**Noname manuscript No.**
(will be inserted by the editor)

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

**Agniva Sengupta · Adrien Bartoli**

Received: date / Accepted: date

**Abstract Purpose.** The visual examination of colonoscopic images fails to extract precise geometric information of the colonic surface. Reconstructing the 3D surface of the colon from colonoscopic image sequences may thus add valuable clinical information. We address this problem of extracting precise spatio-temporal 3D structure information from colonoscopic images.
**Methods.** Using just the intrinsically calibrated monocular image stream, we develop a technique to compute the depth of certain feature points that have been tracked across images. Our method uses the prior knowledge of an approximate geometry of the colon, called the Tubular Topology Prior (TTP). It works by fitting a deformable cylindrical model to points reconstructed independently by Non-Rigid Structure-from-Motion (NRSfM), compromising between the data term and a novel tubular smoothing prior. Our method represents the first method ever to exploit a very weak topological prior to improve NRSfM. As such, it lies in-between standard NRSfM, which does not use a topological prior beyond the mere plane, and Shape-from-Template (SfT), which uses a very strong prior as a full deformable 3D object model.
**Results.** We validate our method on both synthetic images of tubular structures and real colonoscopic data. Our method improves the results obtained by existing NRSfM methods by 71.74% on average on synthetic data and succeeds in obtaining 3D reconstruction from a real colonoscopic sequence defeating the existing methods.
**Conclusion.** Colonoscopic 3D reconstruction is a difficult problem, which is yet unresolved by the existing methods from computer vision. Our proposed dedicated NRSfM method and experiments show that the visual motion might be the right visual cue to use in colonoscopy.

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

Agniva Sengupta . Adrien Bartoli
EnCoV, Institut Pascal, UMR6602 CNRS/UCA
Clermont-Ferrand, France
E-mail: {agniva.sengupta, adrien.bartoli}@uca.fr

**Keywords** Non-Rigid Structure-from-Motion · 3D Reconstruction · Colonoscopy

## 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 NRSfM [4] is thus an appealing idea. Although low-rank NRSfM was attempted on a short beating heart sequence [8], general NRSfM methods have not been applied to endoscopy data in the literature. Modern isometric methods [5,11,6] performed poorly or failed in our experiments, even in simple cases. Stronger template-based methods such as [13] can unfortunately not be used, because a matchable template is not available. Many recent techniques estimate depth from a single or a stream of monocular images using deep learning [14]. The use of these methods in endoscopy is under development. The main problem is the unavailability of labeled data, which prevents conventional supervised learning. Promising attempts were made to train with synthetic data, for which there is a domain adaptation problem, and with self-supervised learning [9]. Unfortunately, there is yet no publicly available monocular 3D reconstruction network for endoscopy. Shape-from-Shading (SfS) methods use a single image to reconstruct the 3D shape of a surface [15, 1]. Endoscopy is a special case for SfS because the attached light source is approximately colocated with the camera and can be calibrated. SfS methods have been applied in laparoscopy [7], which showed that ambiguities remained on the reconstructed surface. In addition, the colon surface typically presents strong specular reflections, which may substantially degrade SfS results. SfS is thus not adapted to the problem of 3D reconstruction in colonoscopy.

Overall, NRSfM thus seems to be a promising and well adapted approach to the problem, but the existing techniques are not ready to cope with the difficulties posed by colonoscopy. 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 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

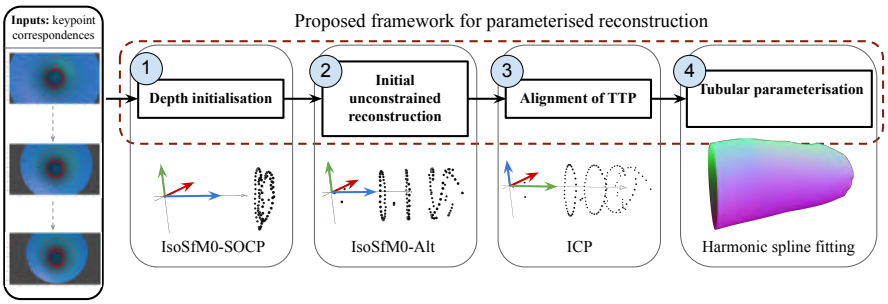


**Fig. 1** Summary of the proposed four-step framework for parameterised deformable surface reconstruction of tubular objects, illustrated on synthetically generated data. We use this framework to reconstruct the structure of the colon from colonoscopic images.

corresponding to a deformed tube in camera coordinates. It works in four steps, as illustrated in Fog. 1.

*Step 1: depth initialisation.* We use the NRSfM method [6] to bootstrap our pipeline, by finding an initial coarse reconstruction. We use the public implementation of [6]. As shown in Sec. 3, this provides an approximate estimate of the scene structure and suffers from inaccuracies. The next steps of our pipeline considerably improve the reconstruction results.

*Step 2: 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 [6]. The NNG depends on the user-chosen number of neighbours, which we chose using a specific experiment given below. The alternate computation of depth and inter-point distance weakly enforces the concept of Maximum Depth Heuristics (MDH), proposed by [12] and later extended to NRSfM by [6], by accepting depth updates that *push* the current estimate of depth away from the camera centre and rejecting every other possible update. 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 3: alignment of TTP.* We align each of the $N$ reconstructed 3D point clouds from *Step 2* to the TTP. This is a problem of finding the rigid transfor-

mation between a tubular surface and a point cloud, which we solve with the Iterative Closest Point (ICP) algorithm. The TTP is defined as a unit circular cylinder in fix world coordinates. We sample it semi-densely using $\mathcal{O}(1000)$ points and run a conventional point-to-point ICP [2].

*Step 4: 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 NRSfM method (second step of our pipeline) is denoted **IsoSfM0-Alt** (for 0-th order, Alternation). We compare it with the authors's implementation of **IsoSfMH** [5] (for homography based), **IsoSfM2** [11] (for 2-nd order) and **IsoSfM0-SOCP** [6] (for 0-th order using Second-Order Cone Programming [3]). 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 2. For *NR-Synth-1*, **IsoSfM0-Alt** is 72.76%, 72.75% and 59.93% 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 80.09% and 73.16% better in $e_p$ than **IsoSfMH** and **IsoSfM0-SOCP** respectively. This is a significant improvement. As shown in figure 2, all three compared state-of-the-art methods fail to recover the cylindrical shape of the data and reconstruct a nearly planar surface instead. It must be noted that the boot-strapping of depth using **IsoSfM0-SOCP**, as proposed in *step 1* of our pipeline, although making the results precisely reproducible, does not constitute a necessary requirement for NRSfM using *step 2*. To highlight this aspect, we present a variant of our proposed method by replacing the depth initialisation of *step 1* by random initialisation of depth (denoted by **IsoSfM0-Alt-II**). The results of comparison between **IsoSfM0-SOCP** and **IsoSfM0-Alt-II** are given in figure 4 for a varying number of neighbourhood points of the NNG. This shows that using at

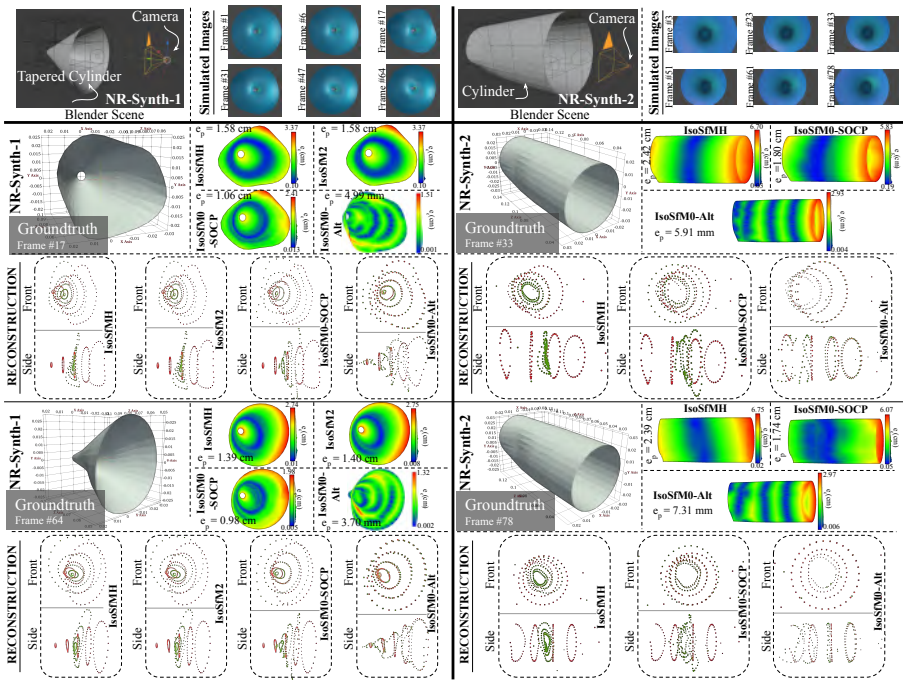


**Fig. 2** Comparison of *initial unconstrained reconstruction (step 2)* results, synthetic sequences. Top row: simulation setup and sample frames. Other parts: reconstruction results from some randomly sampled representative frames; the green and red dots are the reconstructed and groundtruth points respectively.

least 3 neighbours gives a satisfying result. The parameterised reconstruction using TTP (step 4 of our pipeline) is shown in figure 3.

*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 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

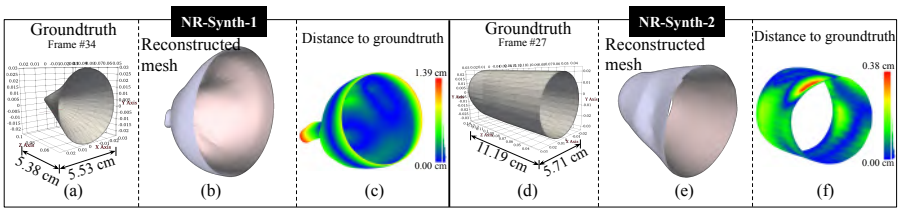


**Fig. 3** Final reconstructed surface using *tubular parameterisation (step 4)*, for the proposed framework on synthetic sequences.

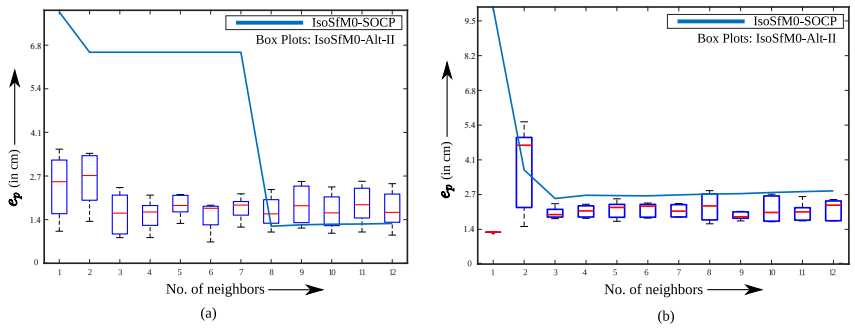


**Fig. 4** Comparison of **IsoSfM0-SOCP** with **IsoSfM0-Alt-II** across a varying number of neighbours of the NNG for *NR-Synth-1* (a) and *NR-Synth-2* (b).

3D reconstruction flatter than **IsoSfM0-Alt**'s, as shown in figure 5. The parameterised reconstruction using TTP is shown in figure 5.

## 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. Using NRSfM is important: as opposed to deep learning, which extrapolates from the training process, NRSfM uses geometric reasoning and can thus obtain a quantitatively certified result with an uncertainty characterisation. The results of NRSfM may also be used to produce training data for deep learning. Future work will involve using automatic correspondences, developing an initialisation and a refinement method exploiting the topology

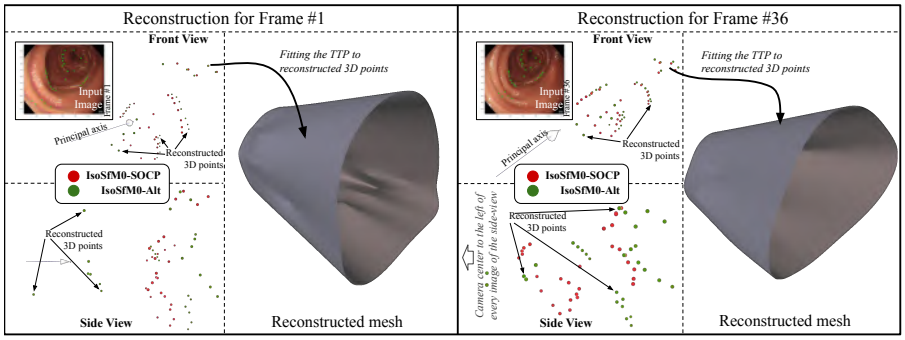


**Fig. 5** Comparison of 3D reconstruction results, real sequence, with the final reconstructed surface for the proposed method. The input images are shown as the top-left insets, along with the feature points (in green) overlaid on the images. The front and side views of the reconstructed points are shown on the left while the reconstructed meshes are shown on the right

prior and comparing to deep learning methods such as [10] on public benchmarks.

## Declarations

The authors declare that they have no conflict of interest. Informed consent was obtained from all individual participants included in the study. This article does not contain any studies with animals performed by any of the authors.

## References


1. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(8), 1670–1687 (2014)
2. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures, vol. 1611, pp. 586–606. International Society for Optics and Photonics (1992)
3. Boyd, S., Boyd, S.P., Vandenberghe, L.: Convex optimization. Cambridge university press (2004)
4. 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)
5. Chhatkuli, A., Pizarro, D., Bartoli, A.: Non-rigid shape-from-motion for isometric surfaces using infinitesimal planarity. In: British Machine Vision Conference (2014)
6. 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)
7. Collins, T., Bartoli, A.: Towards live monocular 3d laparoscopy using shading and specularity information. In: International Conference on Information Processing in Computer-Assisted Interventions, pp. 11–21. Springer (2012)
8. 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)
9. 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)
10. 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)
11. 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)
12. Salzmann, M., Fua, P.: Linear local models for monocular reconstruction of deformable surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(5), 931–944 (2010)
13. 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)
14. Saxena, A., Chung, S.H., Ng, A.Y.: Learning depth from single monocular images. In: Advances in Neural Information Processing Systems, pp. 1161–1168 (2006)
15. Xiong, Y., Chakrabarti, A., Basri, R., Gortler, S.J., Jacobs, D.W., Zickler, T.: From shading to local shape. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(1), 67–79 (2014)