↓	FPS↑
EndoNeRF	EndoNeRF	34.20	93.50	0.1556	2.09	∼23,800	0.025
	EndoSurf	34.84	95.39	0.1125	1.77	∼43,200	0.023
	EndoGaussian	37.88	96.09	0.0510	2.16	146	145
	Deform3DGS	38.59	96.40	0.0531	2.41	**60**	**340**
	Ours	**38.67**	**96.61**	**0.0461**	**1.64**	73	339
StereoMIS	EndoNeRF	27.88	73.47	0.3490	3.53	∼23,800	0.025
	EndoSurf	28.02	78.92	0.2870	3.53	∼43,200	0.023
	EndoGaussian	29.63	81.39	0.2550	5.08	155	131
	Deform3DGS	30.30	83.71	0.2223	5.20	**62**	**303**
	Ours	**30.41**	**84.59**	**0.2073**	**3.37**	89	276

3.2 Main Results

We evaluated our proposed method by comparing its performance with other state-of-the-art approaches: NeRF-based (EndoNeRF [14] and EndoSurf [17]) and Gaussian Splatting-based (EndoGaussian [9] and Deform3DGS [16]) methods. As shown in Table 1, our method achieves superior scores in PSNR, SSIM,

Table 2. An ablation study where each proposed strategy is progressively added to our model's backbone. Bold values highlight the best performance for each metric.

Method	PSNR↑	SSIM↑	LPIPS↓	RMSE↓	Train(s)↓	FPS↑
MAPF	38.55	96.44	0.0495	1.58	**65**	**381**
MOGF	**38.86**	**96.70**	**0.0422**	**1.51**	71	342
MOGF+$\mathcal{L}_{\text{normal}}$	38.67	96.61	0.0461	1.64	72	339

LPIPS, and depth's RMSE compared to both NeRF-based and Gaussian Splatting-based methods in both datasets. In terms of speed, our method maintains training times (73–89s) close to Deform3DGS (60–62s) and is significantly faster than EndoGaussian (146–155s), EndoNeRF (6.5h) and EndoSurf (12h). The rendering rates also remain highly competitive, ranging from 276 to 339 frames per second, on par with Deform3DGS (303–340 FPS) and surpassing EndoGaussian (131–145 FPS). Consequently, our method delivers state-of-the-art reconstruction quality while preserving training and rendering speeds that are suitable for real-time applications.

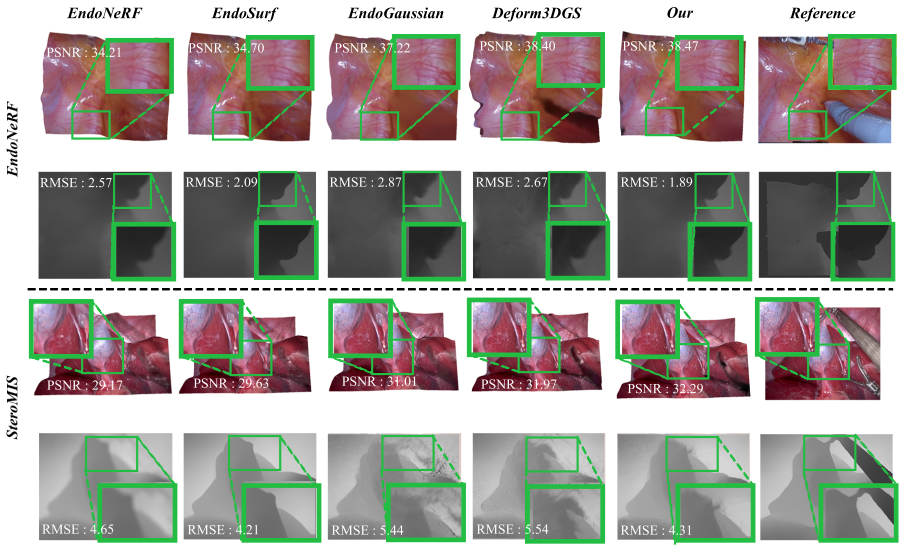

Fig. 2. Qualitative comparison with EndoNeRF, EndoSurf, EndoGaussian, and Deform3DGS. The boxed image is a zoomed-in view showing details in the focused area.

To provide additional qualitative results, we provide qualitative comparisons of several rendered scenes in Fig. 2. It can be observed that our method is better at preserving fine appearance details and offers improved interpolation in occlusion areas, where tissues are temporarily obscured by surgical tools. In Fig.

3, we visualize surface quality using a normal map generated from the reconstructed mesh of StereoMIS, P3. As shown, our method provides smoother and more coherent surfaces than the other Gaussian Splatting–based methods, which exhibit pronounced noisy surfaces. The NeRF based approaches, by virtue of their implicit MLP representation, typically offer smoother surfaces than Gaussian splatting-based methods. However, our approach still surpasses EndoNeRF, particularly in regions where tools occlude underlying tissue, although the performance gap is narrower than with the other Gaussian Splatting baselines. Furthermore, our method achieves surface quality comparable to EndoSurf, which incorporates additional MLPs for a dedicated Surface Density Field. Despite EndoSurf's emphasis on fine-grained surface modeling, our approach achieves similar surface smoothness while maintaining higher photometric rendering quality and computational efficiency.

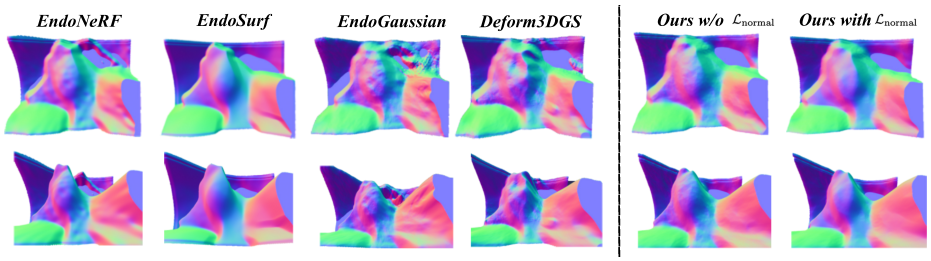

Fig. 3. Qualitative comparison of a normal surface reconstruction methods compared to existing techniques and ablation of normal loss. The first and second rows show the same model from slightly different viewpoints.

3.3 Ablations

We perform comprehensive ablation studies with EndoNeRF dataset, examining two key components of our approach: initialization strategy and normal regularization. As shown in Table 2, both components contribute significantly to our method's performance. For initialization, we compare the MAPF method from [16] with our MOGF initialization. Our approach achieves superior scores across all metrics, with only a slight reduction in training and inference speed. This trade-off results from generating motion awareness initial sparse point clouds, leading to a higher number of Gaussians overall. Furthermore, we examine the impact of our normal regularization term $\mathcal{L}_{normal}$ by comparing models trained with and without it. Incorporating $\mathcal{L}_{normal}$ leads to marginal decreases in rendering metrics (Table 2) but produces noticeably more coherent surfaces (Fig. 3). Thus, while enabling $\mathcal{L}_{normal}$ involves a slight trade-off in appearance metrics, it significantly enhances surface fidelity, which is critical for accurate endoscopic visualization.

4 Conclusion

In this paper, we presented EndoPlanar, a deformable planar-based Gaussian splatting framework for endoscopic stereo video reconstruction. By flattening 3D Gaussians and leveraging our unbiased depth rendering pipeline, we mitigate volumetric biases and achieve more accurate surface geometry. Our Mixture of Gaussians Fusion (MOGF) accelerates convergence and enhances rendering quality, while our edge-aware normal regularization improves surface detail, enabling reliable and efficient endoscopic reconstruction. Evaluated on EndoNeRF and StereoMIS benchmarks, EndoPlanar achieves state-of-the-art reconstruction quality while maintaining training and rendering speeds suitable for real-time applications.

Disclosure of Interests. The authors have no competing interests to declare that they are relevant to the content of this article.

References

1. Cao, A., Johnson, J.: Hexplane: A fast representation for dynamic scenes. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 130–141 (2023)
2. Chen, D., Li, H., Ye, W., Wang, Y., Xie, W., Zhai, S., Wang, N., Liu, H., Bao, H., Zhang, G.: Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. IEEE transactions on visualization and computer graphics **PP** (2024)
3. Chong, N., Si, Y., Zhao, W., Zhang, Q., Yin, B., Zhao, Y.: Virtual reality application for laparoscope in clinical surgery based on siamese network and census transformation. In: MICAD (2021), https://api.semanticscholar.org/CorpusID:238691157
4. Hayoz, M., Hahne, C., Gallardo, M., Candinas, D., Kurmann, T., Allan, M., Sznitman, R.: Learning how to robustly estimate camera pose in endoscopic videos. International Journal of Computer Assisted Radiology and Surgery **18**, 1185 – 1192 (2023)
5. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: SIGGRAPH 2024 Conference Papers. Association for Computing Machinery (2024). https://doi.org/10.1145/3641519.3657428
6. Huang, Y., Cui, B., Bai, L., Guo, Z., Xu, M., Ren, H.: Endo-4dgs: Endoscopic monocular scene reconstruction with 4d gaussian splatting. ArXiv **abs/2401.16416** (2024)
7. Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (TOG) **42**, 1 – 14 (2023)
8. Lin, Y., Dai, Z., Zhu, S., Yao, Y.: Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 21136–21145 (2023)
9. Liu, Y., Li, C., Yang, C., Yuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction (2024)

10. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
11. Ramalhinho, J., Yoo, S., Dowrick, T., Koo, B., Somasundaram, M., Gurusamy, K., Hawkes, D.J., Davidson, B.R., Blandford, A., Clarkson, M.J.: The value of augmented reality in surgery — a usability study on laparoscopic liver surgery. Medical Image Analysis **90** (2023)
12. Saikia, A., Vece, C.D., Bonilla, S., He, C., Magbagbeola, M., Mennillo, L., Czempiel, T., Bano, S., Stoyanov, D.: Robotic arm platform for multi-view image acquisition and 3d reconstruction in minimally invasive surgery. IEEE Robotics and Automation Letters **10**, 3174–3181 (2024)
13. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: European Conference on Computer Vision (2020), https://api.semanticscholar.org/CorpusID:214667893
14. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. ArXiv **abs/2206.15255** (2022)
15. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4d gaussian splatting for real-time dynamic scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20310–20320 (June 2024)
16. Yang, S., Li, Q., Shen, D., Gong, B., Dou, Q., Jin, Y.: Deform3dgs: Flexible deformation for fast surgical scene reconstruction with gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2024), https://api.semanticscholar.org/CorpusID:270067997
17. Zha, R., Cheng, X., Li, H., Harandi, M., Ge, Z.: Endosurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2023), https://api.semanticscholar.org/CorpusID:260091635
18. Zhou, H., Jayender, J.: Emdq-slam: Real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy videos. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention **12904**, 331–340 (2021)
19. Zivkovic, Z.: Improved adaptive gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. vol. 2, pp. 28–31 Vol.2 (2004). https://doi.org/10.1109/ICPR.2004.1333992