Real-time Vision Feedback for Hydraulic Soft Continuum Robotic System in Ureteroscopy

Ruofeng Wei[1], Yiang Lu[2], Philip T. Zhao[4], Rui Li[3], Junichi Tokuda[1]
[1]*Brigham and Women's Hospital, Harvard Medical School*
[2]*Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong*
[3]*Mechanical Engineering Department, NYU Tandon School of Engineering*
[4]*Department of Urology, NYU Grossman School of Medicine*
rwei2@bwh.harvard.edu

Aims: Ureteroscopy is a minimally invasive procedure using a flexible endoscope inserted through the ureter. It is used to diagnose and treat upper tract urothelial carcinoma and remove kidney stones [1]. This project aims to enhance the safety and accuracy of ureteroscopy by developing a hydraulic soft continuum robotic system. This paper focuses on an intelligent visual perception framework designed for soft robotics [2] to facilitate computer-assisted surgical navigation in confined anatomical spaces.

Materials and Methods: We introduce an image-based perception framework (Fig. 1) that reconstructs a 3D map of the ureter and renal pelvis from monocular endoscopic images. Our approach includes 2D and 3D learning-based perception algorithms that estimate depth maps from each image frame and stitch them together to create a 3D point cloud. Due to partial obstruction of the camera view by a guidewire (Fig. 1), the guidewire in the initial frame was masked using the vision-language model (VLM), Claude [3], which processed segmentation masks generated by the SAM [4] model. A long-term tracking model was then employed to track the guidewire in subsequent frames. For depth estimation, we utilized the Depth Anything model, which effectively captures complex surgical environments, including varying ureter wall depths and guidewire placements. A graphics rendering engine with physics simulator based on SofaUnity [5] was set up to collect a dataset of realistic ureteroscope images, allowing simulating the motion of the flexible scope within a simulated environment.

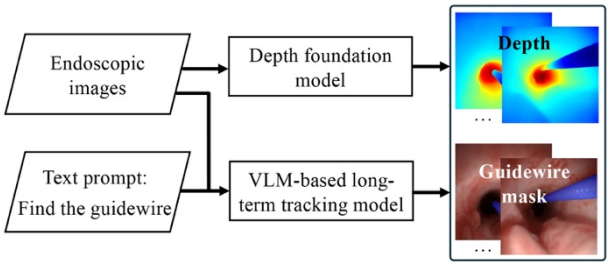

Fig. 1 Overview of the proposed image-based perception framework for robotic ureteroscopy.

Results: Our framework was tested on a clinical ureteroscopy video, where the endoscope was inserted through the ureter, reaching the renal pelvis. The VLM-based tracking model could accurately localize the guidewire for approximately 60 seconds. Our simulation (Fig. 2) demonstrated the configuration of the simulator, including the soft ureteroscope and urinary collecting system, alongside synthetic data examples.

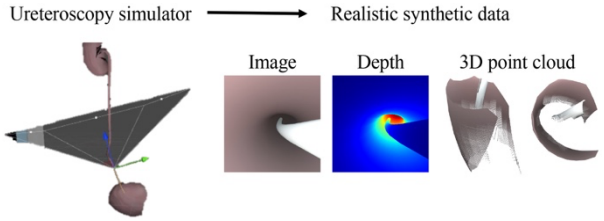

Fig. 2 Simulation of ureteroscopy and example of synthetic data generation.

Conclusions: This paper presents the first monocular endoscopic image-based perception framework for robotic ureteroscopy. We plan three additional studies to enhance the method, including depth estimation fine-tuning, distance calculations, and automated navigation of the hydraulic soft continuum robot using our perception framework.

Acknowledgements: This project was funded in part by NIH R01EB020667, R01EB036015, and R01CA235134.

References:

[1] Browne, Brendan M., Kristian D. Stensland, Matthew J. Moynihan, and David Canes. "An analysis of staging and treatment trends for upper tract urothelial carcinoma in the National Cancer Database." Clinical Genitourinary Cancer 16, no. 4 (2018): e743-e750.
[2] Burgner-Kahrs, Jessica, D. Caleb Rucker, and Howie Choset. "Continuum robots for medical applications: A survey." IEEE transactions on robotics 31, no. 6 (2015): 1261-1280.
[3] https://www.anthropic.com/news/claude-3-5-sonnet. Accessed 10 March 2025.
[4] Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. "Segment anything." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 4015-4026. 2023.
[5] https://infinytech3d.com/solutions-sofaunity/.