

An ORB-SLAM3 Based Approach for Surgical Navigation in Ureteroscopy

Laura Oliva-Maza^a, Florian Steidle^a, Julian Klodmann^a, Klaus Strobl^a and Rudolph Triebel^a

^aDeutsches Zentrum für Luft- und Raumfahrt (DLR), Münchener Str. 20, 82234 Weßling, Germany

ABSTRACT

In ureteroscopy a flexible ureteroscope is used to inspect the kidneys and the ureters. In this context, we investigate robotic solutions that leverage recent developments from Simultaneous Localization and Mapping (SLAM) to provide a 3D map of the organ to be explored, as well as an estimate of the pose of the ureteroscope. With this aid, the surgeon can navigate through the organ more precisely. Additionally, for the final organ inspection, the risk of missing certain regions of the organ is minimized. In this paper we propose a visual, monocular SLAM system based on ORB-SLAM3 that is able to estimate the pose of the ureteroscope's tip. In order to fulfill this task, we introduce two preprocessing steps. The first one aims at increasing the contrast of the image and the second one helps to avoid detecting features in non-desired regions (e.g., reflections). Additionally, we extend ORB-SLAM3 with A-KAZE and SuperPoint features and compare their performances to ORB features. The proposed method is evaluated in two experiments. The first experiment shows that we are able to estimate the trajectory of the ureteroscope and the map of a synthetic kidney with low errors. The second experiment shows that the method also gives promising results in cystoscopy, where we evaluated it with videos captured during a polyp removal procedure in a real bladder.

KEYWORDS

Ureteroscopy; SLAM; Image Processing.

1. Introduction

Surgery is evolving – from open surgery to minimally invasive surgery (MIS) to robotic surgery – in order to generally improve the patient outcome. The driving force of these developments are technological advances such as the miniaturization of mechanical components and endoscopic imaging (Rassweiler et al. 2017), (Klodmann et al. 2021). In laparoscopic MIS, one or more small incisions are made and an endoscope and tools are introduced to the patient's body to allow the surgeon to perform the procedure.

Endoluminal interventions can even further decrease the trauma of the patient. In this types of procedures the whole intervention takes place inside hollow organs (e.g., lung, stomach, colon, urethral tract). To this end, mostly flexible endoscopes are introduced through natural orifices, such as mouth, anus or urethra to inspect the organs, to take biopsies or to treat the tissue from the inside. A special treatment of Urolithiasis (stone formation in the urinary tract) to remove small kidney stones and inspect the renal collecting system is flexible ureteroscopy (fURS), where flexible endoscopes are inserted into the upper urinary tract via ureteral access sheaths.

Even though, the use of fURS increases due to technical improvements of flexible endoscopes and few contraindications (Geraghty et al. 2017), it imposes several challenges for the surgeon. These challenges entail, among others, a steeper learning curve, a smaller field of view, and the difficulty of remote spatial awareness inside the organ by watching endoscopic live video alone (Bergen and Wittenberg 2014). In order to deskill this procedure, computer vision methods based on Simultaneous Localization and Mapping (SLAM) would be useful since they provide a 3D map of the organ that is being explored together with the relative pose of the camera’s tip. This is done in a passive way and without the need of other sensors. Utilizing this information, the navigation through the organ is facilitated for the surgeon, minimizing the risk of not inspecting all renal cavities in fURS to finally guarantee that no stone fragments are left, which cannot pass through the urinary tract.

The use of SLAM in fURS is, however, especially challenging (Bergen and Wittenberg 2014). Endoscopic video suffers from fast camera motion, sudden illumination changes, reflections, lack of texture, low image quality and a small field of view. These are challenges for visual SLAM systems, which is the 2D tracking on the image feed of unknown yet salient, distinct regions of the scene called features. Hence the need of a SLAM system that is robust to the frequent loss of features or to fast camera motion. This is typically achieved by a relocalization module that recognizes regions that have been observed in the past and seamlessly registers back to them.

ORB-SLAM3 (Campos et al. 2021) is a visual SLAM system for monocular, stereo, and RGBD cameras that excels in its place recognition module. It also features a map merging capability, which means that every time that the system gets lost, it will start a new map and, when it eventually detects common regions between the active map and the stored maps, it will fuse them to create a new complete map.

The contribution of this paper is a monocular SLAM system extending the state-of-the-art ORB-SLAM3 system that is able to simultaneously estimate the 3D map of the main compartments of the urinary tract, i.e. bladder and kidney, and the pose of the ureteroscope in fURS procedures. To this end, we enhance the state of the art implementation by adding a pre-processing step to increase the contrast of the image to make the system more robust against illumination changes. In this pre-processing step we also compute a dynamic mask to avoid detecting keypoints in specular reflections and the padding borders of the image. Furthermore, we adapted some hyperparameters (e.g., thresholds) for this type of fURS scenarios. Finally, we provide an evaluation for different feature detection and description methods – ORB (Rublee et al. 2011), A-KAZE (Alcantarilla and Solutions 2011), and SuperPoint (DeTone et al. 2018) – in order to increase the accuracy and robustness of the system as well as to retain real-time performance. To evaluate the new SLAM system we recorded different sequences using a kidney phantom, together with the corresponding 3D ground truth (including the camera motion and the 3D model of the kidney). We also evaluate its generic applicability for the compartments of the urinary tract, in sequences from a real bladder polyp removal procedure.

2. State of the Art

In this section we first reviewed the main approaches to visual mapping and localization, and then focus on the optimal choice of features for fURS.

2.1. Mapping: In the field of urology (involving bladder and kidney) (Bergen and Wittenberg 2014) and (Pentek et al. 2018) reconstruct the 3D model of the bladder

Table 1. Most representative monocular SLAM systems.

Name	Features	Data association	Relocalization	Loop closure	Map merging
MonoSLAM	Shi-Tomasi	Correlation	-	-	-
PTAM	FAST	Correlation	✓	-	-
ORB-SLAM	ORB	Descriptor	✓	✓	-
ORB-SLAM3	ORB	Descriptor	✓	✓	✓

stitching the images captured by the endoscope and using structure from motion (SfM) to estimate a 3D model of the bladder. In our work we furthermore aim at assisting the surgeon in navigating within the organs, hence a localization routine must be considered.

2.2. Simultaneous Localization And Mapping: SLAM is a technique to estimate the 3D map of the environment that surrounds the camera and at the same time the pose of the camera relative to this map. Visual SLAM (vSLAM) uses only the information captured by passive cameras to estimate the map and the poses. MonoSLAM (Davison et al. 2007) was the first system that solved the monocular SLAM problem, using an extended Kalman filter (EKF) and Shi-Tomasi features for tracking.

Accumulating the information of all frames, however, is challenging for long-term computation and accuracy. Likewise, the gold standard bundle adjustment (BA) is computationally unfeasible in real time. The use of a sparse set of representative frames (called keyframes) for BA makes it possible to boost both efficiency and accuracy. PTAM (Klein and Murray 2007) introduced this keyframe Bundle Adjustment approach (kBA), running two threads in parallel to perform local tracking and mapping. It uses FAST (Rosten and Drummond 2006) features and matches them using correlation patches. A more current approach of kBA is ORB-SLAM (Mur-Artal et al. 2015), which uses ORB features instead (Rublee et al. 2011). It has map reusability, loop closure and relocalization capabilities. In particular, (Mur-Artal et al. 2015) only considers monocular cameras, yet ORB-SLAM2 (Mur-Artal and Tardós 2017) allows for stereo and RGB-D cameras. Lastly, ORB-SLAM3 (Campos et al. 2021) included visual-inertial SLAM as well as multiple maps, to eventually merge maps after loop closure. Furthermore, ORB-SLAM3 is thoroughly researched for use in laparoscopy (Mahmoud et al. 2016, 2017, 2018; Song et al. 2018). These approaches focus on increasing the number of features for SLAM tracking; however, in our work, we focus on their quality and recognition rate by additional image pre-processing steps and the use of more advanced feature descriptors.

Another major challenge when using SLAM in soft tissue surgery (e.g., MIS, fURS) is the expected deformation of organs due to manipulation or the breathing and heart-beat motions. Most SLAM systems assume a rigid environment. The methods DefSLAM (Lamarca et al. 2020) and SD-def-SLAM (Rodríguez et al. 2020) judiciously combine Shape-from-Template (SfT) (Bartoli et al. 2015) and Non-Rigid Structure-from-Motion (NRSfM) techniques to model the deformations and will be considered by us in future work.

In this paper we propose a monocular SLAM system enhancing ORB-SLAM3 to aid the surgeon to navigate in fURS procedures without the need of any extra sensors. As it can be seen in Table 1, ORB-SLAM is one of the most comprehensive monocular SLAM systems. We decided to use (Campos et al. 2021) as base to our system since it is an open source state-of-the-art SLAM system, that works in real time and is robust in rigid environments. Besides, it includes the map merging capability to recover from

getting lost after a sudden movement of the camera.

2.3. Salient Points of Interest to Track: A core method to SLAM is detection and tracking of natural, salient points or regions that are rigidly attached to the scene, which are called features. Due to their proven performance, in this work we focus on the three following implementations:

ORB (Oriented FAST and Rotated BRIEF) features (Rublee et al. 2011) are a fusion between the FAST detector (Rosten and Drummond 2006) and the BRIEF descriptor (Calonder et al. 2010). First, FAST corners are detected in each level of the image pyramid using a Harris Corner score to sort the detected features. Then, a modified version of the BRIEF algorithm is used to compute rotation invariant descriptors. ORB is in theory invariant to scale, rotation, and limited affine changes. This is the feature originally used in (Campos et al. 2021).

A-KAZE features (Alcantarilla et al. 2012), (Alcantarilla and Solutions 2011) are detected in a non-linear scalar space in which the noise is reduced in each scale such that distinct points are selected. Then, the Modified-Local Difference Binary (M-LDB) descriptor is used. This is based on LDB (Yang and Cheng 2012), which follows the same principle as BRIEF, except for using the average of areas instead of the binary test between single pixels. A-KAZE is in theory invariant to scale, rotations, and limited affine transformations. Furthermore, it is more distinctive at varying scales due to the non-linear scalar space used (Tareen and Saleem 2018). We expect this to be useful in fURS sequences since the non-linear scalar space will make blurring adaptive to image features, reducing the noise of the image but keeping the object boundaries. We utilize a fast CPU A-KAZE implementation based on https://github.com/h2suzuki/fast_akaze.

SuperPoint (DeTone et al. 2018) is a self-supervised, learned framework for detecting and describing points of interest. It consists of a fully convolutional network architecture including two heads: an interest point decoder and a descriptor decoder. In order to obtain training data without the need of manual labeling, the authors optimize it for a simple set of basic shapes with known expected feature locations; they then distort these and their locations (typically with planar homographies) such that the network learns detection and correspondence search for all distorted and augmented shapes. In literature, SuperPoint has been able to obtain a high number of matches in laparoscopic sequences (Barbed et al. 2021). The input for SuperPoint is the pre-processed image (CLAHE) since after testing both inputs (gray scale and pre-processed image) SuperPoint is able to extract more features in the pre-processed image.

3. The Extended ORB-SLAM3 for Ureteroscopy

As discussed in Sec. 2, we identified ORB-SLAM3 as the most promising SLAM system to solve the challenges of endoscopy in the urinary tract. In this section we detail on the extensions that we realize for ORB-SLAM3 to deliver seamless tracking in this challenging scenario. First, we include a pre-processing step to increase the contrast of the images as well as to compute an image mask to avoid detecting features in potential regions of reflections and to avoid the padding borders of the image. Furthermore, we critically adapt parameters on ORB-SLAM3 (e.g., by decreasing the keypoints thresholds in a ratio of 0.6) to cope with the challenging scenario of tracking the low amount of features typical in fURS. Finally, we evaluate the features types detailed in Sec. 2 to increase the accuracy and robustness of the system.

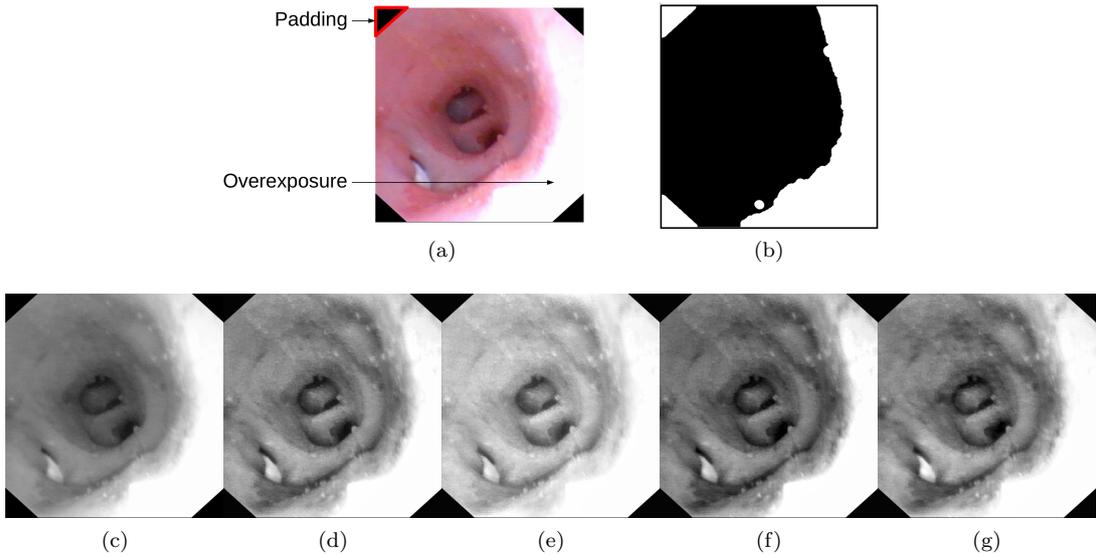


Figure 1. (a) Original image I , (b) mask against reflections and padding, (c) I in grayscale, (d) I after applying CLAHE (I_{CLAHE}) in grayscale, (e)-(g) red, green, and blue channels of I_{CLAHE} .

For pre-processing we apply **Contrast Limit Adaptive Histogram Equalization (CLAHE)** (Zuiderveld 1994) in the L channel of the Lab color space of the image due to the non-uniform illumination typical of fURS – parts of the frame are overexposed, causing reflections, and others are too dark. CLAHE effectively yields a more uniform illumination and improves its contrast (Fig. 4b).

After applying CLAHE, we use the green channel of the RGB color space, since it best preserves contrast in fURS images (cf. Figs. 4c, 4d and 4e). This pre-processing step makes the system more robust against sudden illumination changes typical of fURS procedures due to the proximity of the endoscope’s light to the organ walls and the fast motion of the endoscope.

Next, we compute and apply a **mask to avoid detecting features** on reflections as well as on the padding borders of the endoscopic image (Fig. 1b). This mask is computed using a threshold on the green channel of the image to detect bright pixels which value is between 200 and 255 for specular reflections and darker pixels which value is between 0 and 10 for image padding. It is critical to avoid matching points in such regions since these are not valid 3D quasi-static points. Note that in the context of SLAM these would not only endanger mapping accuracy but also the estimation of the camera pose.

The major modification of the standard ORB-SLAM3 pipeline has been the replacement of the feature detector and descriptor module with the different implementations introduced in Sec. 2. To recap, we decided to evaluate these features since ORB (Rublee et al. 2011) is the one used in ORB-SLAM3, A-KAZE (Alcantarilla and Solutions 2011) uses a non-linear scalar space to make blurring adaptive to image features that can be beneficial in fURS, and SuperPoint (DeTone et al. 2018) since it is a learned feature detector and descriptor which has been proven to deliver a high amount of features in laparoscopic sequences (Barbed et al. 2021).

Lastly, we tackle the lack of natural features typical of ureteroscopy. First, the endoscope has a small field of view; second, its working distance is very short w.r.t. the tissue due to the narrowness of the urinary tract; third, some organs lack of features

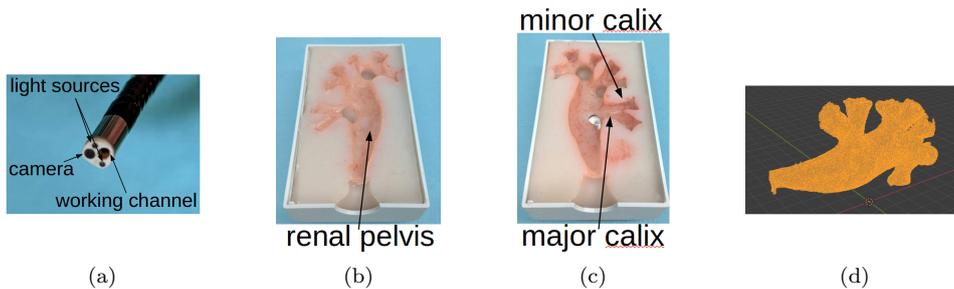


Figure 2. (a) Tip of the endoscope with the monocular camera, the two light sources, and a working channel. Upper (b) and lower (c) halves of the kidney phantom. (c) 3D model of the kidney’s phantom.

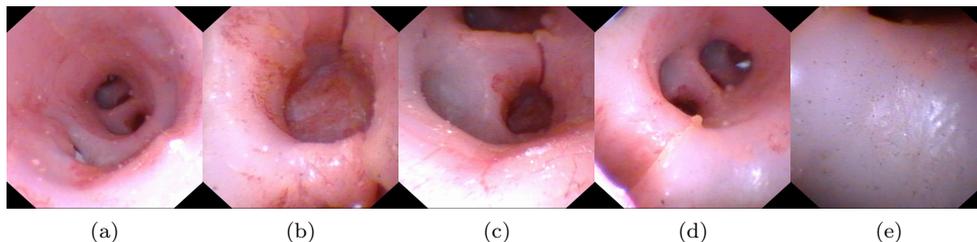


Figure 3. Images from the kidney phantom dataset.

or texture. We therefore had to adapt the parameters of ORB-SLAM3 to ensure a minimum of features that enable seamless tracking of the endoscope pose. Specifically, we reduced the thresholds related with the amount of matched points.

4. Experimental Results and Discussion

A dataset has been recorded using a commercially available flexible ureteroscope Olympus URF-V¹ and a kidney phantom² (60mm×100mm×20mm). To obtain ground truth data, the ureteroscope is electromagnetically tracked in 6 degree of freedom by the NDI Aurora system³. Furthermore, a laser scan was conducted to obtain a 3D model of the kidney phantom.

The dataset consists of five video sequences (Fig. 3). Four of them – k_{01} , k_{02} , k_{03} and k_{04} – are exploration sequences in which the endoscope is introduced in all the cavities of the kidney phantom. The velocity at which the endoscope is moved increased sequentially with the sequences and hence, the difficulty. The sequence k_{05} is captured to evaluate the relocalization module. In this sequence, the endoscope is extracted and re-inserted several times after a small exploration. Additionally, two bladder sequences were recorded by surgeons in a bladder polyp removal procedure (Fig. 4). In these sequences, the bladder is explored and tools, polyps, bubbles and impurities floating in the water can be seen. All experiments were performed with a Intel(R) Core(TM) i7-10700 CPU @ 2.90Hz, 16 cores and 32GB RAM.

For the kidney phantom sequences the following measurements have been evaluated: tracking update rate f_{track} in [fps], Absolute Trajectory Error (ATE) ϵ_{traj} in

¹Olympus Deutschland GmbH, Hamburg, Germany

²SAMED GmbH, Dresden, Germany

³NDI, Waterloo, Ontario, Canada

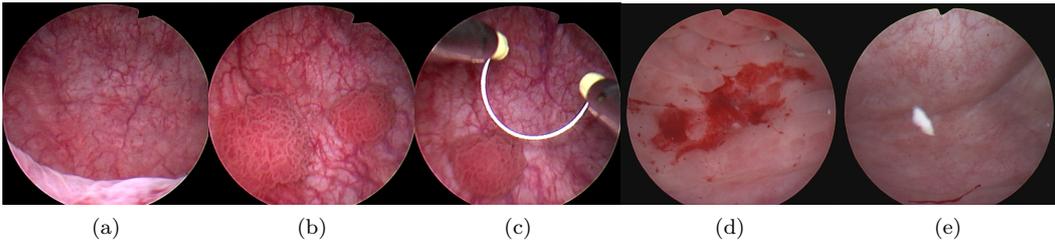


Figure 4. Images from the bladder dataset.

Table 2. Median of f_{track} , n_{kpts} and r_{matched} over 5 runs of the SLAM system.

	ORB-SLAM3 (Campos et al. 2021)	PA (ORB)	PA (A-KAZE)	PA (SuperPoint)
f_{track}	63 fps	35 fps	33 fps	-
n_{kpts}	2214	3303	1723	745
r_{matched}	5.50%	7.94%	21.58%	7.24%

[mm], average number of keypoints detected per image n_{kpts} , Root-Mean Square Error (RMSE) of the map ϵ_{map} in [mm] and ratio of frames providing tracking over the total number of frames in the sequence r_{track} . A requirement for a frame to provide tracking is to be able to find matches between the keypoints detected in the current frame and the map points projected on the current frame and to be able to estimate a pose with those matches. Besides, to evaluate the matching, the ratio r_{matched} between the projected map points successfully matched with the detected keypoints in a new frame over n_{kpts} is calculated. For the bladder sequences b_1 and b_2 , only r_{track} has been computed due to the lack of ground truth. Each experiment has been executed 5 times to compensate for RANSAC by eventually taking the median. In general all performance metrics will be evaluated corresponding to the used features (cf. Sec. 2) and to the original implementation of ORB-SLAM3.

In Table 2, f_{track} , n_{kpts} and r_{matched} are presented. In this case, the goal is to have a high number of keypoints n_{kpts} but not too high since it would in turn decrease f_{track} . Finding a balance between these two values depends on the application. Comparing PA(ORB) (Proposed Approach) and ORB-SLAM3, it can be seen that due to the use of the pre-processing step, PA(ORB) is able to detect more features. Also the adaptation of the ORB-SLAM3 parameters, together with the pre-processing step, allows ORB to have higher r_{matched} . On the other hand, f_{track} is lower. PA(A-KAZE) is able to match the highest percentage of features at a similar f_{track} as PA(ORB). The f_{track} of PA(SuperPoint) is not shown in this table since it is not a fair comparison since the features are computed offline and it is not able to run in real time.

To evaluate the accuracy of the estimated map, first the point cloud of the ground truth (cf. Fig. 2d) and the estimated 3D sparse map are manually aligned to get an initial seed for the iterative closest point (ICP) algorithm to refine this alignment. Then the RMSE error ϵ_{map} between the matched 3D points is computed. Furthermore, the ATE ϵ_{traj} is computed to evaluate the trajectory as proposed in (Sturm et al. 2012). The ATE calculates the RMSE of all global positions \mathbf{x}_t of the frames of the estimated trajectory w.r.t. the ground truth correspondences $\tilde{\mathbf{x}}_t$, after both trajectories have been aligned ($\epsilon_{\text{traj}} = \sqrt{\sum_{t=0}^N (\mathbf{x}_t - \tilde{\mathbf{x}}_t)^2 / N}$, being N the number of frames of the trajectory). For fair comparison, these errors have been computed only for those trajectories with

Table 3. Median of r_{track} over 5 runs of the SLAM system.

Seq.	ORB-SLAM3	PA(ORB)	PA(A-KAZE)	PA(SuperPoint)
k_{01}	67.61%	80.97%	96.50%	80.60%
k_{02}	27.22%	75.52%	87.92%	55.72%
k_{03}	40.65%	51.00%	90.21%	71.34%
k_{04}	4.28%	52.42%	88.63%	21.27%
k_{05}	11.00%	22.20%	76.10%	30.77%
b_{01}	42.50%	60.17%	82.83%	45.33%
b_{02}	14.14%	49.90%	70.88%	53.37%

Table 4. Median of ϵ_{map} and ϵ_{traj} over 5 runs of the SLAM system.

Seq.	PA(ORB)		PA(A-KAZE)		PA(SuperPoint)		GT
	ϵ_{map}	ϵ_{traj}	ϵ_{map}	ϵ_{traj}	ϵ_{map}	ϵ_{traj}	l
k_{01}	1.96	1.84	1.3	1.13	1.56	2.58	1185.83
k_{02}	1.92	2.42	1.46	1.74	-	-	987.95
k_{03}	-	-	2.17	2.48	2.38	4.48	1250.53
k_{04}	-	-	2.53	1.74	-	-	1490.15
k_{05}	-	-	2.02	1.89	-	-	2547.99

$r_{\text{track}} \geq 70\%$. Table 2 shows r_{track} and 4 shows ϵ_{map} and ϵ_{traj} together with the length l in [mm] of each trajectory. Utilizing A-KAZE shows that r_{track} is always highest for all the sequences whereas ϵ_{map} and ϵ_{traj} remain lowest. One of the main reasons is its higher robustness due to a higher r_{matched} , allowing it to continue tracking during longer trajectories. Also for real bladder sequences the extended SLAM system, using A-KAZE features, is capable of a high tracking coverage of over 70%.

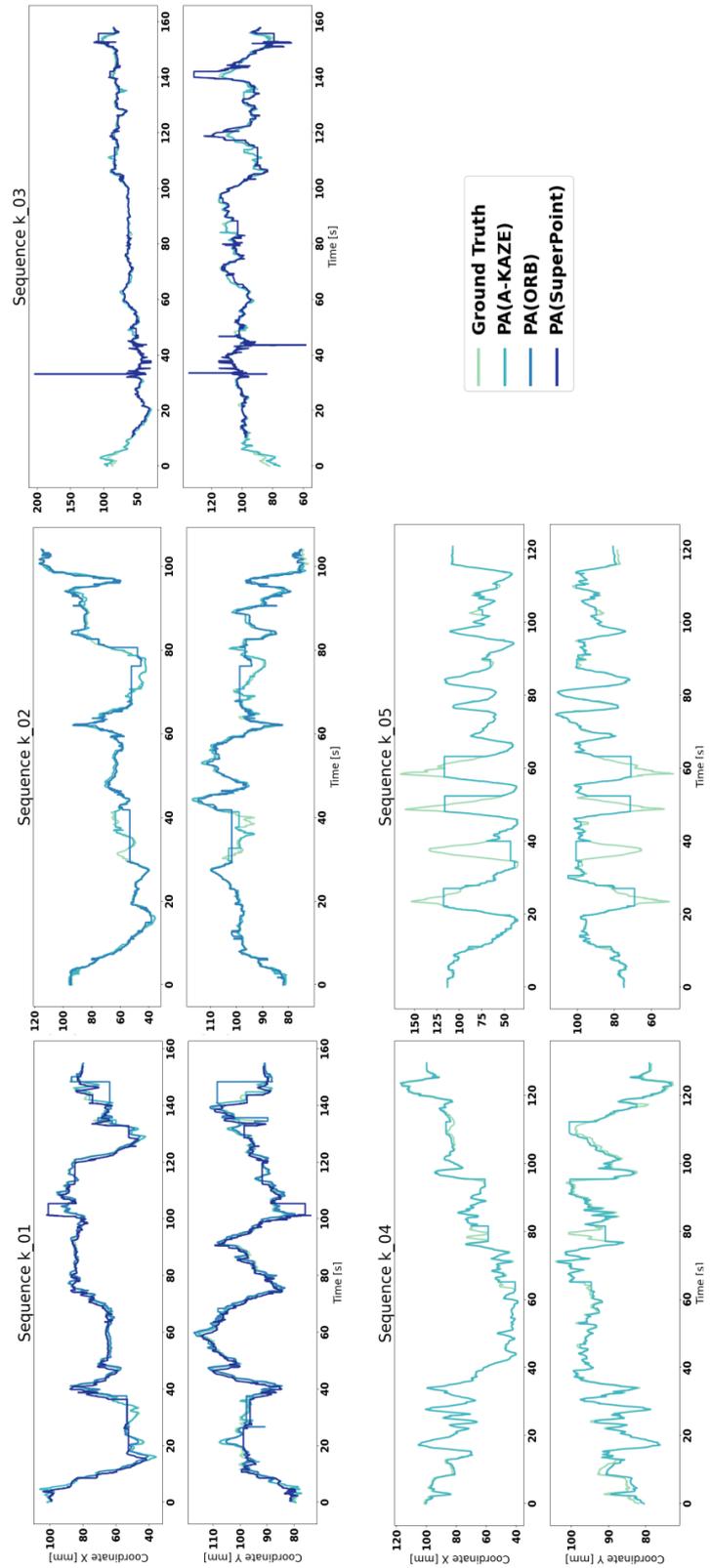
In figure 5 a comparison of the estimated trajectories ($r_{\text{track}} \geq 70\%$) and the ground truth are shown. The straight horizontal lines represent the frames of the sequences in which the system was lost. In sequence k_{05} (relocalization sequence) it can be seen when the endoscope is extracted (low coordinate Y values) and how using A-KAZE features the system is able to relocalize once that the endoscope is re-inserted.

5. Conclusions

In this paper we propose to track the pose of an endoscope during minimally invasive surgery (MIS), specifically flexible ureteroscopy (fURS), using its own video feed. In addition, we generate a sparse 3D model of compartments of the urinary tract for 3D inspection. This is realized by a vSLAM system.

We choose the state-of-the-art ORB-SLAM3 software and first extend it by a pre-processing step in which we apply CLAHE to increase the contrast of the image, we compute and apply a mask to avoid detecting features on reflections and padding borders, and we adapt the parameters of ORB-SLAM3 to cope with less features typical of fURS sequences. We show that with these pre-processing steps ORB-SLAM3 is able to detect and match more features.

The main extension of ORB-SLAM3 has however been the modification of its feature detector and descriptor module with other modern approaches. We have evaluated ORB, A-KAZE, and SuperPoint features. The evaluation has shown that the substitution of ORB features by A-KAZE features increases the accuracy of ORB-SLAM3



(a)

Figure 5. Trajectory comparison.

in fURS sequences Also it allows ORB-SLAM3 to track longer trajectories without losing track.

We recorded two novel datasets for ureteroscopy (using a kidney phantom, its 3D model, and ground-truth endoscope pose) and bladder polyp removal. With these datasets we were able to evaluate our work.

Future work will include designing an AR based user interface aiding the surgeon navigating through the urinary tract and exploiting the real-time capability to enhance the control of a robotic assistance system, which is under development (Schlenk et al. 2022). Besides, deformable SLAM will also be taken into account.

References

- Alcantarilla PF, Bartoli A, Davison AJ. 2012. Kaze features. In: European conference on computer vision. Springer. p. 214–227.
- Alcantarilla PF, Solutions T. 2011. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans Patt Anal Mach Intell.* 34(7):1281–1298.
- Barbed OL, Chadebecq F, Morlana J, Montiel JM, Murillo AC. 2021. Superpoint features in endoscopy.
- Bartoli A, Gérard Y, Chadebecq F, Collins T, Pizarro D. 2015. Shape-from-template. *IEEE transactions on pattern analysis and machine intelligence.* 37(10):2099–2118.
- Bergen T, Wittenberg T. 2014. Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods. *IEEE journal of biomedical and health informatics.* 20(1):304–321.
- Calonder M, Lepetit V, Strecha C, Fua P. 2010. Brief: Binary robust independent elementary features. In: European conference on computer vision. Springer. p. 778–792.
- Campos C, Elvira R, Rodríguez JJG, Montiel JM, Tardós JD. 2021. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics.* 37(6):1874–1890.
- Davison AJ, Reid ID, Molton ND, Stasse O. 2007. Monoslam: Real-time single camera slam. *IEEE transactions on pattern analysis and machine intelligence.* 29(6):1052–1067.
- DeTone D, Malisiewicz T, Rabinovich A. 2018. Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. p. 224–236.
- Geraghty RM, Jones P, Somani BK. 2017. Worldwide trends of urinary stone disease treatment over the last two decades: a systematic review. *Journal of endourology.* 31(6):547–556.
- Klein G, Murray D. 2007. Parallel tracking and mapping for small ar workspaces. In: 2007 6th IEEE and ACM international symposium on mixed and augmented reality. IEEE. p. 225–234.
- Klodmann J, Schlenk C, Hellings-Kuß A, Bahls T, Unterhinninghofen R, Albu-Schäffer A, Hirzinger G. 2021. An introduction to robotically assisted surgical systems: Current developments and focus areas of research. *Current Robotics Reports.* 2(3):321–332.
- Lamarca J, Parashar S, Bartoli A, Montiel J. 2020. Defslam: Tracking and mapping of deforming scenes from monocular sequences. *IEEE Transactions on robotics.* 37(1):291–303.
- Mahmoud N, Cirauqui I, Hostettler A, Doignon C, Soler L, Marescaux J, Montiel J. 2016. Orbslam-based endoscope tracking and 3d reconstruction. In: International workshop on computer-assisted and robotic endoscopy. Springer. p. 72–83.
- Mahmoud N, Collins T, Hostettler A, Soler L, Doignon C, Montiel JMM. 2018. Live tracking and dense reconstruction for handheld monocular endoscopy. *IEEE transactions on medical imaging.* 38(1):79–89.
- Mahmoud N, Hostettler A, Collins T, Soler L, Doignon C, Montiel J. 2017. Slam based quasi dense reconstruction for minimally invasive surgery scenes. arXiv preprint arXiv:170509107.
- Mur-Artal R, Montiel JMM, Tardos JD. 2015. Orb-slam: a versatile and accurate monocular

- slam system. *IEEE transactions on robotics*. 31(5):1147–1163.
- Mur-Artal R, Tardós JD. 2017. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE transactions on robotics*. 33(5):1255–1262.
- Pentek Q, Hein S, Miernik A, Reiterer A. 2018. Image-based 3d surface approximation of the bladder using structure-from-motion for enhanced cystoscopy based on phantom data. *Biomedical Engineering/Biomedizinische Technik*. 63(4):461–466.
- Rassweiler JJ, Autorino R, Klein J, Mottrie A, Goezen AS, Stolzenburg JU, Rha KH, Schurr M, Kaouk J, Patel V, et al. 2017. Future of robotic surgery in urology. *BJU international*. 120(6):822–841.
- Rodríguez JJG, Lamarca J, Morlana J, Tardós JD, Montiel JM. 2020. Sd-defslam: Semi-direct monocular slam for deformable and intracorporeal scenes. *arXiv preprint arXiv:201009409*.
- Rosten E, Drummond T. 2006. Machine learning for high-speed corner detection. In: *European conference on computer vision*. Springer. p. 430–443.
- Rublee E, Rabaud V, Konolige K, Bradski G. 2011. Orb: An efficient alternative to sift or surf. In: *2011 International conference on computer vision*. Ieee. p. 2564–2571.
- Schlenk C, Klodmann J, Hagmann K, Kolb A, Hellings-Kuß A, Steidle F, Schoeb DS, Jürgens T, Miernik A, Albu-Schäffer A. 2022. A robotic system for solo surgery in flexible ureterorenoscopy. *accepted for publication in IEEE RA-L*.
- Song J, Wang J, Zhao L, Huang S, Dissanayake G. 2018. Mis-slam: Real-time large-scale dense deformable slam system in minimal invasive surgery based on heterogeneous computing. *IEEE Robotics and Automation Letters*. 3(4):4068–4075.
- Sturm J, Engelhard N, Endres F, Burgard W, Cremers D. 2012. A benchmark for the evaluation of rgb-d slam systems. In: *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE. p. 573–580.
- Tareen SAK, Saleem Z. 2018. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In: *2018 International conference on computing, mathematics and engineering technologies (iCoMET)*. IEEE. p. 1–10.
- Yang X, Cheng KT. 2012. Ldb: An ultra-fast feature for scalable augmented reality on mobile devices. In: *2012 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE. p. 49–57.
- Zuiderveld K. 1994. Contrast limited adaptive histogram equalization. *Graphics gems*:474–485.