Colonoscopic 3D Reconstruction by Tubular Non-Rigid Structure-from-Motion

Agniva Sengupta and Adrien Bartoli

EnCoV, Institut Pascal, UMR6602 CNRS/UCA, Clermont-Ferrand, France

1 Introduction

Reconstructing the 3D colonic surface and localising the colonoscope's distal end from the video stream would aid the spatial understanding of lesions and hence diagnosis. Using Non-Rigid Structure-from-Motion (NRSfM) [1] is thus an appealing idea. Although low-rank NRSfM was attempted on a short beating heart sequence [4], general NRSfM methods have not been applied to endoscopy data in the literature. Modern isometric methods [2, 7, 3] performed poorly or failed in our experiments, even in simple cases. Stronger template-based methods such as [8] can unfortunately not be used, because a matchable template is not available. Estimating depth from a single or a stream of monocular images using deep learning is challenging in endoscopy due to the unavailability of labelled data. Promising attempts were made to train with synthetic data, for which there is a domain adaptation problem, and with self-supervised learning [5]. Unfortunately, there is yet no publicly available monocular 3D reconstruction network for endoscopy. Colonoscopic images are particularly difficult with NRSfM, because the camera tends to move mainly along its optical axis, creating unstable geometric configurations, and the spreading of the correspondences is often uneven within the images, because of the locally weak texture. We propose to strengthen NRSfM by exploiting the known topology of the surface. Topological information has not been used in NRSfM. We specifically study the Tubular Topology Prior (TTP). Combined with surface smoothness, TTP forms a deformable geometric model, which is tube-shaped in some reference coordinate system. We provide the first isometric NRSfM method for a tubular surface as a monocular camera moves through its inner volume, as in colonoscopy.

2 Proposed Tubular 3D Reconstruction Method

As all NRSfM methods, ours takes M point correspondences over N images and the camera's intrinsic parameters as inputs. It computes a set of N surfaces corresponding to a deformed tube in camera coordinates. It works in two steps.

Step 1: initial unconstrained reconstruction. The initial unconstrained reconstruction is an intermediate step, whose result is a set of N 3D point clouds. It follows the principle of isometric NRSfM. Specifically, it is a zeroth-order method, because differential correspondences are unstable in colonoscopy. In other words, it exploits the raw point correspondences without additional information. Isometry is modeled by preserving the distance between neighbouring 3D points across the point clouds. The inter-point distances are however unknown. Our method thus *alternates* between computing the depth of all image points and the inter-point distances. The two steps are repeated until the estimates converge. The notion of point neighbourhood is defined by a Nearest-Neighbour Graph (NNG), whose nodes are the points and whose edges define the neighbours [3]. Convergence is achieved when the average update on the inter-point distances falls below a predefined threshold $t = 10^{-4}$ or exceeds a maximum number of iterations $h = 14$ in our experiments.

Step 2: tubular parameterisation. The tubular parameterisation upgrades the reconstructed 3D point clouds to smooth surfaces of tubular topology. We represent such a surface by the composition of two maps. The first map, from 2D to 3D, is fixed. It embeds a planar template to a circular cylinder with unit radius. The second map, from 3D to 3D, deforms the cylinder and is represented by a harmonic spline, the 3D equivalent of the classical Thin-Plate Spline, for which we use as many control points as reconstructed 3D points. We fit the maps to the 3D point clouds by minimising the Euclidean distance between the control points and the corresponding reconstructed 3D points and the bending of the unit-circular cylinder, with the Levenberg-Marquardt algorithm.

3 Experimental Results

Synthetic sequences. We simulated two sequences, *NR-Synth-1* and *NR-Synth-2*, using Blender. They contain 69 and 79 frames respectively and 160 3D points each. Our proposed method (first step of our pipeline) is denoted **IsoSfM0-Alt** (for 0-th order, Alternation). We compare it with the authors's implementation of **IsoSfMH** [2] (for homography based), **IsoSfM2** [7] (for 2-nd order) and **IsoSfM0-SOCP** [3] (for 0-th order using Second-Order Cone Programming). We use the mean Euclidean distance e_p between the reconstructed 3D points and the groundtruth as primary evaluation metric and the Euclidean distance e_c between the reconstructed 3D points and the nearest 3D points on the groundtruth shape for error visualisation. A visual comparison of some random representative frames are shown in figure 1. For *NR-Synth-1*, **IsoSfM0-Alt** is 64.36%, 64.32% and 47.57% better in e_p than **IsoSfMH**, **IsoSfM2** and **IsoSfM0-SOCP** respectively. **IsoSfM2** fails to complete the reconstruction of *NR-Synth-2* (the authors' Matlab code crashed while solving for the depth of 3D points), but **IsoSfM0-Alt** is 79.51% and 72.31% better in e_p than **IsoSfMH** and **IsoSfM0-SOCP** respectively. This is a significant improvement. The parameterised reconstruction using TTP (second step of our pipeline) is shown in figure 2.

Real sequence. We extracted a short sequence of 36 frames from *the endoscopy image database for research and training*, approval UK IRAS Project ID 236056, which was kindly provided to us by UCL, and manually annotated 50 points

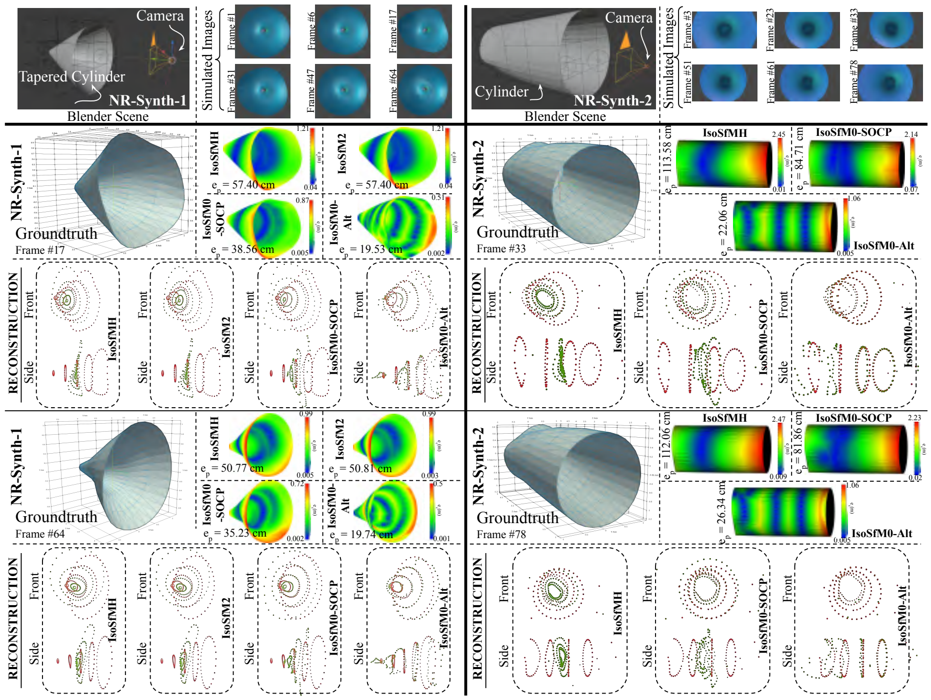

Fig. 1. Comparison of *initial unconstrained reconstruction (step 1)* results, synthetic sequences. Top row: simulation setup and sample frames. Other parts: reconstruction results; the green and red dots are the reconstructed and groundtruth points respectively.

across the sequence. The points are unevenly spread owing to the lack of texture. We ran all four methods. **IsoSfMH** and **IsoSfM2** failed to complete the reconstruction (similarly to *NR-Synth-2*). **IsoSfM0-SOCP** produced a 3D reconstruction flatter than **IsoSfM0-Alt**'s, as shown in figure 3. The parameterised reconstruction using TTP is shown in figure 3.

4 Conclusion

By developing a new method exploiting the tubular topology, we have been able to give initial results of NRSfM in colonoscopy. These results are very encouraging, even if our method and experiments are preliminary. Future work will involve using automatic correspondences, developing an initialisation and a refinement method exploiting the topology prior and comparing to deep learning methods such as [6] on public benchmarks.

Acknowledgments. This work was funded by the FET-Open grant 863146 Endomapper.

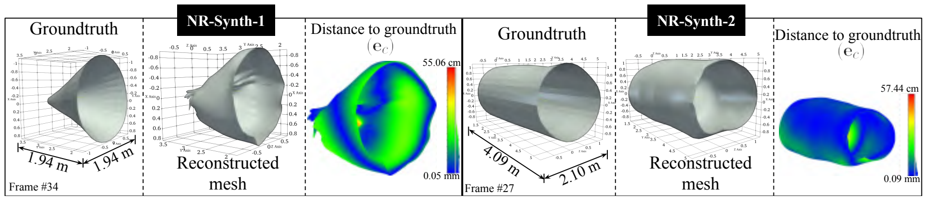

Fig. 2. Final reconstructed surface using *tubular parameterisation (step 2)*, for the proposed method on synthetic sequences.

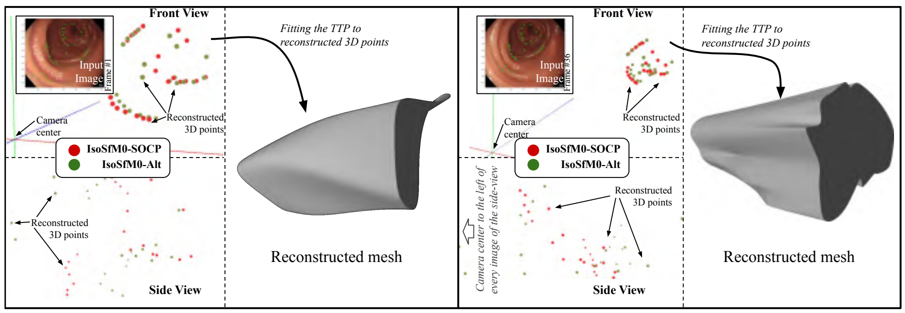

Fig. 3. Comparison of 3D reconstruction results, real sequence, with the final reconstructed surface for the proposed method.

References

1. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 690–696. IEEE (2000)
2. Chhatkuli, A., Pizarro, D., Bartoli, A.: Non-rigid shape-from-motion for isometric surfaces using infinitesimal planarity. In: British Machine Vision Conference (2014)
3. Chhatkuli, A., Pizarro, D., Collins, T., Bartoli, A.: Inextensible non-rigid structure-from-motion by second-order cone programming. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(10), 2428–2441 (2017)
4. Kumar, S., Dai, Y., Li, H.: Spatio-temporal union of subspaces for multi-body non-rigid structure-from-motion. Pattern Recognition **71**, 428–443 (2017)
5. Liu, X., Sinha, A., Ishii, M., Hager, G.D., Reiter, A., Taylor, R.H., Unberath, M.: Dense depth estimation in monocular endoscopy with self-supervised learning methods. IEEE Transactions on Medical Imaging **39**(5), 1438–1447 (2019)
6. Mahmood, F., Durr, N.J.: Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. Medical Image Analysis **48**, 230–243 (2018)
7. Parashar, S., Pizarro, D., Bartoli, A.: Isometric non-rigid shape-from-motion with riemannian geometry solved in linear time. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(10), 2442–2454 (2017)
8. Salzmann, M., Hartley, R., Fua, P.: Convex optimization for deformable surface 3-d tracking. In: 2007 IEEE 11th International Conference on Computer Vision. pp. 1–8. IEEE (2007)