Occlusion-free 4D Gaussians for Open Surgery Videos Using Multi-Camera Shadowless Lamps

Yuna Kato[1], Shohei Mori[2,1], Hideo Saito[1], Yoshifumi Takatsume[1], Hiroki Kajita[1], and Mariko Isogawa[1]

[1] Keio University
[2] University of Stuttgart
{yu01-na10,mariko.isogawa}@keio.jp

Abstract. Video recording of open surgery is in great demand for education and research purposes but is challenging due to the busy and dynamic environment. The state-of-the-art system uses multi-view cameras installed in shadowless lamps (McSL) and implements an automatic camera switching algorithm to avoid disturbances. However, this algorithm leads to missing pixels and distorted projection due to mathematical image warping and does not always provide the best perspective. We propose using 4D Gaussian Splatting (4DGS) to create editable 3D videos and remove Gaussians occluding surgical fields from a perspective. We enable occlusion-free 3D videos by addressing two occlusion removal approaches via (1) occlusion masking and (2) density-based Gaussian filtering. We create a real-surgery dataset and demonstrate that our method outperforms the state-of-the-art auto view-switching approach.

Keywords: Surgical Video Synthesis · 4D Gaussians · Open Surgery.

1 Introduction

Recording open surgeries as videos can provide objective visual information in addition to conventional medical records. While videos are acknowledged for their value in knowledge transfer and research [5, 9], configuring a method to capture video from the desired angle without obstructing is challenging. Endoscopic and robotic surgeries offer egocentric views with minimal obstructions [22]. However, room-installed cameras in open surgeries provide only fixed viewpoints and suffer from potential disturbances (e.g., surgeons) [10].

Egocentric cameras are a standard approach but suffer from frequent and drastic motions, causing unclear and unstable vision [16, 18]. More recent approaches use either a single camera [3] or multiple cameras [25] on a shadowless lamp to record views directly above surgical fields of interest. The multi-camera shadowless lamp (McSL) setup increases the chance of capturing a surgical field of interest with any cameras. Toward occlusion-free video generation using McSL, this unique setup leads to two challenges: automatic multi-camera calibration [11] and occlusion avoidance [11, 25].

In this paper, we address occlusion removal rather than avoidance for dynamic videos [11]. Conventional approaches evaluate individual views for the least occlusions and concatenate selected views for a longer video. However, their 2D image warping for stable view switching results in distorted views and missing pixels around image borders. Instead, we propose using 4D Gaussian Splatting (4DGS) [28] to create fully editable 3D videos and remove Gaussians occluding those of surgical fields from a perspective. We enable occlusion-free 3D videos by addressing two occlusion removal approaches for 4DGS via (1) occlusion masking and (2) density-based Gaussian filtering. As a byproduct of 4DGS, our open surgery videos are in 3D, allowing changing viewpoints around McSL.

To validate our approaches, we create a dataset of videos with and without occlusion. We propose extracting obstacles in a video and synthesizing them into other parts of the video with no occlusions. We use the original and head-synthesized videos as ground truth and input videos, respectively. This approach creates a spatially and temporally consistent evaluation dataset[3].

The contributions of this paper are summarized as follows:

- We propose a 3D video generation pipeline for open surgery with McSL using 4D Gaussian Splatting. This enables both occlusion removal and surgical field visualization from any viewpoint.
- To this end, we introduce occlusion removal modules that eliminate occluded regions at the level of gaussian splats.
- To evaluate our approach for the new task, we created a synthetic dataset.
- We report the results of an expert review study and discuss the challenges of our method and its variants.

2 Related Work

2.1 Surgical Video Recording and Processing

Robotics [4] and virtual/augmented reality (VR/AR) technology [6,13] are known to offer more effective and efficient education than traditional ones, while they have been facing challenges in costs and accessibility. As such, video recording, meanwhile, has been a practical and widely accepted approach.

Head-attached cameras are the most common configuration, but mobile cameras move too fast and have limited battery life [16, 18, 21, 27]. Fixed and wired solutions allow stable recording with enough electricity for hours of recordings [14, 19]. However, positioning is critical to avoid significant interferences during the surgery, which is only feasible with a dedicated operator in reality.

Shimizu et al. developed McSL to directly record open surgeries as the original shadowless lamps are intended [25]. A follow-up work implements automatic calibration and view-switching to reduce the burden for manual view alignment and occlusion avoidance in the original work [11]. Contrary to this 2D approach, we reconstruct 4D (i.e., spatiotemporal Gaussians) of open surgery and remove occlusions from the representation.

[3] project page: https://isogawa.ics.keio.ac.jp/projects/4DGS-McSL

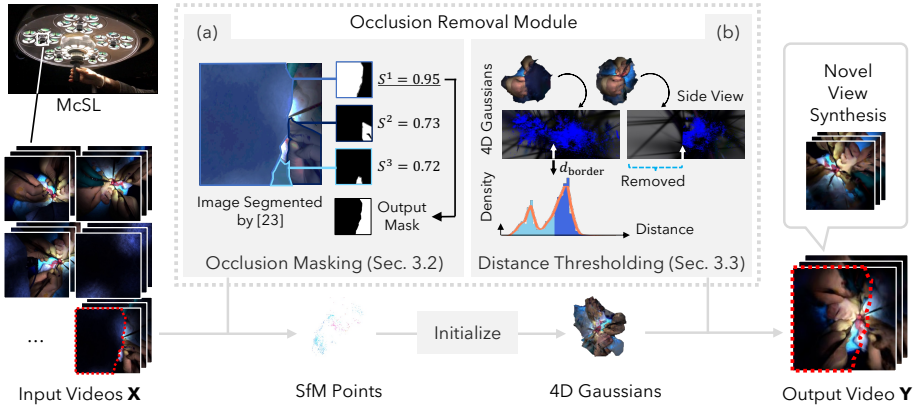

Fig. 1: Overview of our occlusion-free 4D Gaussian generation. Our major contributions lie in the two occlusion removal modules (a, b). The outcome is a 3D video without occlusions (e.g., the head in this figure) and free viewpoint control.

2.2 3D View Synthesis in Medical Domain

3D reconstruction has been actively studied in endoscopic surgery due to the ready-to-use video inputs [2, 26]. Neural rendering [17] allowed researchers to revisit novel view synthesis for improved visual fidelity [1]. More recent 3D Gaussian representation [12] gained rapid attention due to its high-performance rendering in this domain [29]. 4DGS [28] lifted up the application range to the temporal domain [15].

Open surgery 3D video recording was more challenging than endoscopic surgery due to the lack of stable multi-view video resources. With McSL, 3D view synthesis became feasible but was limited to static scenes [20]. Similarly to 4DGS in endoscopic surgery, we leverage 4DGS scene representation for open surgery with McSL video resources but further address its unique challenge of frequent occlusions by surgeons.

3 Method

The proposed method reconstructs surgical scenes in 3D over time and incorporates two different occlusion removal methods: one removes occlusions based on masking (Section 3.2) and the other removes them at the 3D splat level (Section 3.3). We discuss advantages and currently observed limitations, leading to the evaluations on which to select and combine.

3.1 Overview

We record an open surgery as video $\mathbf{X}$ consisting of N_{img} frames from N_{cam} ($= 5$ by default) cameras in the McSL and generate a rendered video $\mathbf{Y}$ of N_{img} frames

from a desired perspective. In summary, $\mathbf{X} = \{X_t^c \mid c \in [1, N_{\text{cam}}],\ t \in [1, N_{\text{img}}]\}$ and $\mathbf{Y} = \{Y_t \mid t \in [1, N_{\text{img}}]\}$ (Fig. 1).

We approach the task of rendering $\mathbf{Y}$ by reconstructing the surgical scene in 4D (xyz-t) space using Gaussian splatting representation and then synthesizing the reconstruction result from the desired viewpoint. We reconstruct our base 4D Gaussians using the state-of-the-art 4DGS approach [28]. 4DGS uses a canonical 3DGS and displaces it for observed dynamics. The original 4DGS approach uses structure-from-motion (SfM) points with SIFT features. Previous works observed McSL calibration failing due to the lack of distinct features in surgical fields and heavy occlusions [11, 20]. Therefore, we use data-driven feature point detection and matching methods, SuperPoint [7] and Superglue [24], respectively, following the best practice in the literature.

3.2 Occlusion Removal via Occlusion Masking

We remove occluding objects by excluding detected mask pixels from 4DGS optimization, aiming to mask out the major disturbances such as the surgeon's head. We use Segment-Anything-2 (SAM2) [23] to segment regions to automate the masking process for all multi-view video inputs. Since the segmented regions do not have any labels, we identify the regions of occluding objects by quantifying occlusion-specific characteristics based on four criteria:

- S_{C} evaluates how convex a segment is. Occlusions such as surgen's head appear mostly round or convex. To measure this, we calculate the intersection over union (IoU) of a segment and its convex hull.
- S_{E} evaluates how much a segment touches the image edges. Our observation is that occluders tend to obscure the edges of camera images since they appear close to McSL. We then calculate $S_{\text{E}} = \min(1, l_{edge}/l_{contour})$ is the length of the occluder's contour, and $l_{contour}$ and l_{edge} is the length of the occluder's contour that touches the image edges, respectively.
- S_{I} evaluates the darkness of a segment. Occlusions appear close to McSL and thus darker than the other areas. We calculate the ratio of the average intensity of a segment over that of the entire image.
- S_{H} is designed to evaluate (i) how close the color is to the target color and (ii) how small the standard deviation of the occluder's color is. This is based on the observation that occluders are often surgen's neck, or head wearing surgical caps, which typically have a roughly known color in advance and are often composed of a single color. Here, the Hue component in the HSV color space is used for robustness, We calculate

$$S_{\text{T}} = \max\left(0, 1 - \frac{\text{std}(c_{seg})}{90}\right)\left(1 - \frac{\text{diff}(\text{ave}(c_{seg}) - c_{targ})}{90}\right), \tag{1}$$

  where c_{seg} and c_{targ} are a set of colors at each pixel in the segmented region and the target color, respectively. std$(\cdot)$, ave$(\cdot)$, and diff$(\cdot)$ represent standard deviation, average, and difference function, respectively. There are two differences in Hue values, we choose the smaller value in absolute differences.

We calculate the final score, S, with a linear combination, $\sum_{k\in\{\mathrm{C,E,I,H}\}} w_k S_k$. We identify an occluder segment where S exceeds a threshold (Fig. 1a). We enlarge the regions to prevent noisy reconstruction around the silhouette.

3.3 Occlusion Removal via Distance Distribution Analysis

When a surgical field is occluded by surgeons, two clusters (i.e., surgical field and occluder) of 3D Gaussians are expected to appear. Given N_{cam} cameras, we approximate their locations by a plane and calculate distances to individual Gaussians. These distance values typically exhibit single peaks in non-occluded scenes and binomial peaks in occluded scenes (Fig. 1b).

We assign the distance values into bins and perform kernel density estimation to approximate the distribution. We identify two distinct peaks that are sufficiently separated. We then determine d_{border} at the midpoint between the farther peak and minimum density between the two peaks to remove all Gaussians that fall below this distance.

4 Experiments

We validate our method via quantitative evaluation using our dataset of real surgeries (Section 4.2) and expert review (Section 4.3).
Hardware. We used a machine with an Intel(R) Xeon(R) w5-3435X with 64 GB RAM running on Ubuntu 22.04 LTS OS, and NVIDIA RTX A6000 48 GB.
Parameters. $w_\alpha = 0.250$, $w_\beta = 0.500$, $w_\gamma = 0.125$, and $w_\delta = 0.125$, and the threshold for classifying the calculated scores was set to 0.85 (Section 3.2).
Contendor. We compare our method with the state-of-the-art view-switching method (`Switch`) [11]. While this approach finds non-occluded frames, we remove objects using the masking and distance thresholding approach (`Ours:Mask` and `Ours:Dist`, respectively) and distance thresholding (`Ours:Mask+Dist`).

4.1 Dataset Generation

To assess our method in various surgery types, we used seven surgical videos: #1 polysyndactyly, #2 external nasal deformity, #3 intramuscular lipoma of the gluteus maximus, #4 Frontal bone fracture, #5 preauricular sinus, #6 scalp scar revision and hair transplantation, and #7 keloid on the anterior chest.

There are no pairs of videos with and without occlusions in the original dataset [11], and no ground truth for our occlusion removal task is provided for quantitative evaluations. Therefore, we generated (c) input videos with occlusions (a) by extracting occluded areas in the same video and superimposing them over (b) the ground truth video without occlusion, as in Fig. 2. We prepared 15 and ten videos (300 frames each) from #1 and #7, respectively. We chose these two scenes because they recorded enough unoccluded videos, which are suitable for ground truth and input compositions.

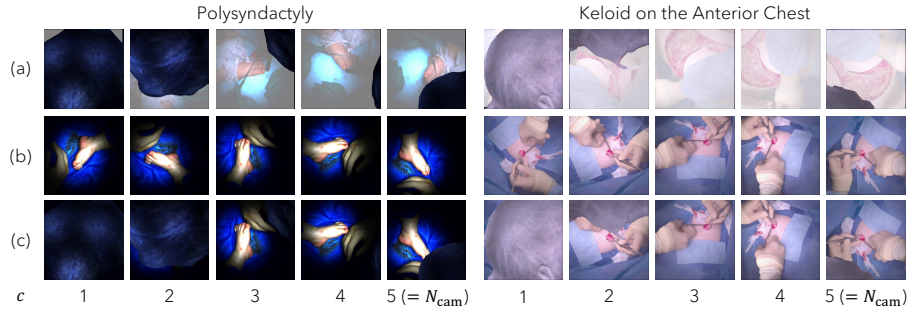

Fig. 2: Creating input and ground truth pairs. We extract (a) occlusion regions and overlay them to (b) frames without occlusions in the same scene to create (c) virtual input frames. Therefore, (b) is the ground truth of (c).

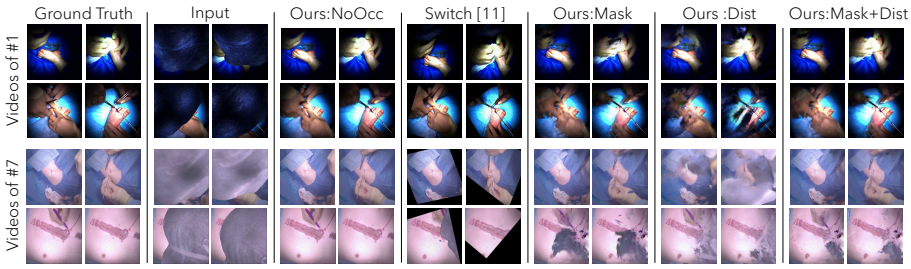

Fig. 3: Comparisons between `Switch` [11] and our method.

4.2 Quantitative Evaluation

Evaluation Metrics. We used standard image metrics including PSNR, SSIM, and LPIPS [30] and also a video metric, AvSpeed [8]. AvSpeed represents the average motion speed of image feature points. A smaller value indicates that the video is temporally smoother. Given the total number of image frames N_f and the number of feature points per frame N_p, AvSpeed is computed as

$$\text{AvSpeed} = \frac{1}{N_p(N_f-1)} \sum_{i=1}^{N_p} \sum_{t=1}^{N_f-1} \|\dot{z}_i(t)\|. \tag{2}$$

Here, the velocity $\dot{z}_i(t)$ is defined as follows, where $z_i(t)$ denotes the position of the i-th feature point in image coordinates:

$$\dot{z}_i(t) = z_i(t+1) - z_i(t). \tag{3}$$

Results. Example results using each method with the composited videos are shown in Fig. 3. Table 1 summarizes the average and standard deviations in individual scores. To highlight the influence of the masking accuracy of Segment-Anything-2 [23], our segment selection criteria, and supportiveness of `Mask+Dist`,

Table 1: Quantitative evaluation of occluded scenes. The best and the second best results are colored by orange and pink, respectively.

Mask Generation Accuracy	Method	PSNR [dB] (↑)	SSIM (↑)	L-PIPS (↓)	AvSpeed [px/frame] (↓)
≥ 98%	`Ours:NoOcc`	31.7 (±2.8)	0.924 (±0.022)	0.160 (±0.072)	2.51 (±2.63)
	`Switch` [11]	16.0 (±4.3)	0.709 (±0.091)	0.380 (±0.097)	3.64 (±3.31)
	`Ours:Mask`	22.7 (±3.1)	0.830 (±0.060)	0.273 (±0.087)	2.79 (±2.67)
	`Ours:Dist`	17.4 (±4.1)	0.763 (±0.078)	0.364 (±0.090)	2.94 (±2.30)
	`Ours:Mask+Dist`	22.0 (±3.9)	0.818 (±0.070)	0.292 (±0.098)	2.53 (±2.47)
< 98%	`Ours:NoOcc`	32.0 (±3.3)	0.925 (±0.029)	0.196 (±0.064)	2.18 (±1.95)
	`Switch` [11]	19.0 (±0.8)	0.807 (±0.067)	0.288 (±0.051)	5.50 (±2.75)
	`Ours:Mask`	16.6 (±1.90)	0.811 (±0.052)	0.385 (±0.043)	3.24 (±1.52)
	`Ours:Dist`	19.1 (±2.4)	0.823 (±0.057)	0.367 (±0.071	2.65 (±2.06)
	`Ours:Mask+Dist`	20.1 (±1.3)	0.824 (±0.052)	0.339 (±0.035)	2.55 (±1.72)

we calculate individual scores depending on the success rates of the masking. The success rate of mask generation was calculated based on the number of matching pixels compared to the occlusion mask used for creating the composited video. The average mask generation accuracy across all 25 scenes was 98%. Therefore, we grouped 20 scenes with the accuracy $\geq 98\%$ and the rest with $< 98\%$. We performed 4DGS on the ground truth videos to confirm the possibly highest performance (`Ours:NoOcc`).

Table 1 shows that our method outperforms the contender while it suffers from inaccurate masks. We observed wrongly segmented head covers that unintentionally included neck and mask straps. In such cases, `Mask+Dist` supplementary removes the frontal cluster of Gaussians, leading to higher performance. The evaluation scores of `Ours:Dist` are relatively low for $\geq 98\%$. This is due to the fact that Gaussians represent scenes using overlapping ellipsoids, which leads to ambiguity in the boundaries between occluders and the background. Consequently, combining both (`Ours:Mask+Dist`) can stably achieve good scores over different conditions. `Switch` [11] can offer clear but distorted views due to plane approximation of Homography warping from a far camera.

`Ours` performs better in the video metric, AvSpeed, than `Switch` regardless the mask generation accuracy. `Switch` applied view switching that resulted in different warping and skewing over time. The missing pixels around the warped views also frequently changed. Overall, the quality of the output videos varied.

4.3 Expert Review

Participants. We collected five medical doctors (D1–D5, all male, age AVG= 42.5, SD= 11.0 years old, performing surgeries regularly) from a medical school. **Visual stimuli.** We performed `Switch` [11], `Ours:Mask`, and `Ours:Mask+Dist` for videos #1–#6 (Fig. 4). To display all videos per method within a screen, we used the first six videos. We excluded `Ours:Dist` due to its low fidelity in the quantitative evaluation. All videos have 300 frames (=10 s), and occlusions occur at least once. Fig. 4 shows example frames. We generated videos with rotating

Fig. 4: Videos for expert review.

cameras to show the motion parallax in our occlusion-free 4DGS. We provide these videos as supplemental material.

Apparatus. We prepared a web page showing videos of three methods in a randomized order. We showed six videos of each method at once per page. For `Ours`, we presented videos with a rotating camera to show the motion parallax.

Task. The participants were asked to provide comments addressing the strengths and weaknesses of individual methods after watching videos of each method.

Comments on `Switch`. The comment was positive, with consistent viewpoints and minimal stress during view-switching (D2, D4, D5). The highest visibility was acknowledged, enhancing the viewing experience (D1, D5). However, some limitations were noted, including occasional visible blurs (D1), missing regions all around the warped view (D4), temporary loss of points of interest (D3, D4), and significant occlusion (when all cameras were blocked) (D2).

Comments on `Ours`. The free viewpoint changes were well-received (D2, D4). D2 specifically noting its usefulness for viewing surgery from different angles. The fewer occlusions than `Switch` were also appreciated (D3). However, some concerns raised, including a lack of clear vision (D2, D3, D4, D5) and occasionally missing surgical instruments (D1).

`Ours:Mask+Dist` was considered superior to `Ours:Mask` due to its cleaner vision (D3, D4). However, some users could not discern a difference between the two (D2, D5). D4 expressed a desire to control the viewpoint, and D5 highlighted the high educational value of reproducing videos from the surgeon's perspective.

5 Conclusion

We proposed occlusion-free 4DGS using McSL to synthesize open surgery videos without visual disturbance at arbitrary viewpoints. We evaluated our method by comparing it with the state-of-the-art auto view-switching method. To this end, we created a new dataset for this unique task. Further, we had an expert review for explanatory investigation of limitations and future challenges.

We aim to improve the rendering fidelity further using prior knowledge, such as hand and instrument models. We are also interested in providing our occlusion-free 4DGS in VR headsets to investigate its educational value.

Ethical approval. Approval for open surgery video recording was obtained from the ethics committee of Keio University under 20180111. Informed consent was obtained from all individual participants included in the study.

Acknowledgments. This work was partially supported by JSPS KAKENHI Grant Number 22H03617, 25K03143, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2120/1 – 390831618.

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

References

1. Batlle, V.M., Montiel, J.M., Fua, P., Tardós, J.D.: Lightneus: Neural surface reconstruction in endoscopy using illumination decline. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 502–512 (2023)
2. Brandao, P., Psychogyios, D., Mazomenos, E., Stoyanov, D., Janatka, M.: Hapnet: Hierarchically aggregated pyramid network for real-time stereo matching. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **9**(3), 219–224 (2021)
3. Byrd, R.J., Ujjin, V.M., Kongchan, S.S., Reed, H.D.: Surgical lighting system with integrated digital video camera. US6633328B1 (2003)
4. Chen, R., Rodrigues Armijo, P., Krause, C., Force, S.R.T., Siu, K.C., Oleynikov, D.: A comprehensive review of robotic surgery curriculum and training for residents, fellows, and postgraduate surgical education. Surgical Endoscopy **34**, 361–367 (2020)
5. Date, I.: Ns now updated no.9 thorough knowledge and application of device and information technology (it) for neurosurgical opperation (2017)
6. Debes, A.J., Aggarwal, R., Balasundaram, I., Jacobsen, M.B.: A tale of two trainers: virtual reality versus a video trainer for acquisition of basic laparoscopic skills. The American Journal of Surgery **199**(6), 840–845 (2010)
7. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 224–236 (2018)

8. Guilluy, W., Beghdadi, A., Oudre, L.: A performance evaluation framework for video stabilization methods. In: 2018 7th European Workshop on Visual Information Processing (EUVIP). pp. 1–6 (2018)
9. Hanada, E.: [special talk] video recording, storing, distributing and editing system for surgical operation. In: ITE Technical Report. pp. 77–80 (2017)
10. Kajita, H.: Surgical video recording and application of deep learning for open surgery. Journal of Japan Society of Computer Aided Surgery **23**(2), 59–64 (2021)
11. Kato, Y., Isogawa, M., Mori, S., Saito, H., Kajita, H., Takatsume, Y.: High-quality virtual single-viewpoint surgical video: Geometric autocalibration of multiple cameras in surgical lights. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 271–280 (2023)
12. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (2023)
13. Kobayashi, L., Zhang, X.C., Collins, S.A., Karim, N., Merck, D.L.: Exploratory application of augmented reality/mixed reality devices for acute care procedure training. Western Journal of Emergency Medicine **19**(1), 158 (2017)
14. Kumar, A.S., Pal, H.: Digital video recording of cardiac surgical procedures. The Annals of Thoracic Surgery **77**(3), 1063–1065 (2004)
15. Liu, H., Liu, Y., Li, C., Li, W., Yuan, Y.: Lgs: A light-weight 4d gaussian splatting for efficient surgical scene reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 660–670 (2024)
16. MC, A.: Intraoperative video production with a head-mounted consumer video camera. Journal of Orthopaedic Trauma **31**, S2–S3 (2017)
17. 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)
18. Nair, A.G., Kamal, S., Dave, T.V., Mishra, K., Reddy, H.S., Della Rocca, D., Della Rocca, R.C., Andron, A., Jain, V.: Surgeon point-of-view recording: Using a high-definition head-mounted video camera in the operating room. Indian journal of ophthalmology **63**(10), 771–774 (2015)
19. Nair, A.G., Kamal, S., Singh, S.: A "flexible tripod" mounted video camera an economical and effective method to record oculoplastic surgeries. Indian Journal of Ophthalmology **67**(9), 1460–1462 (2019)
20. Obayashi, M., Mori, S., Saito, H., Kajita, H., Takatsume, Y.: Multi-view surgical camera calibration with none-feature-rich video frames: Toward 3D surgery playback. Applied Sciences **13**(4) (2023)
21. Ortensi, A., Panunzi, A., Trombetta, S., Cattaneo, A., Sorrenti, S., D'Orazi, V.: Advancement of thyroid surgery video recording: A comparison between two full hd head mounted video cameras. International Journal of Surgery **41**, S65–S69 (2017)
22. Preminger, G.M., Delvecchio, F.C., Birnbach., J.M.: Digital image recording: an integral aspect of video endoscopy. Medicine Meets Virtual Reality **62**, 268–274 (1999)
23. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024), `https://arxiv.org/abs/2408.00714`
24. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4938–4947 (2020)

25. Shimizu, T., Oishi, K., Hachiuma, R., Kajita, H., Takatsume, Y., Saito, H.: Surgery recording without occlusions by multi-view surgical videos. In: VISIGRAPP (5: VISAPP). pp. 837–844 (2020)
26. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: Dynamic reconstruction of deformable soft-tissue with stereo scope in minimal invasive surgery. IEEE Robotics and Automation Letters **3**(1), 155–162 (2017)
27. Wei, N.J., Dougherty, B., Myers, A., Badawy, S.M.: Using google glass in surgical settings: systematic review. JMIR mHealth and uHealth **6**(3), e9409 (2018)
28. 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)
29. 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)
30. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)