T^{2}GS: Comprehensive Reconstruction of Dynamic Surgical Scenes with Gaussian Splatting

Jinjing Xu[1,2,3], Chenyang Li[1,2,3], Peng Liu[1,2,3], Micha Pfeiffer[1,2,3], Liwen Liu[1,2,3,4], Reuben Docea[1,2,3], Martin Wagner[3,5], and Stefanie Speidel[1,2,3,5] (✉)

[1] Department of Translational Surgical Oncology, National Center for Tumor Diseases (NCT/UCC Dresden), Dresden, Germany
[2] Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Germany
[3] University Hospital Carl Gustav Carus, TU Dresden, Dresden, Germany
[4] Technische Universität Dresden, Dresden, Germany
[5] The Centre for Tactile Internet with Human-in-the-Loop (CeTI), TU Dresden, Dresden, Germany
stefanie.speidel@nct-dresden.de

Abstract. Surgical scene reconstruction from endoscopic video is crucial for many applications in computer- and robot-assisted surgery. However, existing methods primarily focus on soft tissue deformation while often neglecting the dynamic motion of surgical tools, limiting the completeness of the reconstructed scene. To bridge the aforementioned research gap, we propose **T^{2}GS**, a novel and efficient surgical scene reconstruction framework that enables efficient spatio-temporal modelling of both deformable tissues and dynamically interacting surgical tools. T^{2}GS leverages Gaussian Splatting for dynamic scene reconstruction, and it integrates a recent tissue deformation modelling technique while most importantly, introduces a novel efficient tool motion model (ETMM). At its core, ETMM disambiguates the modelling process of tool's motion as global trajectory modelling and local shape-change modelling. We additionally propose pose-informed pointcloud fusion (PIPF), holistically initialized of tools' gaussians for improved tool motion reconstruction. Extensive experiments on public datasets demonstrate T^{2}GS's superior performance for comprehensive endoscopic scene reconstruction compared to previous methods. Moreover, as we specifically design our method with efficiency in concern, T^{2}GS also showcases promising reconstruction efficiency (3mins) and rendering speed (71fps), highlighting its potential for intraoperative applications. Our code is available at `https://gitlab.com/nct_tso_public/ttgs`.

Keywords: Scene reconstruction · Gaussian splatting · Dynamic scenes.

1 Introduction

Information about tool and tissue interaction in surgical scenes can benefit many downstream tasks, e.g., VR-based surgical training[2], surgical skill assessment[1,20] and visual enhancement[12]. In order to achieve a more comprehensive revelation

of the tool-tissue interaction, it is necessary to model the dynamics of both the tissue and the tool.

Recent advancements in dynamic surgical scene reconstruction have seen a shift from conventional methods[8] to modern learning-based approaches[21], such as NeRF [11] based methods and Gaussian splatting[7] technique, as these solutions greatly enhance the compactness of the representation, meanwhile facilitate high fidelity novel view rendering. Pioneer works along this path include EndoNeRF[14] and EndoSurf[23], they require high overhead for training and showcase slow rendering. Despite efforts[19,18] for acceleration, the efficiency of these group of methods is not comparable with the Gaussian spaltting based alternatives[21,5,10,16]. Among these GS methods, at core they leverage various 4D modeling techniques[9,15,22] to effectively reconstruct deforming tissue whereas ignoring interacted surgical tools, leading to incompletely reconstructed scenes. Our preliminary studies confirm that current methods are inadequate for tool motion modelling task, highlighting the need for specialized solutions.

On the other hand, proper Gaussian initialization is critical for convergence and reconstruction quality in 3DGS [7], while conventional SfM[13] methods assume scene rigidity thereafter being limited for surgical usage. Despite the efforts[21,5,10] in Gaussian initialization adapted for surgical scene, they focus on tissue reconstruction rather than tools, being improper to use for tools gaussian initialization, another critical component to concern in surgical scenes.

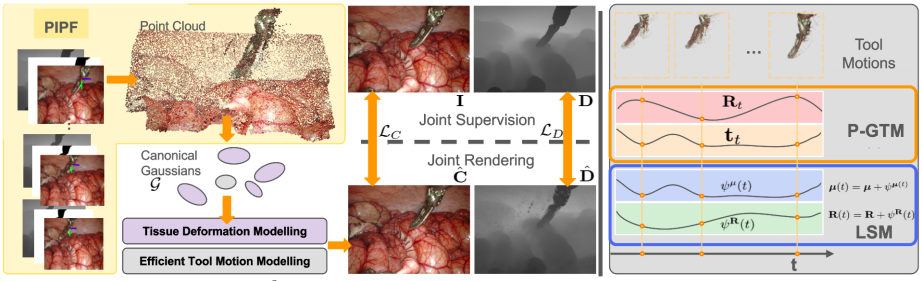
Therefore, to address the current limitation, we propose the framework T^2GS to model dynamics from both tissue and surgical tools. Our main contributions include:

- The first 4DGS-based framework with special designs for tool motion reconstruction and Gaussian initialization, that help to holistically reconstruct surgical scenes.
- Efficient Tool Motion Modelling module (ETMM), composed of Pose-guided Global Trajectory Modelling (P-GTM) and Local shape-change Modelling (LSM), together effectively model the deformation field of surgical tools.
- Efficient and high-quality reconstruction of comprehensive surgical scenes compared with other state-of-the-art methods, especially regarding tool reconstruction quality.

2 Method

2.1 Overview

As shown in Fig.1 (a) of T^2GS pipeline, during training stage, our framework starts with separate Gaussian initialization for tissue and individual tool, for the later we propose PIPF for holistic tools Gaussian initialization (Sec.2.3). Then tissue deformation and tool motion field are modelled separately with previous state of art tissue reconstruction method(Sec.2.2) and our proposed tool motion model ETMM (Sec.2.4). Then the composed gaussians of tissue and tools are jointly rendered (Sec.2.2) to obtain the color image and depth map, being supervised with groundtruth correspondence (Sec.2.5).



(a)T^{2}GS pipeline. (b)Efficient Tool Motion Modelling.

Fig. 1: Overview. In Fig.(a), we introduce the proposed complete pipeline, containing: 1) Gaussian initialization. 2) Deformation field modelling. 3) Joint rendering of composed gaussian. 4) Supervision. We detailed ETMM in Fig.(b).

2.2 Preliminaries

3D Gaussian Splatting (3DGS). In 3DGS[7], a scene is represented by a dense set of 3D Gaussians $\{\mathcal{G}_j\}$, each with a center $\boldsymbol{\mu}_j \in \mathbb{R}^3$, a decomposed covariance $\boldsymbol{\Sigma}_j = \mathbf{R}_j\mathbf{S}_j^2\mathbf{R}_j^\top$ represented as rotation $\mathbf{R}_j$ and scaling $\mathbf{S}_j$, per-Gaussian color $\mathbf{SH}_j$ and opacity $\boldsymbol{\alpha}_j$. Under a view transformation and projection, each 3D Gaussian projects to an elliptical 2D splat. The final color at a pixel $\mathbf{p}$ is obtained by alpha compositing these splats in front-to-back order via alpha blending:

$$\hat{\mathbf{C}}(\mathbf{p}) \;=\; \sum_j \Big(\boldsymbol{\alpha}_j \prod_{k<j}(1-\boldsymbol{\alpha}_k)\Big)\,\mathbf{c}_j, \tag{1}$$

where $\boldsymbol{\alpha}_j$ and $\mathbf{c}_j$ are the projected opacity and color of Gaussian $\mathcal{G}_j$. The depth image and the opacity map can be derived using analogous alpha blending formulations. By jointly optimizing all Gaussians' positions, shapes, and appearance, this method enables efficient and realistic rendering of 3D scenes.

Flexible Deformation Modeling. FDM[21] are proposed to be coupled with 3DGS and facilitate superior fast reconstruction of deforming tissue surface. We keep this method for the tissue reconstruction component in T^{2}GS framework. In addition, we adapt it to facilitate local shape-change modelling (LSM) of surgical tools in our proposed tool motion model ETMM, as later detailed in Sec.2.4. The FDM framework represents temporal deformations through a set of learnable basis functions that modulate Gaussian attributes. For each attribute type $\phi \in \{\boldsymbol{\mu}, \mathbf{R}, \mathbf{S}\}$, a deformation function $\psi^\phi(t)$ is defined as a linear combination of learnable basis functions: $\psi^\phi(t) = \sum_{k=1}^{K} w_k^\phi B_k(t)$ where $\{B_k(t)\}_{k=1}^K$ are the basis functions and $\{w_k^\phi\}_{k=1}^K$ are learnable weights. The final Gaussian attributes $\mathcal{G}(t)$ at time t are computed by adding these deformation functions to its canonical attributes: $\boldsymbol{\mu}(t) = \boldsymbol{\mu} + \psi^{\boldsymbol{\mu}(t)}$, $\mathbf{R}(t) = \mathbf{R} + \psi^{\mathbf{R}}(t)$, $\mathbf{S}(t) = \mathbf{S} + \psi^{\mathbf{S}}(t)$ where $\boldsymbol{\mu}, \mathbf{R}, \mathbf{S}$ are the aforementioned canonical position, rotation and scale at-

tributes. The weights are optimized per-gaussian during training to capture the observed deformations in the input sequence.

2.3 Pose Informed Initialization

Initial Pose Estimation The initial tool pose is estimated using 2D images. To obtain the 6DoF pose $\mathbf{T}_t^n$ of a tool instance $\mathbf{O}_{id=n}$, we use the state-of-the-art dense tracker CoTracker[6] to track tool points in 2D space (query points are initialized based on its instance mask), yielding 2D trajectories with visibility $\mathbf{v}_j^t \in \{0, 1\}$ at time t. For consecutive frames $(t, t+1)$, given established 2D-2D data-association of 2D points $\mathbf{p}_i^t$ and $\mathbf{p}_i^{t+1}$ from trajectory, we obatin 3D keypoints $\mathbf{p'}_i^t$ at time t, computed as $\mathbf{p'}_i^t = \mathbf{D}^t \cdot \mathbf{K}^{-1}[\mathbf{p}_i^t; 1]^T$ given $\mathbf{p}_i^t \in \mathcal{P}^t = \{\mathbf{p}_i^t \,|\, \mathbf{v}_i^t = 1\}$ and camera intrinsic $\mathbf{K}$. The relative pose of tool object $\mathbf{O}_{id=n}$ is estimated by solving:

$$\mathbf{T}_{t,t+1}^n = \arg\min_{\mathbf{R},\mathbf{t}} \sum_i \left\| \pi(\mathbf{R}\mathbf{p'}_i^t + \mathbf{t}) - \mathbf{p}_i^{t+1} \right\|^2 \tag{2}$$

using PnP algorithm. Global poses are accumulated as $\mathbf{T}_t^n = \prod_{k=0}^{t-1} \mathbf{T}_{k,k+1}$. We employ RANSAC for improved robustness to potential tool shape changing.

Pose Informed Point Fusion (PIPF) Our Pose Informed Point Fusion constructs the tool $\mathbf{O}_{id=n}$'s canonical point cloud $\mathbf{P}_{id=n}$ by aggregating depth observations across frames using estimated poses $\{\mathbf{T}_t^n\}$:

$$\mathbf{P}_{id=n} = \bigcup_{t=1}^{T} inv(\mathbf{T}_t^n) \left(\mathbf{D}_t \odot \mathbf{K}^{-1}[\mathbf{I}_t \odot \mathbf{M}_t^n] \right) \tag{3}$$

where $inv(\mathbf{T}_t^n)$ back-projects observed tool object pointcloud into the object's canonical space. This pose-aware fusion enables coherent initialization despite large tool movements.

2.4 Efficient Tool Motion Modelling

The dynamics of tools can be decomposed into global motions and local shape changes, modeled separately by pose-guided global trajectory modeling and local shape-change modeling, together forming our efficient tool motion modelling module.

Pose-guided Global Trajectory Modelling (P-GTM) We model the global trajectory of tool motion through a set of learnable object poses $\{\mathbf{T}_t\}$. The Gaussians for the object are transformed as a whole in this modelling process. For tool object Gaussians with mean $\boldsymbol{\mu}_o$ and rotation $\mathbf{R}_o$ in its local canonical space, their world-space position $\boldsymbol{\mu}_w^t$ and orientation $\mathbf{R}_w^t$ at time t are:

$$\boldsymbol{\mu}_w^t = \mathbf{R}_t \boldsymbol{\mu}_o + \mathbf{t}_t, \qquad \mathbf{R}_w^t = \mathbf{R}_t \mathbf{R}_o \tag{4}$$

where $\mathbf{T}_t = (\mathbf{R}_t, \mathbf{t}_t) \in \mathrm{SE}(3)$ is initialized from our aforementioned 6DoF pose estimates (as also used for PIPF) and refined during training. The world covariance $\boldsymbol{\Sigma}_w^t$ derives from $\mathbf{R}_w^t$ and canonical scale $\mathbf{S}_o$:

$$\boldsymbol{\Sigma}_w^t = \mathbf{R}_w^t \mathbf{S}_o \mathbf{S}_o^\top {\mathbf{R}_w^t}^\top \tag{5}$$

This global trajectory modelling significantly reduces learnable parameters to 6DoF per frame, enabling real-time optimization while maintaining physical plausibility through the rigidity assumption of surgical tools. Our experiments demonstrate the effective guidance from the known object pose for improved reconstruction quality.

Local shape-change Modelling(LSM) In addition to the global movement, surgical tools potentially present geometry shape changing (e.g. during its interaction with tissue). We propose to model these part of geometry changing with adapted-FDM. Specifically, we only keep the temporal parametrization of changes in rotation μ and position $\mathbf{r}$ among Gaussians attributes, while disabling the modelling of scale attribute $\mathbf{s}$ for improved convergence and reconstruction quality, as we demonstrate in our experimental session. This is intuitively reasonable because surgical tools consist of particles featuring negligible non-rigidity. Despite the constraint, our LSM still involves much higher flexibility compared to the aforementioned P-GTM, thereafter, it potentially can help to migrate the inaccuracy in global trajectory modelling, meanwhile being jointly optimized with P-GTM for overall optimal reconstruction.

2.5 Joint Optimization on Tool and Tissue Reconstruction

We jointly supervise both the color and depth predictions as conducted in [21], for which we define our color and depth losses as below:

$$\mathcal{L}_C \;=\; \sum_{i \in \mathbf{I}} \left\| \hat{\mathbf{C}}_i - \mathbf{I}_i \right\|_1, \qquad \mathcal{L}_D \;=\; \sum_{i \in \mathbf{I}} \left\| \hat{\mathbf{D}}_i - \mathbf{D}_i \right\|_1. \tag{6}$$

To better distinguish foreground tool objects from background tissue, as inspired from [17], we additionally apply an entropy-based regularization on the rendered alpha values of tool Gaussians, the term is formulated as below :

$$\mathcal{L}_{reg} \;=\; -\sum \Big(\mathbf{O}_{\mathrm{obj}} \log \mathbf{O}_{\mathrm{obj}} \;+\; \big(1 - \mathbf{O}_{\mathrm{obj}}\big) \log\big(1 - \mathbf{O}_{\mathrm{obj}}\big) \Big), \tag{7}$$

where $\mathbf{O}_{\mathrm{obj}}$ is the accumulated opacity of the decomposed foreground. We then optimize the total loss $\mathcal{L} = \lambda_C \mathcal{L}_C + \lambda_D \mathcal{L}_D + \lambda_{reg} \mathcal{L}_{reg}$, jointly refining the canonical Gaussians and the motion models for more accurate reconstruction.

3 Experiments

3.1 Experimental Setting

Datasets and Evaluation. We evaluate our method on EndoNeRF dataset[14] and StereoMIS dataset[4]. EndoNeRF comprises stereo endoscopic videos from robotic prostatectomy cases, featuring surgical tool movement and dynamic tissue deformations. StereoMIS includes in-vivo porcine stereo sequences with large-scale anatomical variations. We conduct experiments on the public "pulling" and "cutting" sequence from EndoNeRF (63/156 frames each), and select 2 clips from StereoMIS containing " deforming bowel" and " breathing liver" (80/68 frames each) and capturing moving surgical tools at the same time. For each scene, we split frames into 7:1 training/testing sets following[14,21,10,16]. We compute PSNR and SSIM to evaluate reconstruction quality, meanwhile, report training time and rendering speed.

Implementation Details. We train our model for 3000 iterations with loss weights for color (λ_C), depth (λ_D), and regularization (λ_{reg}) set to 1.0, 1.0, and 10^{-2}, respectively. All experiments are conducted on a single NVIDIA RTX 2080 with PyTorch.

3.2 Comparison Results

Table 1: Performance comparison of our method with baselines on whole scene reconstruction task and tool reconstruction task. We conduct experiments on both EndoNeRF and StereoMIS datasets. We report metric computed on complete frame, and the starred ($*$) ones computed on tool-area only.

Method	PSNR↑	SSIM↑	PSNR*↑	SSIM*↑	min↓	FPS↑	PSNR*↑	SSIM*↑	min↓	FPS↑
	EndoNeRF: Whole Scene Reconstruction						**Moving Tool Reconstruction**			
ForPlane[18]	32.34	88.52	36.97	97.57	12:24	0.20	36.91	97.51	12:29	0.41
SurgicalGS[16]	32.79	92.85	36.13	97.85	3.46h	27.95	34.44	97.59	3.3h	142.8
EndoG[10]	31.78	91.14	34.52	97.18	3:06	114.7	34.47	97.06	2:44	83.0
Deform3DGS[21]	32.33	91.81	35.07	97.21	**1:10**	**174.0**	35.02	97.10	**1:28**	**200.2**
ETMM(ours)	-	-	-	-	-	-	**36.94**	**97.98**	2:20	71.2
T^{2}GS(ours)	**32.95**	**92.12**	**36.91**	**98.14**	3:36	50.7	-	-	-	-
	StereoMIS: Whole Scene Reconstruction						**Moving Tool Reconstruction**			
ForPlane[18]	30.25	85.03	36.20	98.40	10:08	0.34	40.83	99.09	11:34	0.71
SurgicalGS[16]	**32.48**	**93.37**	36.55	98.59	4.45h	13.61	**37.51**	**98.83**	1.6h	110.7
EndoG[10]	29.94	89.56	34.45	98.25	3:17	85.5	33.07	97.54	3:42	23.9
Deform3DGS[21]	30.48	89.88	34.33	98.23	**1:06**	**174.6**	34.29	97.78	**1:25**	**213.6**
ETMM(ours)	-	-	-	-	-	-	36.04	98.49	1:23	71.4
T^{2}GS(ours)	31.55	90.48	**39.83**	**99.27**	3:08	58.3	-	-	-	-

We compare our proposed method with baseline methods for 1) whole scene reconstruction task (reconstruct both tissue deformation and tool motion as a complete scene) and 2) Moving Tool reconstruction task (tool-only). The baselines include Deform3DGS [21], EndoGaussian [10] and SurgicalGS [16], being previous state of art GS methods for surgical scene reconstruction, which at core

translate varying representative dynamic GS modelling strategies [9,15,22] in surgery. In addition, we also compare with Forplane[18], a state of art NeRF[11] based method for fast surgical scene reconstruction.

For **moving tool reconstruction task**, as reported in the right column of Tab. 1, our tool motion model ETMM outperforms EndoG and Deform3DGS with significant margins in reconstruction quality across both datasets, meanwhile being competitive in training time and rendering speed. As discussed in Sec. 1, this limitation may stem from Deform3DGS [21] adopting FDM for 4D modelling. While highly flexible, this modelling mechanism is prone to overparameterization, making it more susceptible to noise and potentially leading to performance degradation. EndoGaussian's limitation lies on its low rank tensors representation [3], being incorporated as an encoder for accelerated reconstruction[15]. On the other hand, SurgicalGS, an MLP decoder-only 4D GS reconstruction method, significantly outperforms EndoGaussian. Notably, it occasionally outperforms our method but comes at the cost of extensive computational overhead, requiring several hours. In general, NeRF based method Forplane is not comparable with Gaussian based methods in terms of both reconstruction efficiency and quality.

For **whole Scene Reconstruction task**, our T^{2}GS framework achieves the best reconstruction performance, outperforming both EndoNeRF and StereoMIS in terms of PSNR and SSIM. Additional evaluation metrics (PSNR*,SSIM*) highlight its success stems from accurate reconstruction of the moving tools. We also compare T^{2}GS with Deform3DGS on the tissue reconstruction task, namely only reconstructing tissue by masking tools. As indicated by the consistent PSNR and SSIM values computed on tissue-area only in Tab.2, T^{2}GS 's overall reconstruction improvement does not substantially impair tissue reconstruction before/after the incorporation of our motion model. As also visually justified in Fig.2 — T^{2}GS clearly showcases least artifacts in tool area, meanwhile remain intricate tissue reconstruction details. Additionally, we observe our outperformance is more significant on StereoMIS than on EndoNeRF, where the captured surgical tools tend to display stronger motion, justifying the effectiveness of LSM and LSP in our proposed motion model.

3.3 Ablation Study

We conduct ablation studies on our ETMM's key components using the EndoNeRF dataset, focusing on tool motion reconstruction to validate our design. As shown in Tab. 3, initializing tool Gaussians from the sequence's initial frame, as conducted in [5], without PIPF (w/o PIPF) results in degraded performance, underscoring the importance of our pose-informed point fusion (PIPF) strategy for holistic tool initialization. To evaluate pose-guided global trajectory modeling (P-GTM), we initialize the learned transformation with an identity matrix (w/o PoseInit), which eliminates pose guidance and enforces learning from scratch, leading to performance deterioration. Removing local shape modeling (w/o LSM) also causes a performance drop due to incomplete motion modeling of local tool-shape changes. Additionally, we compare our adapted-FDM with

Table 2: Additional comparison. We reported metrics on tissue-area only with denotation(+). The left/right four columns are on EndoNeRF/StereoMIS.

Deform3DGS[21]		T^2GS		Deform3DGS[21]		T^2GS	
$PSNR^+\uparrow$	$SSIM^+\uparrow$	$PSNR^+\uparrow$	$SSIM^+\uparrow$	$PSNR^+\uparrow$	$SSIM^+\uparrow$	$PSNR^+\uparrow$	$SSIM^+\uparrow$
36.53	95.04	35.91	94.69	34.08	92.25	33.64	91.60

Table 3: Ablation on ETMM.

Method	PSNR*↑	SSIM*↑	min↓	FPS ↑
w/o PIPF	34.43	97.31	2:45	61.5
w/o PoseInit	34.24	97.37	2:55	61.3
w/o LSM	33.17	97.02	2:04	**72.3**
w/o Reg	36.54	97.87	**1:48**	57.5
Ours	**36.94**	**97.98**	2:20	71.2

Table 4: Ablation on our adapted-FDM used in LSM.

Method	PSNR*↑	SSIM*↑	min↓	FPS↑
Ours(Poly)	34.32	97.29	2:48	61.6
Ours(FDM)	36.86	97.92	2:56	60.1
Ours	**36.94**	**97.98**	**2:20**	**71.2**

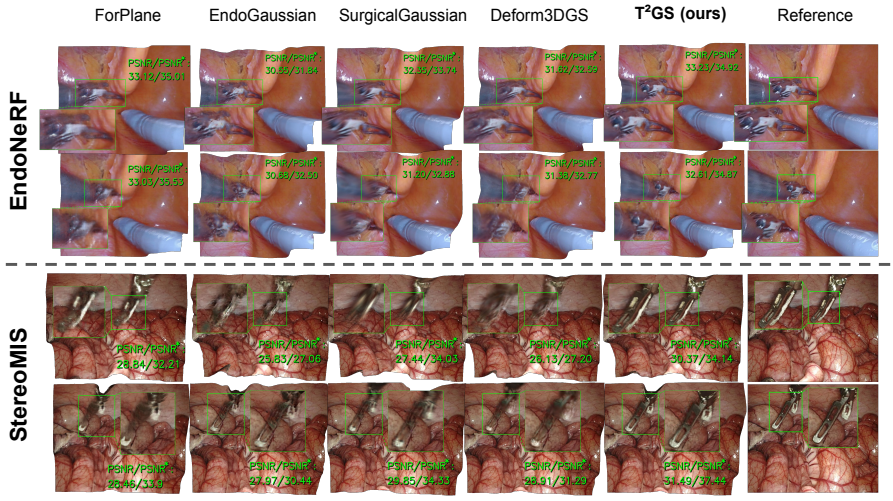


Fig. 2: **Visualisation on samples from EndoNeRF and StereoMIS.** We denote PSNR/PSNR*.

alternatives including FDM (Our-FDM) and FDM with polynomial basis functions (Our-Poly), as shown in Tab. 4. Their sub-optimal results highlight the effectiveness of our design.

4 Conclusion

In conclusion, we introduce T^2GS, a novel framework designed to address the limitations of current methods by holistically modeling the dynamics of both tissues and surgical tools in surgical scenes. As the first Gaussian Splatting-

based framework with specialized designs for surgical tool motion reconstruction, T^{2}GS integrates the Efficient Tool Motion Modeling (ETMM) module, which combines Pose-guided Global Trajectory Modeling (P-GTM) and Local Shape-change Modeling (LSM) to effectively capture the deformation field of surgical tools. Our framework achieves efficient and high-quality reconstruction of comprehensive surgical scenes, outperforming state-of-the-art methods, particularly in tool reconstruction quality. By bridging the gap in surgical tool motion modeling, T^{2}GS sets a new benchmark for dynamic MIS scene reconstruction, offering significant advancements in both efficiency and accuracy.

Acknowledgments. This work was funded by the German Federal Ministry of Education and Research (BMBF, grant 01IS23070, Software Campus 3.0, TUD Dresden, project 'NeuralNodes'), and by the Bavarian and Saxon Ministries for "Next Generation AI Computing (GAIn)". It also received funding from the European Union (CloudSkin project, ID: 101092646), the German Research Foundation (DFG, EXC 2050/1, Project ID 390696704, CeTI), and the BMBF (6G-life project, ID: 16KISK001K). We acknowledge Isabel Funke for her valuable input during the writing.

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

References

1. Aghazadeh, F., Zheng, B., Tavakoli, M., Rouhani, H.: Surgical tooltip motion metrics assessment using virtual marker: An objective approach to skill assessment for minimally invasive surgery. International Journal of Computer Assisted Radiology and Surgery **18**(12), 2191–2202 (2023)
2. Boedecker, C., Huettl, F., Saalfeld, P., Paschold, M., Kneist, W., Baumgart, J., Preim, B., Hansen, C., Lang, H., Huber, T.: Using virtual 3d-models in surgical planning: workflow of an immersive virtual reality application in liver surgery. Langenbeck's archives of surgery **406**, 911–915 (2021)
3. Cao, A., Johnson, J.: Hexplane: A fast representation for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 130–141 (2023)
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**(7), 1185–1192 (2023)
5. Huang, Y., Cui, B., Bai, L., Guo, Z., Xu, M., Islam, M., Ren, H.: Endo-4dgs: Endoscopic monocular scene reconstruction with 4d gaussian splatting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 197–207. Springer (2024)
6. Karaev, N., Rocco, I., Graham, B., Neverova, N., Vedaldi, A., Rupprecht, C.: Cotracker: It is better to track together. In: European Conference on Computer Vision. pp. 18–35. Springer (2024)
7. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)

8. Lamarca, J., Parashar, S., Bartoli, A., Montiel, J.: Defslam: Tracking and mapping of deforming scenes from monocular sequences. IEEE Transactions on robotics **37**(1), 291–303 (2020)
9. Lin, Y., Dai, Z., Zhu, S., Yao, Y.: Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21136–21145 (2024)
10. Liu, Y., Li, C., Yang, C., Yuan, Y.: Endogaussian: Gaussian splatting for deformable surgical scene reconstruction. arXiv e-prints pp. arXiv–2401 (2024)
11. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
12. Rodby, K.A., Turin, S., Jacobs, R.J., Cruz, J.F., Hassid, V.J., Kolokythas, A., Antony, A.K.: Advances in oncologic head and neck reconstruction: systematic review and future considerations of virtual surgical planning and computer aided design/computer aided modeling. Journal of Plastic, Reconstructive & Aesthetic Surgery **67**(9), 1171–1185 (2014)
13. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016)
14. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: 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)
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. pp. 20310–20320 (2024)
16. Xie, W., Yao, J., Cao, X., Lin, Q., Tang, Z., Dong, X., Guo, X.: Surgicalgaussian: Deformable 3d gaussians for high-fidelity surgical scene reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 617–627. Springer (2024)
17. Yan, Y., Lin, H., Zhou, C., Wang, W., Sun, H., Zhan, K., Lang, X., Zhou, X., Peng, S.: Street gaussians: Modeling dynamic urban scenes with gaussian splatting. In: European Conference on Computer Vision. pp. 156–173. Springer (2024)
18. Yang, C., Wang, K., Wang, Y., Dou, Q., Yang, X., Shen, W.: Efficient deformable tissue reconstruction via orthogonal neural plane. IEEE Transactions on Medical Imaging (2024)
19. Yang, C., Wang, K., Wang, Y., Yang, X., Shen, W.: 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)
20. Yang, J.H., Goodman, E.D., Dawes, A.J., Gahagan, J.V., Esquivel, M.M., Liebert, C.A., Kin, C., Yeung, S., Gurland, B.H.: Using ai and computer vision to analyze technical proficiency in robotic surgery. Surgical Endoscopy **37**(4), 3010–3017 (2023)
21. 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. pp. 132–142. Springer (2024)

22. Yang, Z., Gao, X., Zhou, W., Jiao, S., Zhang, Y., Jin, X.: Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20331–20341 (2024)
23. 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. pp. 13–23. Springer (2023)