Endoscopic Monocular Scene Reconstruction with Dynamic Gaussian Splatting and Motion Tracking

Gou Hao; Wang Changmiao; Yang Jiahao; Liu Yaoqun; Jia Fucang; Xiao Deqiang; Qin Feiwei*; Luo Huoling*

Contract: mipha0612@gmail.com

Highlights

- ❑ Compact Motion Representation: We design a set of low-dimensional Sim(3) motion bases to efficiently represent complex per-point motion as a linear combination, accurately modeling tissue deformation.

$$\mathbf{T}_{base} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & s \end{bmatrix} \in \mathrm{Sim}(3) \qquad \mathbf{T} = \sum \mathbf{w}^{(b)} \mathbf{T}_{base}^{(b)}$$

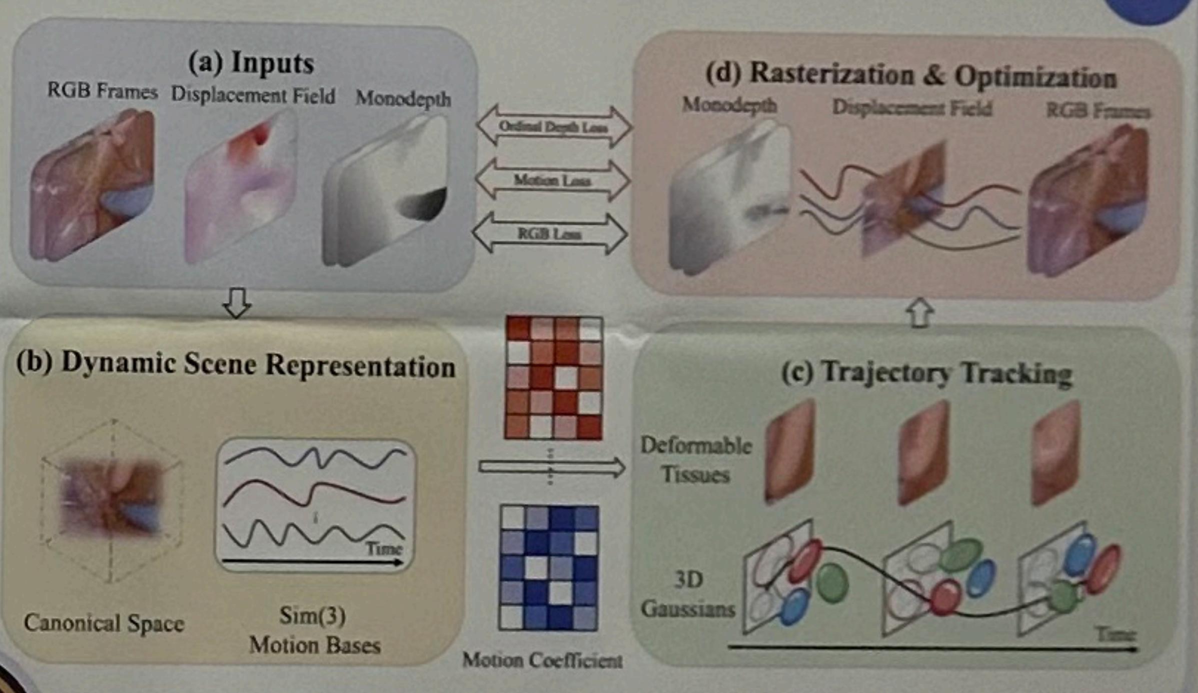
- ❑ A novel order-based depth loss function: While absolute depth values from monocular priors are scale-inconsistent, the relative depth order between pixels remains stable over time.

$$\mathcal{L}_{\mathrm{ordinal}} = \left\| \min(0, \mathrm{sign}(\hat{D}_t(p_1) - \hat{D}_t(p_2))) \times \mathrm{sign}(D_t(p_1) - (D_t(p_2))) \right\|$$

I Background & Challenges

- Monocular dynamic surgical scene reconstruction is highly challenging due to limited perspectives and complex- non-rigid tissue deformations.
- Existing methods often fail to fully leverage temporal correlations in video frames, leading to suboptimal performance.
- While many methods utilize depth priors, image-based approaches suffer from inter-frame inconsistency, causing depth flickering. Video-based methods overcome this issue but at a high computational cost.

Endo-GSMT Framework II



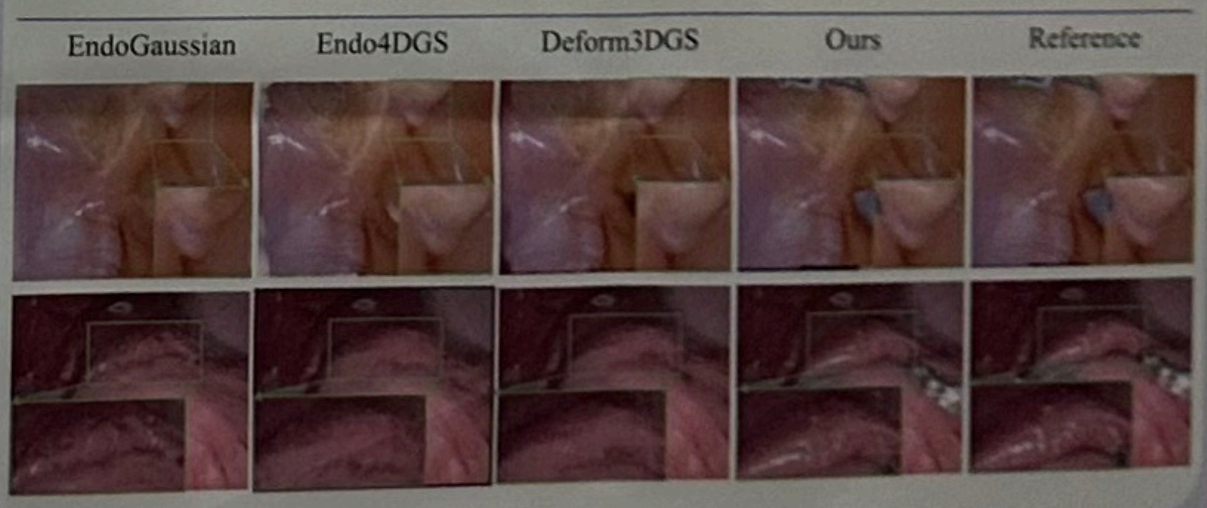
III Experiment Results

Dataset	Method	Frame Extraction Eva. PSNR↑	Frame Extraction Eva. SSIM↑	Frame Extraction Eva. LPIPS↓	NVS Eva. PSNR↑	NVS Eva. SSIM↑	NVS Eva. LPIPS↓
EndoNeRF (Normal)	EndoNeRF	28.355	0.918	0.090	-	-	-
	LerPlane-32k	38.238	0.948	0.055	-	-	-
	EndoGS	35.616	0.952	0.059	24.566	0.882	0.115
	EndoGaussian	35.522	0.957	0.103	27.949	0.905	0.096
	Endo-4DGS	36.945	0.957	0.037	28.318	0.909	0.092
	Deform3DGS	38.259	0.960	0.062	30.469	0.921	0.083
	Oours	**38.783**	**0.968**	**0.028**	**30.735**	**0.928**	**0.063**
StereoMIS (Hard)	EndoNeRF	31.922	0.857	0.146	-	-	-
	LerPlane-32k	31.679	0.845	0.113	-	-	-
	EndoGS	32.819	0.907	0.099	20.714	0.755	0.200
	EndoGaussian	29.191	0.827	0.181	23.098	0.721	0.227
	Endo-4DGS	32.580	0.862	0.124	27.461	0.802	0.156
	Deform3DGS	32.209	0.863	0.124	22.131	0.702	0.214
	Oours	**34.703**	**0.917**	**0.060**	**29.699**	**0.863**	**0.091**

Quantitative evaluation of our Endo-GSMT framework against existing two NeRF-based methods and four 3DGS-based methods. While other are evaluated using frames from the same viewpoint (Eva 1), we introduce a superior novel-view systhesis evaluation to the field (Eva 2).

Conclusion & Visualization IV

We propose Endo-GSMT, a novel framework based on 3DGS for the 3D reconstruction of dynamic scenes from monocular surgical videos.



Daejeon REPUBLIC OF KOREA MICCAI2025