This version of the contribution has been accepted for publication at MICCAI 2023, after peer review, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/[insert DOI].
Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms.

LightNeuS: Neural Surface Reconstruction in Endoscopy using Illumination Decline

Víctor M. Batlle[1], José M. M. Montiel[1], Pascal Fua[2], and Juan D. Tardós[1]

[1] Inst. Investigación en Ingeniería de Aragón, I3A, Universidad de Zaragoza, Spain
{vmbatlle,josemari,tardos}@unizar.es
[2] CVLab, École Polytechnique Fédérale de Lausanne, Switzerland
pascal.fua@epfl.ch

Abstract. We propose a new approach to 3D reconstruction from sequences of images acquired by monocular endoscopes. It is based on two key insights. First, endoluminal cavities are watertight, a property naturally enforced by modeling them in terms of a signed distance function. Second, the scene illumination is variable. It comes from the endoscope's light sources and decays with the inverse of the squared distance to the surface. To exploit these insights, we build on NeuS [25], a neural implicit surface reconstruction technique with an outstanding capability to learn appearance and a SDF surface model from multiple views, but currently limited to scenes with static illumination. To remove this limitation and exploit the relation between pixel brightness and depth, we modify the NeuS architecture to explicitly account for it and introduce a calibrated photometric model of the endoscope's camera and light source.
Our method is the first one to produce watertight reconstructions of whole colon sections. We demonstrate excellent accuracy on phantom imagery. Remarkably, the watertight prior combined with illumination decline, allows to complete the reconstruction of unseen portions of the surface with acceptable accuracy, paving the way to automatic quality assessment of cancer screening explorations, measuring the global percentage of observed mucosa.

Keywords: Reconstruction · Photometric multi-view · Endoscopy

1 Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and is the second most common cause of cancer death [23]. Early detection is crucial for a good prognosis. Despite the existence of other techniques, such as virtual colonoscopy (VC), optical colonoscopy (OC) remains the gold standard for colonoscopy screening and the removal of precursor lesions. Unfortunately, we do not yet have the ability to reconstruct densely the 3D shape of large sections of the colon. This would usher exciting new developments, such as post-intervention diagnosis, measuring polyps and stenosis, and automatically evaluating exploration thoroughness in terms of the surface percentage that has been observed.

This is the problem we address here. It has been shown that the colon 3D shape can be estimated from single images acquired during human colonoscopies [3]. However, to model large sections of it while increasing the reconstruction accuracy, multiple images must be used. As most endoscopes contain a single camera, the natural way to do this is to use video sequences acquired by these cameras in the manner of structure-from-motion algorithms. An important first step in that direction is to register the images from the sequences. This can now be done reliably using either batch [21] or SLAM techniques [8]. Unfortunately, this solves only half the problem because these techniques provide very sparse reconstructions and going from there to dense ones remains an open problem. And occlusions, specularities, varying albedos, and specificities of endoscopic lighting make it a challenging one.

To overcome these difficulties, we rely on two properties of endoscopic images:

- Endoluminal cavities such as the gastrointestinal tract, and in particular the human colon, are watertight surfaces. To account for this, we represent its surface in terms of a signed distance function (SDF), which by its very nature presents continuous watertight surfaces.
- In endoscopy the light source is co-located with the camera. It illuminates a dark scene and is always close to the surface. As a result, the irradiance decreases rapidly with distance t from camera to surface; more specifically it is a function of $1/t^2$. In other words, there is a strong correlation between light and depth, which remains unexploited to date.

To take advantage of these specificities, we build on the success of Neural implicit Surfaces (NeuS) [25] that have been shown to be highly effective at deriving surface 3D models from sets of registered images. As the Neural Radiance Fields (NeRFs) [15] that inspired them, they were designed to operate on regular images taken around a scene, sampling fairly regularly the set of possible viewing directions. Furthermore, the lighting is assumed to be static and distant so that the brightness of a pixel and its distance to the camera are unrelated. Unfortunately, none of these conditions hold in endoscopies. The camera is inside a cavity (in the colon, a roughly cylindrical tunnel) that limits viewing directions. The light source is co-located with the camera and close to the surface, which results in a strong correlation between pixel brightness and distance to the camera. In this paper, we show that, far from being a handicap, this correlation is a key information for neural network self-supervision.

NeuS training selects a pixel from an image and samples points along its projecting ray. However, the network is agnostic to the sampling distance. In LightNeuS, we explicitly feed to the renderer the distance of each one of these sampled points to the light source, as shown in Fig. 1. Hence, the renderer can exploit the inverse-square illumination decline. We also introduce and calibrate a photometric model for the endoscope light and camera, so that the inverse square law discussed above actually holds. Together, these two changes make the minimization problem better posed and the automatic depth estimation more reliable.

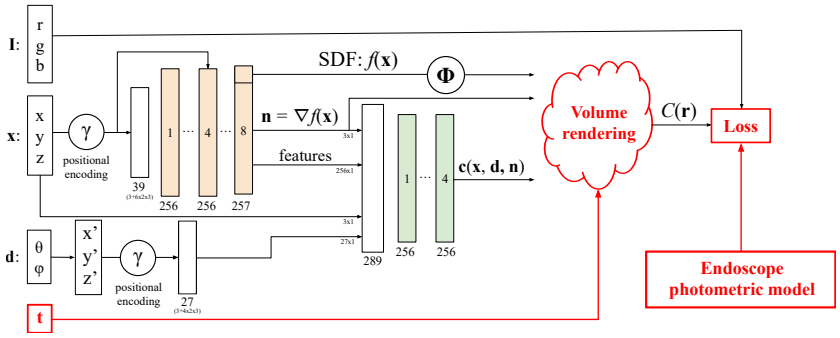

Fig. 1. From NeuS to LightNeuS. The original NeuS architecture is depicted by the black arrows. In LightNeuS, when training the network with a sampled point, we provide the sampling distance t to the renderer, that takes into account illumination decline. We also incorporate a calibrated photometric endoscope model that is used to correctly compute the photometric loss. The changes are shown in red.

Our results show that exploiting the illumination is key to unlocking implicit neural surface reconstruction in endoscopy. It delivers accuracies in the range of 3 mm, whereas an unmodified NeuS is either 5 times less accurate or even fails to reconstruct any surface at all. Earlier methods [3] have reported similar accuracies but only on very few synthetic images and on short sections of the colon. By contrast, we can handle much longer ones and provide a broad evaluation in a real dataset (C3VD) over multiple sequences. This makes us the first to show accurate results of extended 3D watertight surfaces from monocular endoscopy images.

2 Related Works

3D Reconstruction from Endoscopic Images. It can help with the effective localization of lesions, such as polyps and adenomas, by providing a complete representation of the observed surface. Unfortunately, many state-of the-art SLAM techniques based on feature matching [5] or direct methods [7, 6] are impractical for dense endoscopic reconstruction due to the lack of texture and the inconsistent lighting that moves along with the camera. Nevertheless, sparse reconstructions by classical Structure-from-Motion (SfM) algorithms can be good starting points for refinement and densification based on Shape-from-Shading (SfS) [24, 28]. However, classical multi-view and SfS methods require strong suboptimal priors on colon surface shape and reflectance.

In monocular dense reconstructions, it is common practice to encode shape priors in terms of smooth rigid surfaces [17, 20, 14]. Recently, [22] proposes a tubular topology prior for NRSfM aimed to process endoluminal cavities where these tubular shapes are prevalent. In contrast, for the same environments, we propose the watertight prior coded by implicit SDF representations.

Recent methods for dense reconstruction rely on neural networks to predict per-pixel depth in the 2D space of each image and fuse the depth maps by using multi-view stereo (MVS) [2] or a SLAM pipeline [12, 13]. However, holes in the reconstruction appear due to failures in triangulation and inaccurate depth estimation or in areas not observed in any image. Wang et al. [27] show the potential of neural rendering in reconstruction from medical images, although they use a binocular static camera with fixed light source, which is not feasible in endoluminal endoscopy. Unfortunately, most of the previous 3D methods do not provide code [14, 22], are not evaluated in biomedical settings [17, 20], or do not report reconstruction accuracy [12, 13].

Neural Radiance Fields (NeRFs) were first proposed to reconstruct novel views of non-Lambertian objects [15]. This method provides an *implicit neural representation* of a scene in terms of local densities and associated colors. In effect, the scene representation is stored in the weights of a neural network, usually a multilayer perceptron (MLP), that learns its shape and reflectance for any coordinate and viewing direction. NeRFs use volume rendering [9], based on ray-tracing from multiple camera positions. The volume density $\sigma(\mathbf{x})$ can be interpreted as the differential probability of a ray terminating at an infinitesimal particle at location $\mathbf{x}$. The expected color $C(\mathbf{r})$ of the pixel with camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ is the integration of the radiance emitted by the field at every traveled distance t from near to far bounds t_n and t_f, such that

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\, \sigma(\mathbf{r}(t))\, \mathbf{c}(\mathbf{r}(t), \mathbf{d})\, \mathrm{d}t \quad \text{where } T(t) = \exp\left(- \int_{t_n}^{t} \sigma(\mathbf{r}(s))\, \mathrm{d}s \right) \tag{1}$$

where $\mathbf{c}$ stands for the color. The function T denotes the accumulated transmittance along the ray from t_n to t, that is the probability that the ray travels from t_n to t without hitting any other particle. The authors propose two MLPs to estimate the volume density function $\sigma : \mathbf{x} \to [0, 1]$ and the directional emitted color function $\mathbf{c} : (\mathbf{x}, \mathbf{d}) \to [0, 1]^3$, so the density of a point does not depend on the viewing direction $\mathbf{d}$, but the color does. This allows them to model non-Lambertian reflectance. In addition, they propose a positional encoding for location $\mathbf{x}$ and direction $\mathbf{d}$, which allows high-frequency details in the reconstruction.

Neural Implicit Surfaces (NeuS) were introduced in [25] to improve the quality of NeRF representation modelling watertight surfaces. For that, the volume density σ is computed so as to be maximal at the zero-crossings of a signed distance function (SDF) f:

$$\sigma(\mathbf{r}(t)) = \max\left(\frac{\Phi_s'(f(\mathbf{r}(t)))}{\Phi_s(f(\mathbf{r}(t)))}, 0 \right) \quad \text{where } \Phi_s(x) = \frac{1}{1 + e^{-sx}} \tag{2}$$

The SDF formulation makes it possible to estimate the surface normal as $\mathbf{n} = \nabla f(\mathbf{x})$. The reflectance of a material is usually determined as a function of the incoming and outgoing light directions with respect to the surface normal. Therefore, the normal is added as an input to the MLP that estimates color $\mathbf{c} : (\mathbf{x}, \mathbf{d}, \mathbf{n})$, as shown in Fig. 1.

3 LightNeuS

In this section, we present the key contributions that make *LightNeuS* a neural implicit reconstruction method suitable for endoscopy in endoluminal cavities. In this context, the light source is located next to the camera and moves with it. Furthermore, it is close to the surfaces to be modeled. As a result, for any surface point $\mathbf{x} = \mathbf{o} + t\mathbf{d}$, the irradiance decreases with the square of the distance to the camera t. Hence, we can write the color of the corresponding pixel as [3]:

$$\mathcal{I}(\mathbf{x}) = \left(\frac{L_e}{t^2} \, \mathrm{BRDF}(\mathbf{x}, \mathbf{d}) \, \cos(\theta) \, g \right)^{1/\gamma} \tag{3}$$

where L_e is the radiance emitted by the light source to the surface point, that was modeled and calibrated in the EndoMapper dataset [1] according to the SLS model from [16]. The bidirectional reflectance distribution function (BRDF) determines how much light is reflected to the camera, and the cosine term $\cos(\theta) = -\mathbf{d} \cdot \mathbf{n}$ weights the incoming radiance with respect to the surface normal $\mathbf{n}$. Equation (3) also takes into account the camera gain g and gamma correction γ.

3.1 Using Illumination Decline as a Depth Cue

The NeuS formulation of Section 2 assumes distant and fixed lighting. However, in endoscopy inverse-square light decline is significant, as quantified in Eq. (3). Accounting for this is done by modifying the original NeuS formulation as follows. Fig. 1 depicts the original NeuS network in black. It uses a SDF network—shown in orange—to estimate a view-independent geometry and only the final RGB color depends on the viewing direction $\mathbf{d}$. It is estimated by the network shown in green. Thus, this second network $\mathbf{c}(\mathbf{x}, \mathbf{d}, \mathbf{n})$ may learn to model non-Lambertian $\mathrm{BRDF}(\mathbf{x}, \mathbf{d})$, including specular highlights, and the cosine term of Eq. (3). However, if the distance t from the light to the point $\mathbf{x}$ is not provided to the color network, the $1/t^2$ dependency cannot be learned, and surface reconstruction will fail. Our key insight is to explicitly supply this distance as input to the volume rendering algorithm, as shown in red in Fig. 1 and reformulate Eq. (1) as

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \, \sigma(\mathbf{r}(t)) \, \frac{\mathbf{c}(\mathbf{r}(t), \mathbf{d}, \mathbf{n})}{t^2} \, \mathrm{d}t \tag{4}$$

This conceptually simple change, using illumination decline while training, unlocks all the power of neural surface reconstruction in endoscopy.

3.2 Endoscope Photometric Model

Apart from illumination decline, there are several significant differences between the images captured by endoscopes and those conventionally used to train NeRFs and NeuS: fish-eye lenses, strong vignetting, uneven scene illumination, and post-processing.

Endoscopes use fisheye lenses to cover a wide field of view, usually close to 170 degrees. These lenses produce strong deformations, making it unwise to use the standard pinhole camera model. Instead, specific models [19, 10] must be used. Hence, we also modified the original NeuS implementation to support these models.

The light sources of endoscopes behave like spotlights. In other words, they do not emit with the same intensity in all directions, so L_e in Eq. (3) is not constant for all image pixels. This effect is similar to the vignetting effect caused by conventional lenses, that is aggravated in fisheye lenses. Fortunately, they can be accurately calibrated [1, 16] and compensated for.

The post-processing software of medical endoscopes is designed to always display well-exposed images, so that physicians can see details correctly. An adaptive gain factor g is applied by the endoscope's internal logic and gamma correction is also used to adapt to non-linear human vision, achieving better contrast perception in mid tones and dark areas. Endoscope manufacturers know the post-processing logic of their devices, but this information is proprietary and not available to users. Again, gamma correction can be calibrated assuming it is constant [3], and the gain change between successive images can be estimated, for example, by sparse feature matching.

All these factors must be taken into account during network training. Thus, our photometric loss is computed using a normalized image:

$$I' = \left(\frac{I^\gamma}{L_e g} \right)^{1/\gamma} \tag{5}$$

4 Experiments

We validate our method on the C3VD dataset [4], which covers all different sections of the colon anatomy in 22 video sequences. This dataset contains sequences recorded with a medical video colonoscope, Olympus Evis Exera III CF-HQ190L. The images were recorded inside a *phantom*, a model of a human colon made of silicone. The intrinsic camera parameters are provided. The camera extrinsics for each frame are estimated by 2D-3D registration against the known 3D model. In an operational setting, we could use a structure-from-motion approach such as COLMAP [21] or a SLAM technique such as [8], which have been shown to work well in endoscopic settings. The gain values were easily estimated from the dataset itself. For vignetting, we use the calibration obtained from a colonoscope of the same brand and series from the EndoMapper dataset [1].

During training, we follow the NeuS paper approach of using a few informative frames per scene, as separated as possible, by sampling each video uniformly. For each sequence, we train both the vanilla NeuS and our LighNeuS using 20 frames each time. They are extracted uniformly over the duration of the video. We use the same batch size and number of iterations as in the original NeuS paper, 512 and 300k respectively. Once the network is trained, we can extract triangulated meshes from the reconstruction. Since the C3VD dataset comprises

Table 1. Reconstruction error [mm] on the C3VD dataset. Surveyed: points seen at least once. **Extended:** points within 20 mm of a visible point. Anatomical regions: Cecum, Descending, Sigmoid and Transverse. For NeuS, we provide two sets of numbers because the optimization failed on the other sections. In *italics* we mark the sequences where the camera moves less than 1 cm yielding higher errors.

		NeuS		**LightNeuS (ours)**										
Sequence		**C1a**	**C4b**	**C1a**	**C1b**	**C2a**	**C2b**	**C2c**	**C3a**	**C4a**	**C4b**	**D4a**	**S1a**	**S2a**
Sur.	MedAE	4.53	10.6	0.95	4.85	1.40	3.26	2.57	1.12	1.90	1.41	2.66	4.23	1.19
	MAE	5.07	10.6	1.48	5.11	1.54	3.65	3.00	2.54	2.14	1.63	3.26	4.33	1.89
	RMSE	6.40	11.6	2.01	5.63	1.87	4.39	3.74	5.49	2.92	2.10	4.08	4.96	2.78
Ext.	MedAE	4.68	5.35	0.83	4.89	1.41	3.32	2.54	1.27	1.91	1.45	4.50	4.01	1.40
	MAE	6.24	6.74	1.26	5.10	1.56	3.70	3.01	3.83	2.18	1.72	6.61	4.19	2.36
	RMSE	8.77	8.56	1.72	5.60	1.90	4.42	3.77	7.96	2.95	2.20	9.32	4.87	3.96

LightNeuS (ours)												
S3a	**S3b**	**T1a**	**T1b**	**T2a**	**T2b**	**T4a**	**Mean**	***T2c***	***T3a***	***T3b***	***T4b***	***Mean***
2.57	3.63	3.43	2.33	2.24	2.16	1.15	**2.39**	*5.07*	*6.39*	*11.0*	*1.75*	***6.04***
2.68	4.16	3.47	2.72	2.28	2.30	2.31	**2.80**	*5.45*	*8.65*	*12.1*	*6.70*	***8.23***
3.18	4.81	4.07	3.34	2.58	2.70	3.79	**3.58**	*6.48*	*10.7*	*14.4*	*11.3*	***10.7***
2.87	3.54	3.38	2.69	2.19	2.12	1.29	**2.53**	*4.44*	*6.54*	*13.6*	*8.00*	***8.16***
3.27	4.64	3.31	3.21	2.22	2.28	2.22	**3.15**	*5.36*	*8.10*	*14.1*	*10.4*	***9.47***
4.04	6.10	3.86	3.96	2.55	2.69	3.32	**4.18**	*6.78*	*9.94*	*15.9*	*13.9*	***11.6***

a ground-truth triangle mesh, we compute point-to-triangle distances from all the vertices in the reconstruction to the closest ground-truth triangle.

In the first rows of Table 1, we report median (MedAE), mean (MAE), and root mean square (RMSE) values of these distances for all vertices seen in at least one image. Columns show the result for 22 sequences. We note 18 sequences where the camera moved at least 1 cm, and the reconstruction yielded a mean error of 2.80 mm. The other four smaller trajectories (<1cm) lack parallax and the mean error is higher (8.23mm).

This is in the range of reported accuracy in the literature for monocular dense non-watertight depth estimation, 1.1 mm in [14] for high parallax geometry in laparoscopy, which is a much more favorable geometry than the one we have here, or 0.85 mm for the significantly smaller-size cavities of endoscopic endonasal surgery (ESS) [11].

In contrast, vanilla NeuS assumes constant illumination. The strong light changes typical of endoscopy fatally mislead the method. We only report numerical results of NeuS in two sequences because in all the rest, the SDF diverges and ends up blown out of the rendering volume, giving no result at all.

We provide a qualitative result in Fig. 2 and additional ones in the supplementary material. Note that the watertight prior inherent to an SDF allows the network to hallucinate unseen areas. Remarkably, these unsurveyed areas continue the tubular shape of the colon and we found them to be mostly accurate when compared to the ground truth. For example, the curved areas of the colon where a wall is occluded behind the corner of the curve is reconstructed,

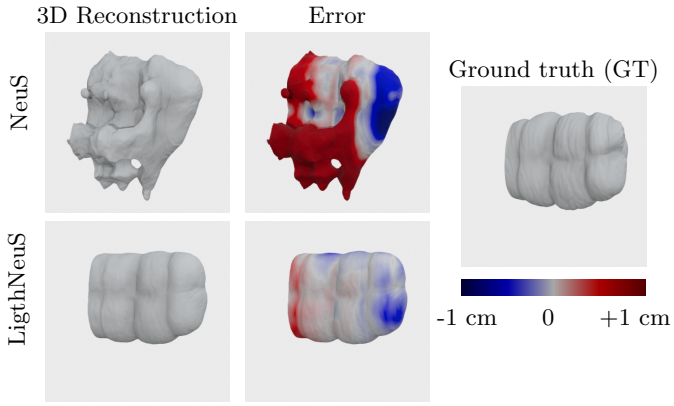

Fig. 2. Benefits of illumination decline. Result on the *"Cecum 1 a"* sequence. **Top:** The NeuS reconstruction exhibits multiple artifacts that make it unusable. **Bottom:** Our reconstruction is much closer to the ground truth shape. The error is shown in blue if the reconstruction is inside the surface, and in red otherwise. A fully saturated red or blue denotes an error of more than 1cm and grey denotes no error at all.

as shown in Fig. 3. This ability to "fill in" observation gaps may be useful in providing the endoscopist with an estimate of the percentage of unsurveyed area during a procedure.

We hypothesize that this desirable behavior stems from the fact that the network learns an empirical shape prior from the observed anatomy of the colon. However, we don't expect this behavior to hold for distant unseen parts, but only for regions closer than 20 mm to one observation. In the last rows of Table 1, we compute accuracy metrics for this *extended* region. It includes not only surveyed areas, but also neighboring areas that were not observed.

5 Conclusion

We have presented a method for 3D dense multi-view reconstruction from endoscopic images. We are the first to show that neural radiance fields can be used to obtain accurate dense reconstructions of colon sections of significant length. At the heart of our approach, is exploiting the correlation between depth and brightness. We have observed that, without it, neural reconstruction fails.

The current method could be used offline for post-exploration coverage analysis and endoscopist training. But real-time performance could be achieved in the future as the new NeuS2 [26] converges in minutes, enabling automatic coverage reporting. Similar to other reconstruction methods, for now our approach works in areas of the colon where there is little deformation. Several sub-maps of non-deformed areas can be created if necessary. However, this limitation could be overcome by adopting the deformable NeRFs formalism [18].

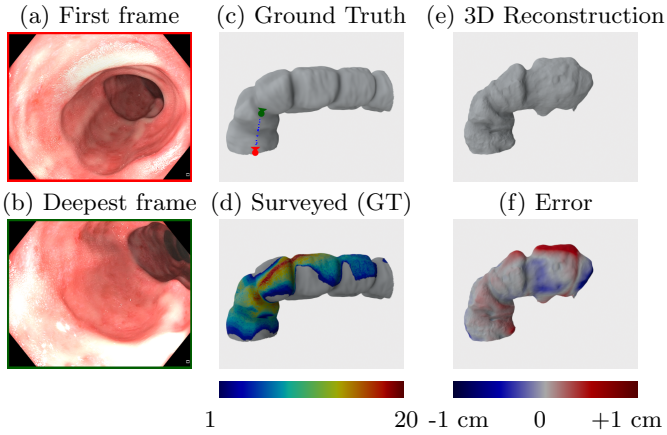

Fig. 3. Reconstructing partially observed regions. Results on *"Transcending 4 a"* sequence. The camera performs a short trajectory from (a) to (b). In (c) we represent both frames and intermediate camera poses. (d) Number of frames seeing each surface point, with GT unobserved areas shown in gray. (e) We managed to reconstruct a curved section of the colon. (f) Our method plausibly estimates the wall of the colon at the right of camera (b), although it was never seen in the images.

Acknowledgement

This work was supported by EU-H2020 grant 863146: ENDOMAPPER, Spanish government grants PID2021-127685NB-I00 and FPU20/06782 and by Aragón government grant DGA_T45-17R.

References

1. Azagra, P., Sostres, C., Ferrandez, A., Riazuelo, L., Tomasini, C., Barbed, O.L., Morlana, J., Recasens, D., Batlle, V.M., Gómez-Rodríguez, J.J., Elvira, R., López, J., Oriol, C., Civera, J., Tardós, J.D., Murillo, A.C., Lanas, A., Montiel, J.M.M.: EndoMapper dataset of complete calibrated endoscopy procedures. arXiv:2204.14240 (2022)
2. Bae, G., Budvytis, I., Yeung, C.K., Cipolla, R.: Deep multi-view stereo for dense 3D reconstruction from monocular endoscopic video. In: Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 774–783 (2020)
3. Batlle, V.M., Montiel, J.M.M., Tardós, J.D.: Photometric single-view dense 3D reconstruction in endoscopy. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). pp. 4904–4910 (2022)
4. Bobrow, T.L., Golhar, M., Vijayan, R., Akshintala, V.S., Garcia, J.R., Durr, N.J.: Colonoscopy 3D video dataset with paired depth from 2D-3D registration. arXiv:2206.08903 (2022)
5. Campos, C., Elvira, R., Gómez-Rodríguez, J.J., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. IEEE Transactions on Robotics **37**(6), 1874–1890 (2021)

6. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(3), 611–625 (2018)
7. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-scale direct monocular SLAM. In: European Conf. on Computer Vision (ECCV). pp. 834–849 (2014)
8. Gómez-Rodríguez, J.J., Lamarca, J., Morlana, J., Tardós, J.D., Montiel, J.M.M.: SD-DefSLAM: Semi-direct monocular SLAM for deformable and intracorporeal scenes. In: IEEE Int. Conf. on Robotics and Automation (ICRA). pp. 5170–5177 (2021)
9. Kajiya, J.T., Von Herzen, B.P.: Ray tracing volume densities. SIGGRAPH Comput. Graph. **18**(3), 165–174 (jan 1984)
10. Kannala, J., Brandt, S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(8), 1335–1340 (2006)
11. Liu, X., Li, Z., Ishii, M., Hager, G.D., Taylor, R.H., Unberath, M.: Sage: Slam with appearance and geometry prior for endoscopy. In: IEEE Int. Conf. on Robotics and Automation (ICRA). pp. 5587–5593 (2022)
12. Ma, R., Wang, R., Pizer, S., Rosenman, J., McGill, S.K., Frahm, J.M.: Real-time 3D reconstruction of colonoscopic surfaces for determining missing regions. In: Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 573–582 (2019)
13. Ma, R., Wang, R., Zhang, Y., Pizer, S., McGill, S.K., Rosenman, J., Frahm, J.M.: RNNSLAM: Reconstructing the 3D colon to visualize missing regions during a colonoscopy. Medical image analysis **72**, 102100 (2021)
14. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE Transactions on Medical Imaging **38**(1), 79–89 (2019)
15. 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 (dec 2021)
16. Modrzejewski, R., Collins, T., Hostettler, A., Marescaux, J., Bartoli, A.: Light modelling and calibration in laparoscopy. Int. Journal of Computer Assisted Radiology and Surgery **15**(5), 859–866 (2020)
17. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense tracking and mapping in real-time. In: IEEE Int. Conf. on Computer Vision (ICCV). pp. 2320–2327 (2011)
18. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: IEEE/CVF Int. Conf. on Computer Vision (ICCV). pp. 5865–5874 (2021)
19. Scaramuzza, D., Martinelli, A., Siegwart, R.: A toolbox for easily calibrating omnidirectional cameras. In: IEEE/RJS Int. Conf. on Intelligent Robots and Systems (IROS). pp. 5695–5701 (2006)
20. Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conf. on Computer Vision (ECCV). pp. 501–518 (2016)
21. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2016)
22. Sengupta, A., Bartoli, A.: Colonoscopic 3D reconstruction by tubular non-rigid structure-from-motion. Int. Journal of Computer Assisted Radiology and Surgery **16**(7), 1237–1241 (2021)

23. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **71**(3), 209–249 (2021)
24. Tokgozoglu, H.N., Meisner, E.M., Kazhdan, M., Hager, G.D.: Color-based hybrid reconstruction for endoscopy. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 8–15 (2012)
25. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In: Advances in Neural Information Processing Systems. vol. 34, pp. 27171–27183 (2021)
26. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: NeuS2: Fast learning of neural implicit surfaces for multi-view reconstruction. arXiv:2212.05231 (2022)
27. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3D reconstruction of deformable tissues in robotic surgery. In: Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 431–441 (2022)
28. Zhao, Q., Price, T., Pizer, S., Niethammer, M., Alterovitz, R., Rosenman, J.: The endoscopogram: A 3D model reconstructed from endoscopic video frames. In: Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 439–447 (2016)

Supplementary material

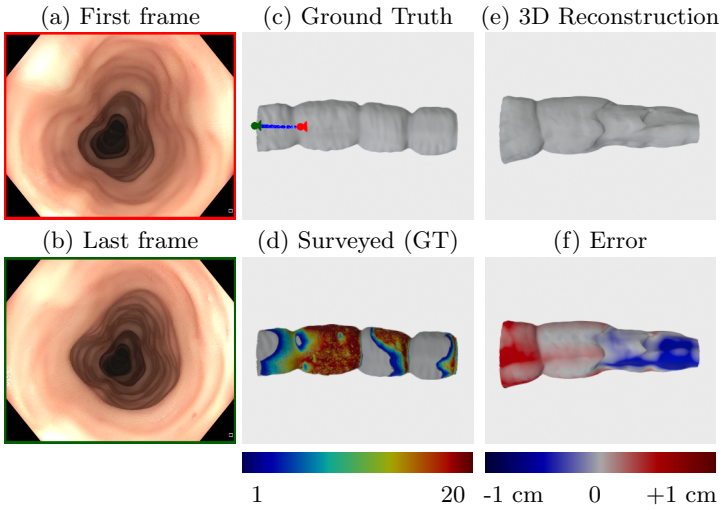

Fig. 4. Reconstructing with low parallax. Results on *"Transcending 1 a"* sequence. (c) The camera travels in a straight line, covering less than a third of the section. As shown in (a) and (b), the haustra completely hide the background walls. (e) Consequently, the reconstruction underestimates the diameter of the end of the tube. However, the three characteristic folds in our reconstructed colon match the ground truth in number and location. In addition, areas observed multiple times —red in (d)— are reconstructed with high accuracy —gray in (f).

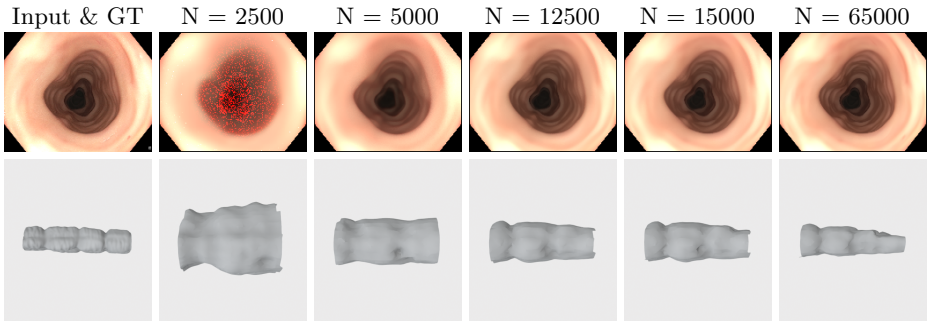

Fig. 5. Reconstruction convergence. Results on *"Transcending 1 a"* sequence. We show the intermediate results for N optimisation iterations. We see how the reconstruction converges quickly. In 65k iterations we already have a reasonable solution, compared to the 300k iterations proposed by the authors of NeuS.

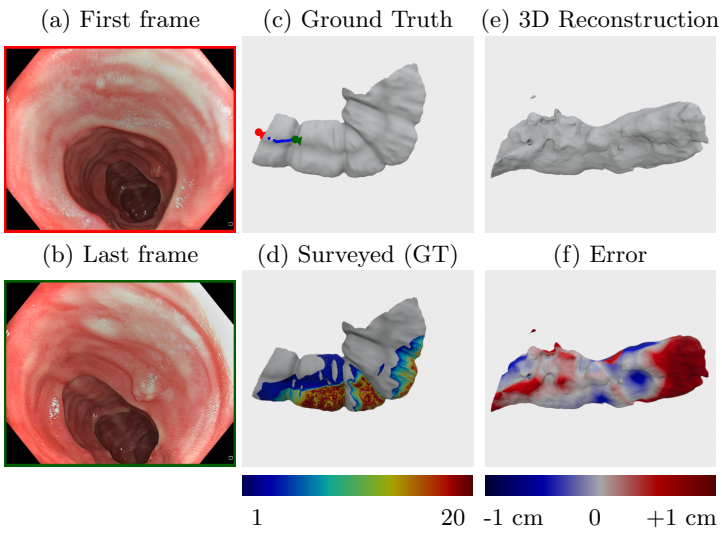

Fig. 6. Reconstructing congruent shapes. Results on *"Descending 4 a"* sequence. (c) The camera takes the most challenging route: the shortest translation of the sequences tested, turning towards the right wall. (d) This results in very poor coverage, especially to the left of the camera. The curve to the left is never seen. (e) In this way we check that our method only "hallucinates" partially observed areas, based on the structure of the region it has actually observed. As a result, the reconstruction continues as a straight tube. Again, areas observed multiple times —red in (d)— are reconstructed with high accuracy —gray in (f).

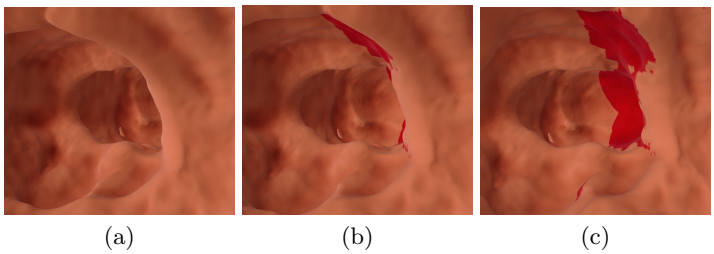

Fig. 7. Post-intervention 3D visualization. Results on *"Transcending 4 a"* sequence. Our reconstructions would allow physicians to revisit the area explored during the endoscopy. This opens the door to augmented reality (AR) in post-intervention diagnostics. As an example, we show a visualisation of the area not surveyed during the procedure, marked in red. After inspecting a colon section, our watertight surface displays (a) visualized and (b), (c) non-visualized mucosa. The doctor can analyse non-visualized areas and make decisions about subsequent exploration. A video of this demonstration is included in the supplementary material.