



Proceeding of the 11th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery

April 25 - 27, 2022

Napoli, Italy

CRAS

2022

**Proceeding of the 11th Joint Workshop on New
Technologies for Computer/Robot Assisted Surgery**

Proceedings of the 11th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery,
Napoli, Italy, April 25 - 27, 2022

This work is subjected to copyright. It is published as an open-access publication under the “Creative Commons Attribution 4.0 International” license.

www.cras-eu.org

Robotic surgery is one of the most appealing fields of modern robotics. With over 3 decades of history, more than 3.800 systems installed worldwide, and over 600.000 robot-assisted interventions conducted per year, the field of robotic surgery is well-established. Despite these impressive figures and increasing popularity in research labs all over the world, the list of technological advances that made it into the operating room (OR) during this last decade remains limited. Long-awaited techniques such as 3D reconstruction, motion compensation, virtual guidance, haptic feedback, understudy in many labs all over the planet did not make their appearance into the market yet. CRAS strives to overcome this status-quo by strengthening the collaboration between the different research groups to boost the efficacy and shorten the development cycle.

CRAS 2022 is the 11th edition of this successful event. Formerly labeled as a workshop, from this 10th edition onwards CRAS has graduated and will continue to move forward as a ‘conference’.

The 11th edition of CRAS will take place from April 25-27, 2022, at Centro Congressi Federico II in Napoli, Italy. Further information about the current health emergency due to the COVID-19 pandemic will be provided on this website.

CRAS 2022 Support



UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II



Venue

CRAS 2022 was held in Centro Congressi Federico II, Via Partenope 36, Napoli, Italy. The metropolitan city of Napoli can be reached in several ways.



About Napoli



The Province of Napoli (Naples) is a magical place where colors, flavors, culture, and history are intertwined in a charming mix of knowledge, joy, and fun. The area is loomed over by Mt. Vesuvius and overlooks a marvelous bay, whose beauty has served as driving inspiration for many an artist. The best characteristic of Napoli is that it has still held on to much of its character and old-world charm. Layers of history are visible as you wander Napoli's narrow, cobbled streets, dipping in and out of the many bars, restaurants, and galleries started by the new generation of entrepreneurs, chefs, and artists reinvigorating the city once again. Of course, it also has the advantage of having one of the most stunning backdrops in the world. With Vesuvius to the east, the ancient port of Pozzuoli to the west, and the timeless islands of Ischia, Procida, and Capri nestled in the glittering bay, it's no wonder the Romans nicknamed the area Campania Felix – the happy land.

CRAS 2022 Organizing Committee

The organizational board of CRAS consists - in alphabetical order - out of the below members.

Salih Abdelaziz LIRMM	Alperen Acemoglu IIT	Kaspar Althoefer Queen Mary London	F.R. Y Baena ICL
Christos Bergeles KCL	Gianni Borghesan KU Leuven	Marta Capiluppi Univ. Verona	Alicia Casals UPC
Zhuoqi Cheng SDU	Gastone Ciuti SSSA	Giulio Dagnino ICL	Jenny Dankelman TU Delft
Diego Dall’Alba Univ. Verona	Brian Davis ICL - IIT	Elena De Momi POLIMI	Kathleen Denis KU Leuven
Nikhil Despande IIT	Giancarlo Ferrigno POLIMI	Fanny Ficuciello Univ. Napoli	Paolo Fiorini Univ. Verona
Ioannis Georgilas Univ. of Bath	Edward Grant NCSU	Yassine Haddab LIRRM	Tamas Haidegger Obuda Univ.
Benoit Herman UCLouvain	Albert Hernansanz UPC	Mohammad H. Hamedani Univ. Napoli	Cristina Iacono Univ. Napoli
Pierre Jannin U.Rennes	Chao Liu LIRMM - CNRS	Leonardo Mattos IIT	E. Mazomenos Leeds
Arianna Menciassi SSSA	Rocco Moccia Univ. Napoli	Sara Moccia Univ. Pol. delle Marche, IIT	Guillaume Morel ISIR
Riccardo Muradore Univ. Verona	George Mylonas UCL	Florent Nageotte Ben-Gurion Univ.	Ilana Nisky Ben-Gurion Univ.
Mouloud Ourak KU Leuven	Sebastien Ourselin KCL	Nicolas Padoy UNISTRA	Veronica Penza IIT
Philippe Poignet LIRRM	Pierre Renaud UNISTRA	Benoit Rosa UNISTRA	Jerome Szewczyk ISIR
Bruno Siciliano Univ. Napoli	Giannarou Stamatia ICL	Stefanie Speidel University Dresden	Danail Stoyanov UCL
Selene Tognarelli SSSA	E. Vander Poorten KU Leuven	Tom Vercauteren KCL	Helge Wurdemann UCL
Nabil Zemiti LIRRM			

Program

In this joint conference, all science and engineering presentations from CRAS were embedded into closely related medically oriented sessions. The session topics included:

- Minimally invasive surgery
- Robotic surgery
- Surgical staplers and suturing
- Trauma and resuscitation
- Endoscopy and endoscopic surgery
- Microsurgery
- Endovascular surgery
- Rehabilitation and enhanced recovery after surgery
- Imaging and diagnosis
- Health Technology Assessment (HTA)
- High energy surgical devices
- Surgical training
- Tracers, surgical guidance and augmented reality
- Transplantation and regenerative medicine
- "Low cost" technologies
- New technologies, novel methods and future applications
- New technologies demonstrations and competition

For details on the CRAS+SPIGC 2022 conference program, please refer to CRAS website <https://cras-eu.org/program-2/>.

Open-Access License Information

All articles included in these proceedings are published as open-access publication under the “***Creative Commons Attribution 4.0 International***” license.

This means that the materials included here can be shared and adapted for any purpose. However, you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

For more details on the “Creative Commons Attribution 4.0 International” license, please refer to: <https://creativecommons.org/licenses/by/4.0/>

CRAS 2022 Abstract index

Robot Assisted Electrical Impedance Scanning for tissue detection and imaging	10
Real-time retina and instrument segmentation of OCT images based on GPU accelerated inference	12
Controlling of a Micropositioning Piezoelectric Actuator using an LSTM Network for Robotic Eye Surgery	14
Experimental assessment of a fabric-based wearable haptic softness display for the localization of uterine leiomyomas in robot-assisted surgical palpation tasks	16
Towards a Handheld Robotic Instrument for Minimally Invasive Neurosurgery	18
A Temporal Learning Approach to Inpainting Endoscopic Specularities and Its effect on Image Correspondence.....	20
Automatic Robotic Scanning for real-time 3D Ultrasound Reconstruction in Spine Surgery	22
Detector-Free Dense Feature Matching for Fetoscopic Mosaicking	24
Smartawake: a real-time decision support system for intraoperative assessment in awake surgery	26
Recovering 3D information of human soft tissue using stereo endoscopic images	27
Autonomous Pick and Place using RL: Demonstrations Optimization	29
Deep Learning-based Plane Pose Regression: a First Step Towards Training and Guidance in Freehand Obstetric Ultrasound	31
OCT as tool for laser ablation monitoring applied to cholesteatoma.....	33
Augmented reality-assisted markerless hip centre measurement for total knee arthroplasty	35
WAYNet: Weakly-Supervised Multi-Task Anastomoses Detection in Fetoscopy	37
An optimization method for surgical targeting planning	39
Surgical knot training in ophthalmic surgery: Skill assessment with eye-tracking	41
Fast and Automatic Optical 3D Cameras Calibration for Contactless Surface Registration in Computer Assisted Surgery	43
Tool motion estimation from internal pressure variation in a parallel, fluid-driven endoluminal soft robot	45
Acoustic or Visual? A User Study on the Effectiveness of Sensor-based Feedback to Train Tissue Manipulation in Robotic Surgery	47
Experimental Evaluation of the SoftSCREEN Capsule System in a Colon Phantom	49
Unconstrained Rotation for Control of Concentric Tube Robots with Deep Reinforcement Learning.....	51
Evaluation of a driving mechanism for percutaneous needle insertion	53
A Linguistic Comparison Between Textual Datasets to Assess the Complexity of Surgical Robotic Procedural Descriptions	55

Predicting Image Relationships with Image Box Embeddings During Endoscopy	57
Accuracy Assessment of a Forward Kinematic Quasi-Static Model of a Concentric Tube Robot for Spina Bifida Repair.....	59
A Modular Organ Perfusion Research Platform for Computer Assisted Imaging and Intervention	61
Concurrent Hierarchical Autonomous Surgical State Estimation during Robot-assisted Surgery	63
Proof-of-Concept Medical Robotic Platform for Endovascular Catheterization	66
Autonomous image guided control of endoscopic orientation for OCT scanning.....	68
3D Reconstruction of Local Anatomy based on 2D Ultrasound for Fetal Laser Surgery.....	70
Unsupervised Denoise and Deblur of OCT Images via Disentangled Representation	72
Study of the curling behaviour of a self-shaping cochlear implant during insertion using a teleoperated robotic system	74
Estimation of a Vessel Side-Branches Model for Robotic Intravascular Navigation	76
One-Shot Boundary Detection Network for Multi-Modal Side-Viewing Imaging	78
Semi-Supervised Placental Vessel Segmentation with Majority Voting for Pseudo-Label Updating.....	80
DESIGN AND EVALUATION OF A HAPTIC CERVIX PHANTOM FOR HYSTEROSCOPY TRAINING	82
DESIGN AND EVALUATION OF A HAPTIC CERVIX PHANTOM FOR HYSTEROSCOPY TRAINING	84
Automatic bone breach detection for spine surgery based on bio-electrical conductivity sensing: Ex-vivo experimental validation.....	86
Force control of the KUKA LBR Med without external force sensor	88
Multi-level-assistance Robotic Platform for Navigation in the Urinary System: Design and Preliminary Tests	90
Real-Time Detection of Laryngeal Squamous Cell Carcinoma via Video Laryngoscopy with WL and NBI using YOLO.....	92
Toward automatic bone breach detection for spine surgery using tissue bio-electrical conductivity sensing	94
Augmented Reality and Robotic Navigation System for Spinal Surgery	96
Surgical image generation for binary tool segmentation	98
Spatio-temporal models for online activity recognition in cataract surgical videos	100
Prototyping a Laparoscopic Smart Trocar for Semi-Autonomous Surgical Robotics	102
Design, realization and preliminary validation of an active physical simulator for the study of pelvic floor damages during childbirth	104
End-of-instrument Stiffness Perception Device Based on Sensory Augmentation for Laparoscopic Surgery.....	106
Online detection of disturbances in model-based weight compensation for robotic endoscope holders..	108

Towards an autonomous robotic platform for percutaneous procedures	110
Hybrid controllers for middle ear surgery: towards efficient cholesteatoma removal.....	112
A posterior landmark-based registration-free method to identify pedicle screw trajectories for robot-based navigation: A proof-of-concept.....	114
Intra-Domain Adaptation for Robust Visual Guidance in Intratympanic Injections	116
A Fabric Soft Robotic Exoskeleton with Novel Elastic Band Integrated Actuators for Hand Rehabilitation	118
Electrical Impedance Tomography Sensing for Soft Continuum Surgical Manipulators.....	120
A Linguistic Comparison Between Textual Datasets to Assess the Complexity of Surgical Robotic Procedural Descriptions	122
Deep learning-based localization of the biliary tract on white-light images acquired during laparoscopic cholecystectomy.....	124
Material Characterization for Magnetic Soft Robots	126
Dual-Arm Platform for Control of Magnetically Actuated Soft Robots.....	128
Fiber Reinforcement for Miniaturized Magnetically Actuated Soft Continuum Robots.....	130
Geometric Reinforcement towards Fully Soft Magnetic Endoscopic Endonasal Surgical Manipulators.	132
Simulated Force Control of the da Vinci PSM for Autonomous Intraoperative Ultrasound Scanning	134
Vision-based Assessment and Adaption of Ultrasound Probe Tissue Coupling during Robotic Surgery	136
Autonomous Tissue Detection in Robotic Assisted Minimally Invasive Surgery	138
Human-Robot Collaboration System Using Virtual Fixture Method and Image-Guided for Dental Implant Surgery	140
Reduced mathematics model of an under-actuated soft hand for Robot-assisted Laparoscopic Surgery: The MUSHA Hand II.....	142

Robot Assisted Electrical Impedance Scanning for tissue detection and imaging

Zhuoqi Cheng, Thiusius Rajeeth Savarimuthu

Mærsk Mc-Kinney Møller Institute, University of Southern Denmark, 5230 Odense, Denmark

INTRODUCTION

Nowadays Robotic assisted Minimally Invasive Surgery (RMIS) is regularly exploited for various medical treatments. During RMIS, it is often challenging for surgeons to detect critical tissue types in the field of view. Particularly, the clinically available RMIS systems have very limited capability in localizing critical subsurface tissues such as early stage cancer, lymph nodes and blood vessels. External sensing probes, such as ultrasound transducer or palpation probes are thus developed to assist in this task. Since different tissue types can have very different electrical characteristics, Electrical Bio-Impedance (EBI) sensing technology is integrated to the bipolar forceps of RMIS system to provide real-time and on-site tissue identification [1]. However, because of the large contact impedance, the bipolar EBI measurement system is very sensitive to the pressing depth of the forceps, which is very difficult to control accurately. Also, the bipolar sensing method has concentrated sensitivity about the forceps tips, constraining its sensing capability and performance. To overcome these limitations, an improved robotic EBI sensing system based on a tripolar configuration, named Robot Assisted Electrical Impedance Scanning (RAEIS), is developed. In addition to accurate tissue identification [2], the new system is able to perform flexible, autonomous scanning, image non-homogeneous tissue region and potentially detect subsurface critical tissue.

MATERIALS AND METHODS

A. Tripolar sensing configuration

The utilized tripolar configuration is constructed by one current source electrode (CSE), one voltage measurement electrode (VME), and one ground electrode (GND). Excitation current I is injected into the tissue via the CSE at point A. If the tissue under measurement is homogeneous, the current flow disperses from the source point radially into the tissue. The tissue surface is assumed to be flat and the tissue conductivity is σ . The electric potential at a position M on the tissue surface can be calculated as

$$V_M = \frac{I}{2\pi\sigma} \left(\frac{1}{d} - \frac{1}{r_0} \right) \quad (1)$$

where d represents the distance of $|AM|$. r_0 is the equivalent radius of CSE and it is difficult to obtain. Thus, differential voltage is used by using the VME to measure the electric potential at 2 different positions M and N . By this means, the conductivity of tissue can be calculated as

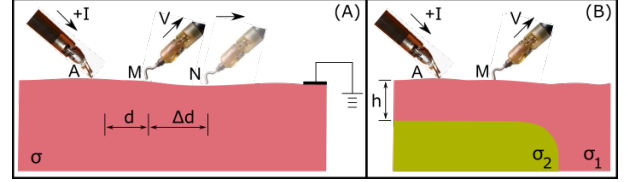


Fig. 1: Modeling the tripolar configuration based on two robotic instruments when the sensing is conducted on homogeneous material (A) and double layer material (B).

Eq. (2), where Δd denotes the distance of $|MN|$.

$$\sigma = \frac{I}{2\pi V_{MN}} \left(\frac{1}{d} - \frac{1}{d + \Delta d} \right) \quad (2)$$

Assuming that the material under measurement has a superficial layer of material σ_1 and a subsurface layer of material σ_2 as shown in Fig. 1(B), a part of the current reaching the interface between 2 materials would reflect. The intensity of the reflected current is reduced by a factor $k = \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2}$. Given the thickness of the top layer as h , the voltage difference between M and N can be calculated as:

$$V_{MN}^* = \frac{I}{2\pi\sigma_1} \left[\frac{\Delta d}{d \cdot \Delta d} + 2 \sum_{n=1}^{\infty} \left(\frac{k^n}{\sqrt{d^2 + (2nh)^2}} - \frac{k^n}{\sqrt{(d + \Delta d)^2 + (2nh)^2}} \right) \right] \quad (3)$$

By substituting V_{MN} in Eq.(2) with Eq.(3), the calculated conductivity is a function of d , and it can reflect the present of the subsurface tissue. For this, an artificial neural network (ANN) is used. The ANN structure is designed as a simple multilayer perceptrons with one hidden layer. The $\sigma(d)$ values acquired at a position are inputted to the ANN. The final layer consists of a single neuron with a Sigmoid activation to output a likelihood value between 0 and 1.

The proposed system can be implemented on most existing robotic systems for RMIS. Two independent robotic instruments which are used for electrosurgery are used as the CSE and VME. A big GND electrode is attached on the tissue and relatively far from the measurement site. Also, the system requires to use an impedance spectroscopy for current excitation and voltage measurement.

B. Scanning strategy: grid scanning v.s. active search

The RAEIS system can sense the electrical property of a small region for each measurement. To inspect a relatively big area, 2 scanning methods are developed including grid

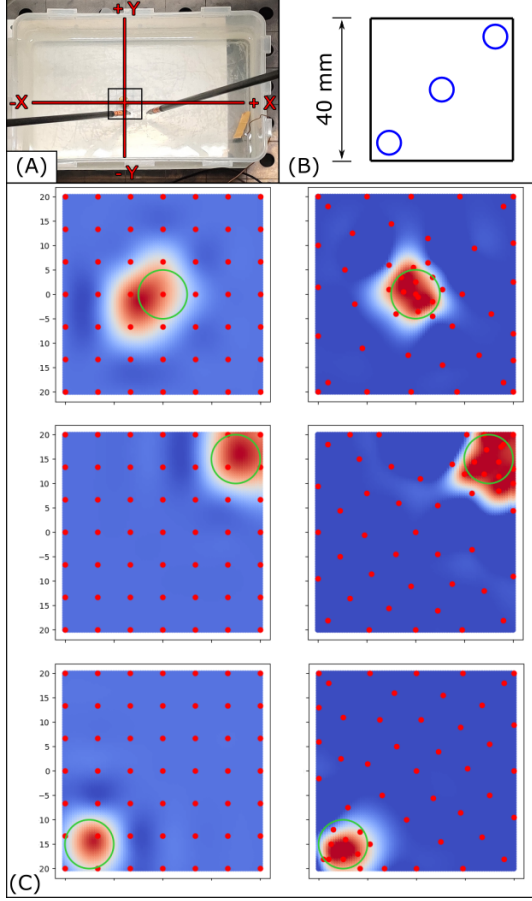


Fig. 2: (A) The water tank experimental setup; (B) Position of the metal object for the experiment; and (C) Results of grid scanning (left) and AS setting (right).

scanning and active search (AS). Specifically, the grid scanning is based on a 2D rectangle mesh over the user-defined area and sampling density. The CSE is controlled to every vertex of the grid and then the VME performs the scanning along the $X+$ direction. Inspired by [3], the AS method aims to increase the sampling number where the target tissue is more likely to present or where sensing information is lacking, and reduce the sampling priority in the region of lower likelihood and higher certainty. The AS method uses a Gaussian Process (GP) to estimate the likelihood distribution of the searching region. Then, a Bayesian optimizer is used for determining the subsequent sensing position automatically.

C. Experiment and results

Experiment was carried out on a water tank setting where tap water was used to simulate the background material and a metal cylinder (10 mm diameter) immersed inside water simulating the target object. Fig. 2(A) shows the water container and the coordinate definition. A 40×40 mm square region was define as the searching area which was highlighted by a black box. A copper plate was used as the GND electrode and it was placed on the right side of the container.

We first performed the data collection for ANN training.

No. of Sample	16	25	36	49
$precision_{Grid}$	6.8%	43.2%	33.8%	69.2%
$recall_{Grid}$	33.3%	66.5%	61.8%	90%
$precision_{AS}$	76.9%	70.7%	68.6%	70.3%
$recall_{AS}$	92.2%	95%	91.4%	91.4%

TABLE I: Precision and recall of 2 scanning methods.

The metal object was placed in the center and a grid setting with a resolution of 2 mm in both X and Y direction was used. The measurements collected above the object were labelled as '1', and '0' for the other data. Different immersed depths (2.5 mm, 5 mm, 7.5 mm and 10 mm) were used and all the measurements were put together to train the ANN model. Subsequently, the metal cylinder was placed in one of the three different locations: (0,0), (-15 mm, 15 mm), and (15 mm, 15 mm) as shown in Fig. 2(B). Both scanning methods were tested. For the grid scanning, measurements were performed on equal steps in both X and Y direction inside the searching area. Different grid density from 4×4 to 7×7 were investigated. For the AS strategy, the GP model was initialised with the ANN predictions at 5 positions including (-18, -18), (-18, 18), (18, -18), (18, 18) and (0, 0). Then the optimiser was queried for a new sampling position. The values collected at this position were used to make a prediction, update the GP model, and so on so forth. To compare the sensing performance between two methods, the AS process was performed until the sampling number reached 49.

The results of 7×7 grid scanning and AS are shown in the left column and right column of Fig. 2(C) respectively. Results of all three conditions are put together for analysis. Table I presents the precision and recall associated to the number of scanning points. To achieve both high precision and recall, the AS requires a much shorter sensing time of 1.9 min (12 sampling points), compared to 7.6 min for the grid scanning method (49 sampling points).

CONCLUSIONS AND DISCUSSION

The experimental results show that both scanning methods can effectively display the non-homogeneous region, but a generally higher precision and recall is achieved by using the AS strategy. Also, the active search method can greatly reduce the number of sampling points compared to a grid based searching method for achieving a similar detection accuracy. Compared to the other medical sensing technologies, the proposed RAEIS technology is non-invasive, fast, effective, and most importantly, without introducing additional instruments to the operational site. [4] provides more technical details and tissue experiments.

REFERENCES

- [1] Z. Cheng and et al., "Design and integration of electrical bio-impedance sensing in surgical robotic tools for tissue identification and display," *Frontiers in Robotics and AI*, 2019.
- [2] —, "Robot assisted electrical impedance scanning for tissue bioimpedance spectroscopy measurement," *Measurement*, 2022.
- [3] H. Salman and et al., "Trajectory-optimized sensing for active search of tissue abnormalities in robotic surgery," *IEEE ICRA*, 2018.
- [4] Z. Cheng and et al., "Active search of subsurface lymph nodes using robot-assisted electrical impedance scanning," *IEEE Transactions on Instrumentation and Measurement*, 2022.

Real-time retina and instrument segmentation of OCT images based on GPU accelerated inference

Pengwei Xu¹, Mouloud Ourak¹, Oscar Caravaca-Mora²,
Gianni Borghesan^{1,3}, Dominiek Reynaerts^{1,4}, and Emmanuel Vander Poorten¹

¹ Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Belgium

² Tearney Lab, Mass General Hospital, Harvard University, United States

³ Core Lab ROB, Flanders Make, Belgium

⁴ Core Lab MaPS, Flanders Make, Belgium

INTRODUCTION

Subretinal injection (SI) is a procedure where the subretinal space in human eyes is accessed for drug delivery and gene therapy purposes [1]. Through subretinal injection these materials can be injected nearby the retinal pigment epithelium cells and the plasma membrane of the photoreceptor. Furthermore, the technique is able to regress the consequent subretinal blebs rapidly [2]. Although important for treating and curing related eye diseases, subretinal injection is very demanding, even for skilled surgeons, due to the microscopic dimensions of the anatomic structures involved as well as due to the delicate nature of the retina. This motivates the development of robot-assisted approaches.

Surgical instrument tracking is a core component for robotic surgery, because it can provide feedback for guidance or closed-loop control [3]. Nowadays, optical coherence tomography (OCT) is gaining popularity as imaging modality in eye surgery. OCT has been used to visualise anatomic structures. Others used it to estimate the distance between the surgical tool and the retina without [4] or with [3] instrument tracking. For intra-operative use of OCT it is vital that the signals are processed fast and reliable.

Deep Learning (DL) inference is the process of using a trained neural network to make predictions against new data. Previous work that focused on accelerating inference of OCT reported already very short inference times namely 3.5ms to segment the retinal layer boundaries [5]. However, performance for joint segmentation of retina and instrument takes much longer. E.g. works from Peng and Sommersperger [3] required at least 13ms per B-scan. This abstract introduces optimization methods for DL inference for segmenting both retina and surgical instruments, aiming to deliver real-time low-latency instrument tracking. First results on an in-silico eye model (Fig.1) showed inference time below 5.5 ms which is less than half of the previous state-of-the-art.

MATERIALS AND METHODS

DL segmentation starts from generating ground truth data for retina and needle segmentation. After that, a U-

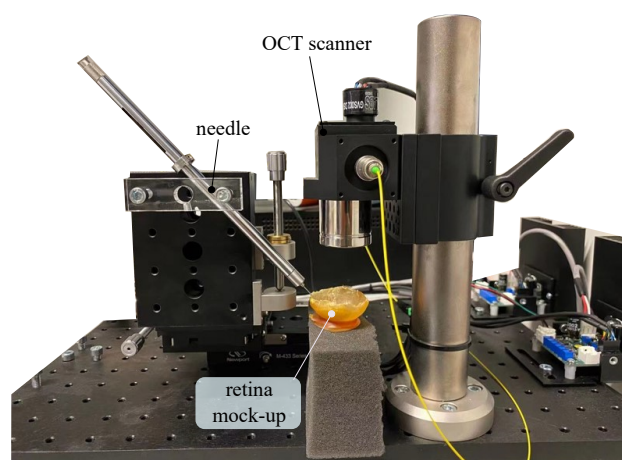


Fig. 1: Experimental setup, a surgical needle is attached to an XYZ positioning stage, an eye model is placed below the OCT scanner. Multiple positions of needle and eye model were set for training and testing.

Net model is trained to classify each pixel of the cross-sectional tomography (B-scan). The B-scan is achieved by laterally combining a series of axial depth OCT scans (A-scan). The OCT pixels are labelled into three classes, namely: retina, needle, and background. Proposed by Ronneberger *et al.* [6], U-Net is widely used for image segmentation, especially in medical applications where the amount of available data is usually limited. It has also been used previously for OCT image segmentation. E.g. Sommersperger reported accuracy of classifying pixels of the B-scan exceeding 95% [3]. During training, the parameters of the neural network are optimized based on the error between prediction and ground truth. After training, inference in milliseconds can be obtained by exploiting GPU acceleration. For real-time inference, a trained neural network, has to be further compressed. The precision of the network parameters may need to be reduced, so that computational cost can be reduced and real-time inference becomes possible.

This work exploited TensorRT, which is a platform for

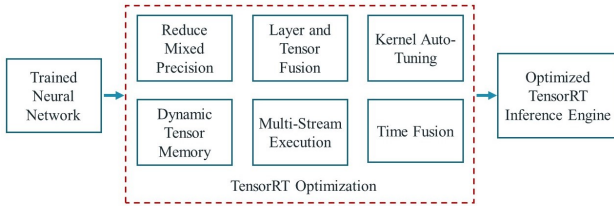


Fig. 2: TensorRT optimization pipeline. During the optimization routine, the throughput is maximized by quantizing models to mixed precision while preserving accuracy. GPU memory and bandwidth are optimized intensively.

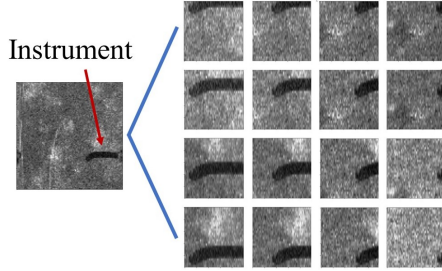


Fig. 3: Example of data augmentation from a dense-resolution OCT volume. Pictures are en face projection views of the OCT volume. The left picture shows the entire volume. The instrument is visible as the darker region. 16 sub-volumes of size $1\text{mm} \times 1\text{mm}$ are extracted from the complete volume $5\text{mm} \times 5\text{mm}$ based on different window positions. Afterwards, 7 equally-spaced B-scans are extracted from each sub-volume, simulating the data that would be acquired during RT approaches.

high-performance DL inference developed by NVIDIA. TensorRT is developed based on CUDA, NVIDIA's parallel programming model, and allows us to significantly accelerate DL inference [5], [7]. A typical optimization workflow of TensorRT is shown in Fig.2. This abstract reports on a TensorRT inference engine that was set up for efficient segmentation of OCT images. High-resolution OCT images were acquired from an eye mock-up with a swept source OCT system which was assembled in our group based on Axsun OCT (SS-OCT, 1310 OEM, Axsun Technologies, MA, US). The system has a 100kHz sampling rate and an axial resolution of $5.6\text{ }\mu\text{m}$. Figure 1 shows the experimental setup. 1000 B-scan images of size 1000×1024 were acquired over a field of view of $5 \times 5\text{ mm}$ at 8 positions. Then the acquired B-scans are augmented into 448 B-scans of size 256×1024 as shown in Fig.3 for training and validation of the U-Net model. A computer with Intel i9-9880H CPU, an NVIDIA Quadro RTX 3000 GPU, and 32 GB of RAM was used.

RESULTS

The original U-Net model and optimized TensorRT inference engine, were tested by segmenting 100 test B-scans for 100 iterations respectively. Both ran on the GPU, using the same FP32 precision. The results show that after optimization (Table I), the inference time for segmenting one B-scan image was reduced as much as 35 times. The

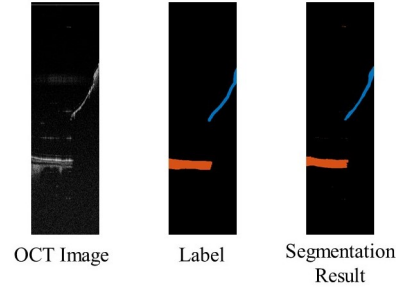


Fig. 4: An example of segmentation result by the U-Net. The B-scan is of size 256×1024 pixels. The instrument is labeled in blue, while the retina is labeled in red.

TABLE I: Comparison of inference time shown as mean time \pm standard deviation, over 100 iterations.

	Mean Inference	Maximum inference
Original U-Net	$187.16 \pm 3.48\text{ ms}$	194.24 ms
Optimized TensorRT	$5.39 \pm 0.02\text{ ms}$	5.47 ms

forementioned U-Net inference optimization significantly reduced the inference time, while it was proven to preserve overall estimation accuracy as well [5]. Fig.4 provides a view of the obtained segmentation accuracy.

CONCLUSIONS AND DISCUSSION

This abstract presented a preliminary study of real time retina and instrument segmentation based on optimized DL inference model built with TensorRT. It was shown that it is possible to segment the OCT image of the retina and the instrument in several milliseconds. This result is supportive for real time instrument tracking in robot-assisted subretinal surgery. A limitation of the current work is that training of the DL networks requires a sufficiently large data-set covering various situations, while in this abstract the proposed method is only tested on different positions of a simple eye model.

REFERENCES

- [1] Y. Peng *et al.*, "Subretinal injection: a review on the novel route of therapeutic delivery for vitreoretinal diseases," *Ophthalmic research*, vol. 58, no. 4, pp. 217–226, 2017.
- [2] C.J. Johnson *et al.*, "Technical brief: Subretinal injection and electroporation into adult mouse eyes," *Molecular Vision*, vol. 14, p. 2211, 2008.
- [3] M. Sommersperger *et al.*, "Real-time tool to layer distance estimation for robotic subretinal injection using intraoperative 4d oct," *Biomed Opt Express*, vol. 12, no. 2, pp. 1085–1104, 2021.
- [4] M. Ourak *et al.*, "Combined oct distance and fbg force sensing cannulation needle for retinal vein cannulation: in vivo animal validation," *IJCARS*, vol. 14, no. 2, pp. 301–309, 2019.
- [5] S. Borkovkina *et al.*, "Real-time retinal layer segmentation of oct volumes with gpu accelerated inferencing using a compressed, low-latency neural network," *Biomed. Opt. Express*, vol. 11, no. 7, pp. 3968–3984, Jul 2020.
- [6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [7] A. Houman *et al.*, "Speeding Up Deep Learning Inference Using NVIDIA TensorRT (Updated)," <https://developer.nvidia.com/blog/speeding-up-deep-learning-inference-using-tensorrt-updated/>, 2021, [Online; accessed 20-June-2021].

Controlling of a Micropositioning Piezoelectric Actuator using an LSTM Network for Robotic Eye Surgery

Yao Zhang¹, Mouloud Ourak¹, Eric Verschooten², Kenan Niu¹, Di Wu¹, Gianni Borghesan^{1,3}, Philip X. Joris², Emmanuel Vander Poorten¹

¹KU Leuven, Department of Mechanical Engineering, Leuven, Belgium

²KU Leuven, Laboratory of Auditory Neurophysiology, Leuven, Belgium

³Core Lab ROB, Flanders Make, Belgium

INTRODUCTION

Retinal venous occlusion (RVO) is the second most common retinal vascular disease. In total, more than 16.4 million people suffer from it worldwide [1]. RVO can lead to severe vision-impairing damage due to neovascularisation, ischemia, and edema. Currently, retinal vein cannulation (RVC) provides a promising solution. However, it remains a challenging operation for surgeons. During the procedure, a clot-dissolving drug that can cannulate the clotted retinal vein is injected into an obstructed vein through a micro-scale cannulation needle [2]. However, due to the small scale of the retinal veins (30 to 400 μ m), reliable manual injection is extremely challenging [3].

With advancements in robotic eye surgery [4], the feasibility of robot-assisted cannulation has been demonstrated recently [5]. However, precise control of the insertion depth remains a challenge. Too deep insertion means that the targeted vein could be pierced and the active agent could be injected below the vein into a highly sensitive region. In order to avoid this problem, this abstract investigates the use of piezoelectric actuation to produce precise insertion. However, piezoelectric actuator exhibits hysteresis between the applied input voltage and output displacement. To cope with the complex non-linear relationship generated by hysteresis, mathematical modeling approaches, such as the Rate-Dependent Prandtl-Ishlinskii (RDPI), e.g. used in [6], were proposed. These methods require careful identification of quite a number of parameters. Recently, deep learning (DL) has also shown good potential to learn complex hysteresis behaviour [7]. This work investigates accurate control of the micropositioning piezoelectric actuator for RVC using such DL-method.

MATERIALS AND METHODS

Hysteresis is typically described as time series data, because it is affected by both current and previous inputs. To cope with this type of data, a Long Short-Term Memory (LSTM) artificial neural network, which uses historical data as a starting point and applies this knowledge to predict future outputs, is proposed [8].

To collect training data and evaluate the performance of the LSTM, an experimental setup was developed (see Fig.

This work was supported by a grant from the Belgian FWO [SB/1S96922N], and an internal KU Leuven C3 project 3E200471. Corresponding author: yao.zhang@kuleuven.be

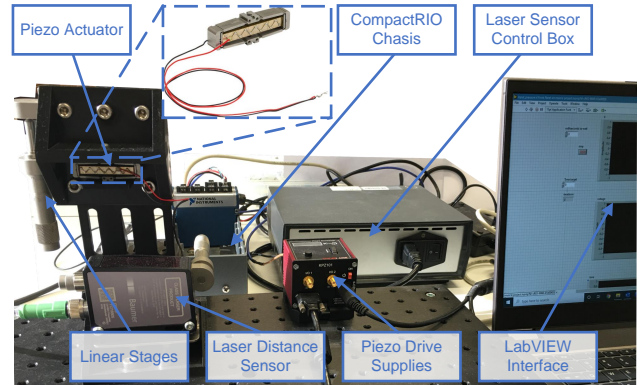


Fig. 1: Experimental setup, a piezo actuator realizes 1-dimensional motion, a laser distance sensor captures the achieved motion for evaluation purposes.

1). A micropositioning piezoelectric actuator (APF503, Thorlabs) is used to generate motion with amplitude up to 390 μ m. The piezoelectric actuator is actuated by piezoelectric drive supplies (KPZ101, Thorlabs) up to 150V. The piezoceramic is a smart material that expands or contracts when an electrical voltage is applied. The expansion or contraction of the piezoceramic is amplified into a larger linear movement through a flexure mounted on the actuator (visible in the insert of Fig. 1). The setup also contains a high-resolution (0.7 μ m) laser distance sensor (OM70-11216505, Baumer Group, Switzerland) that captures the motion of the piezoelectric actuator. The applied electrical voltage and resulting displacement is collected through LabVIEW[®]. Electrical voltages as descending sinusoidal waves following:

$$v(t) = Ae^{-\tau t} \left(\sin(2\pi ft - \frac{\pi}{2}) + 1 \right) \quad [V] \quad (1)$$

were used to drive the piezoelectric actuator and generate multi-loop hysteresis training data. The resulting displacement and collected electrical voltage was used as input and label to train the LSTM, respectively. The LSTM network contains 4 stacked layers. Each LSTM cell is made up of 64 neurons. Because of the small scale of the retinal veins, the amplitude A was set to 7.5V to ensure the peak value of the resulting motion would stay below 30 μ m. To enrich the training data, the descending constant τ was set to -0.08 and -0.12, the frequency f was set to 0.4Hz, 0.8Hz, 1.2Hz, and 1.6Hz. As a result, eight

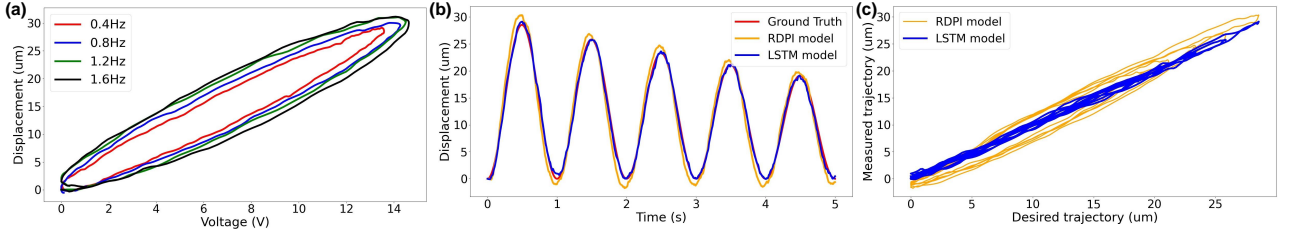


Fig. 2: (a): Rate-dependant hysteresis. (b): Desired trajectory (red) and measured trajectory achieved by RDPI (orange) and LSTM (blue). (c): Relation between desired trajectory and measured trajectory.

groups containing 19216 data points were used to train the LSTM. Figure 2(a) shows the major hysteresis loops of each frequency. The width of the hysteresis loops increases as the excitation frequency increases. This behavior is known as rate-dependent hysteresis. The whole training process takes around 25 to 30 minutes with 200 epochs on a 6 GB CUDA-capable GPU. To test the performance of the trained LSTM, the desired trajectory data were used as input, while needed control voltage was predicted as output. The following trajectory:

$$d(t) = 15e^{-0.1t} \left(\sin(2\pi t - \frac{\pi}{2}) + 1 \right) \quad [] \quad (2)$$

was used to test the trained LSTM model. The amplitude was set to 15 keeping the peak value of the test trajectory below 30. The test trajectory data were reshaped into a window size of 50. Each group acts like $d(t-49)$, $d(t-48)$, ..., $d(t)$, and was used to predict $v(t)$. Afterwards, the reshaped trajectory data were fed into the trained LSTM model. The average prediction time of each Δ_{point} in LSTM is 4.5 ms. The output of the LSTM was read as control signal and sent to the piezo drive to generate the corresponding voltage. The measurements from the laser distance sensor served as ground truth to calculate the three types errors in Table I. To assess the quality of the LSTM-based controller, a state-of-the-art RDPI-based controller, introduced in [6], was set up as well. To quantitatively evaluate the performance of both controllers, three metrics are used, namely Maximum Absolute Error (MAE), Root Mean Square Error (RMSE), and Normalized Root Mean Square Error (NRMSE).

TABLE I: Results, over 5 groups of experiments

Model	RMSE ()		MAE ()		NRMSE (%)	
	Mean	STD	Mean	STD	Mean	STD
RDPI	1.76	0.08	3.66	0.25	6.17	0.01
LSTM	0.56	0.08	1.48	0.23	1.95	0.01

RESULTS

The experiments were repeated five times. Figure 2(b) shows one example of the five experiments. One can observe that the LSTM-based model can accurately control the piezoelectric actuator, allowing it to follow precisely the desired trajectory. The three metrics and standard deviation over five groups of experiments are shown in Table I. The average RMSE, MAE, and NRMSE of the LSTM are respectively 0.56, 1.48, and 1.95%. Compared to the RDPI model, the LSTM model improves performance by 68%, 60%, and 69%, respectively, offering a lower

standard deviation over the three metrics, the LSTM-based controller shows good repeatable performance. The compensated input-output relationship is shown in the Fig. 2(c) by comparing the measured and desired trajectories. Compared to the RDPI model, the LSTM-based controller establishes a more linear relationship. This 1-to-1 response shows that the hysteresis is adequately compensated by the controller.

CONCLUSIONS AND DISCUSSION

An LSTM-based controller was introduced to precisely control the piezoelectric actuator in this abstract. The proposed LSTM model was first trained under eight groups of descending sinusoidal waves. Following that, the trained LSTM model was tested with the desired trajectory under different frequencies f and descending constant τ . The performance of the LSTM model was evaluated with RMSE (0.56), MAE (1.48), and NRMSE (1.95%). The errors are less than half of those achieved by the RDPI model. The good linearity shown in Fig. 2(c) demonstrates the feasibility of the proposed LSTM model in compensating for hysteresis under micrometer scale in a micropositioning piezoelectric actuator, which offers an alternative control scheme for minimally invasive eye surgery.

Future work will focus on investigating generalization ability of the LSTM-based controller over different test trajectories. We also aim to move to a realistic pre-clinical phantom.

REFERENCES

- [1] S. Rogers *et al.*, “The prevalence of retinal vein occlusion: pooled data from population studies from the United States, Europe, Asia, and Australia”, *Ophthalmology*, Vol. 117, pp. 313-319, 2010.
- [2] WM. Tang, *et al.*, “A study of surgical approaches to retinal vascular occlusions”, *Archives of Ophthalmology*, Vol. 118, pp. 138-143, 2000.
- [3] M. Ourak, *et al.*, “Combined oct distance and fbg force sensing cannulation needle for retinal vein cannulation: in vivo animal validation”, *International journal of computer assisted radiology and surgery*, Vol. 14, pp. 301-309, 2019.
- [4] E. Vander Poorten, *et al.*, “Robotic retinal surgery”, *In Handbook of Robotic and Image-Guided Surgery*, Elsevier, pp.627-672, 2020.
- [5] A. Gijbels, *et al.*, “In-human robot-assisted retinal vein cannulation, a world first”, *Annu. of Biomed. Eng.*, Vol. 46, pp. 1676-1685, 2018.
- [6] S. Shakiba, *et al.*, “Tracking control of an SMA-driven actuator with rate-dependent behavior using an inverse model of hysteresis”, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, Vol. 42, No. 8, pp. 1-15, 2020.
- [7] D. Wu, *et al.*, “Hysteresis Modeling of Robotic Catheters Based on Long Short-Term Memory Network for Improved Environment Reconstruction”, *IEEE Robotics and Automation Letters*, Vol. 6, pp. 2106-2113, 2021.
- [8] S. Hochreiter, *et al.*, “Long short-term memory”, *Neural computation*, Vol. 9, pp. 1735-1780, 1997.

Experimental assessment of a fabric-based wearable haptic softness display for the localization of uterine leiomyomas in robot-assisted surgical palpation tasks

S. Fani^{1,2}, A. Giannini³, D. Doria¹, A. Bicchi^{1,2,4}, T. Simoncini³, and M. Bianchi^{1,2}

¹Research Center E. Piaggio, University of Pisa, Pisa, Italy

²Information Engineering Department, University of Pisa, Pisa, Italy

³Division of Obstetrics and Gynecology, Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

⁴Soft Robotics for Human Cooperation and Rehabilitation, Fondazione Istituto Italiano di Tecnologia, Genova, Italy

INTRODUCTION

Uterine leiomyomas or fibroids are benign tumors that generate on the surface or within the muscular tissues of the uterus. They appear as lumps that are usually stiffer than the surrounding structures [1] and are associated with symptoms in about 80% of women affected by fibromatosis, which include bleeding, menstrual and pelvic pain and infertility. For these reasons, they often require surgical treatment, which usually relies on the usage of commercial robotic platforms for Robot-assisted Minimally Invasive Surgery (RMIS), such as the da Vinci Xi surgical system (Intuitive Surgical, CA, USA). The goal is the removal of the uterine fibroid or, in a precautionary manner, of the whole uterus [2]. This condition is extremely negative for fertile women. On the contrary, if the surgeon could correctly identify the position of the myomas through tactile feedback, which should be ideally close to the cutaneous information acquired during physical palpation, the impact of the surgical procedure would be minimized, together with the possible damages to the other uterine tissues. In literature there are several examples of haptic devices, both kinaesthetic (force-feedback) and cutaneous, which provides cues for lump localization in both laparoscopic and robotics-enabled surgery - see [3] for a discussion. However, these systems were not specifically designed for, and tested in, gynecologic RMIS applications, where the absence of haptic feedback in RMIS procedures represents a well-documented cause of incomplete removal of myomas and of higher incidence of recurrent myomas [4]. In [3] we proposed to use the W-FYD [5], a wearable fabric-based softness display for softness rendering - see fig. 1, to reproduce the stiffness characteristic of fibroids - and hence to enhance their localization in a teleoperation architecture that simulates a robot-assisted surgical palpation of leiomyomas. Relying on *sensory substitution* the teleoperation loop is intrinsically stable. In our work, we first performed an *ex-vivo* stiffness characterization of the uterine tissues, to properly map it onto the characteristics of our device. We tested our integrated system with gynecologic surgeons in palpation tasks with silicone specimens, which replicated the characteristics of uterine tissues with embedded leiomyomas. Results show

that our system enables a softness-based discrimination of the embedded fibroids comparable to the one that physicians would achieve using directly their fingers in palpation tasks. Furthermore, the feedback provided by the haptic interface was perceived as comfortable, intuitive, and highly useful for fibroid localization. In the following we report the details of the teleoperation platform and of the experimental protocol. For further information, please refer to [3].

MATERIALS AND METHODS

The teleoperation system we designed consists of two main components; the teleoperated part, or *Indenting System*; and the W-FYD integrated with the console, or teleoperating part.

A. The Indenting System

The *Indenting System* is a two Degrees of Freedom (DoFs) mechatronic device, which can perform a controlled indentation on an object (in our work we considered fibromatosis uterine tissues and silicone specimens), while gathering force and displacement information.

B. W-FYD: Wearable Fabric Yielding Display and the console

The W-FYD uses a layer of isotropic elastic fabric as interaction surface for the user. The device is composed of two main parts: the *base* and the *frame*. The base, directly fixed on the user's finger, hosts the lifting system that allows the passive interaction. The frame, that is moved by the lifting mechanism, hosts the two DC motors used to move and stretch the fabric band allowing the stiffness control. The device can enable two different modes of interaction: a passive mode, in which the user is stimulated via the lifting mechanism pressing the contact interface against the finger-pad (the finger is not moving - see Fig. 1(b)); and an active one, in which the finger actively probes the interaction surface for softness. In this work, we considered the passive mode to deliver only tactile cues to the skin [6].

The change in the stretching level of the fabric is obtained controlling the angle of the two DC motors, producing different levels of stiffness to be provided on the user's finger-pad. In the passive mode, the stiffness is managed considering the indentation h_p is directly generated by the system and the stretching state of the fabric is controlled accordingly. In this manner, the W-FYD can convey softness information - computed from

This work has received funding from the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence), and in the framework of PRIN (Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale) 2017 with the project TIGHT: Tactile InteGration for Humans and arTificial systems (Grant number 818 2017SB48FP)

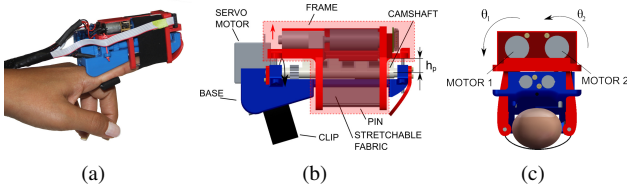


Fig. 1. W-FYD functioning principles [7]. (a) W-FYD Prototype worn on the user's finger. (b) Passive Mode: Lifting mechanism (in red the moving parts). (c) Motor angles definition. .

the force and indentation information acquired by the *Indenting System* - to the human operator (relying on a suitable scaling coefficient). The W-FYD is attached to the stylus of the Geomagic Touch, which acts as teleoperating console and provides position commands to the indenting system. The entire control cycle works at a frequency of 200 Hz.

C. The protocol

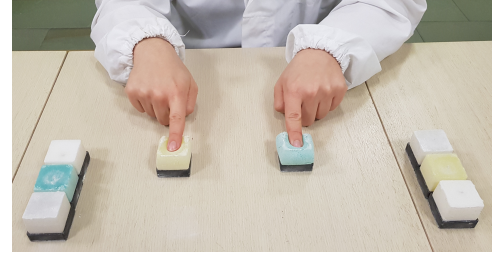
The task of the experiment consisted in the evaluation of the user capability in discriminating the presence of the myomas using stiffness tactile feedback in real-time in two different conditions. The first condition, named the *Ground Truth Condition*, consisted in the discrimination of the lump in silicone samples replicating the biological tissues without the use of the system, but relying on direct palpation. In the second condition, named *Evaluating Condition*, the surgeons were required to perform the same task using the teleoperation architecture with the W-FYD. For each trial, the test sample was placed below the index finger of the user's right hand/the *Indenting System* and the participant was asked to perform some exploration on it. At the same time, the user was allowed to perform an exploration with the left hand index finger on all the four comparison samples (a cubic shape of 35 mm containing a spherical shape with 6.5 mm diameter at three different depth lengths, 5 mm, 10 mm and 15 mm, respectively. There is an additional sample without any lump), one at the time and in the same order for each trial, and was asked to identify which of these corresponded to the test sample.

Thirteen participants (5 Female, Age mean \pm SD: 29.77 \pm 3.22), of which twelve medical residents and one senior experienced clinician, took part to the study. No one had any physical limitation which would have affected the experimental outcomes. They gave their informed consent to participate to the experiments. A total of 40 random trials were presented to the user (10 times for each sample) for each condition, see fig. 2.

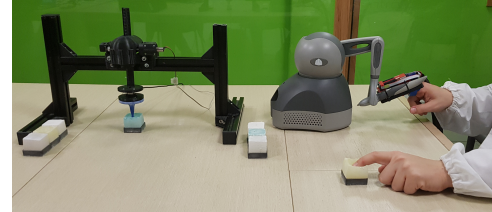
At the end of the experiment, the participants underwent through a subjective quantitative evaluation test using a seven-point Likert-type scale (1: Strongly disagree, 7: Strongly agree).

D. Results

A Friedman test was used to compare the results (recognition accuracy) in the two conditions considering four repeated measures for each subject (one for each experimental condition of the position of myomas). No statistical difference was found between subjects' performances while interacting directly with the samples and while interacting through the presented device ($p > 0.05$).



(a)



(b)

Fig. 2. Experimental setup (a) Ground Truth Condition (b) Evaluating Condition. In (b) it is also possible to observe the architecture with the W-FYD system worn by a surgeon and fixed to the teleoperating console, i.e. a Geomagic Touch stylus.

Results from the subjective evaluation showed that participants perceived the tactile feedback from the W-FYD intuitive and easy to distinguish with respect to the different stimuli. Furthermore participants agreed on the utility of the integration of the W-FYD with the telepresence robotic system in laparoscopic robotic myomectomy.

DISCUSSION AND CONCLUSION

The outcomes of our work [3] suggest that the integration of the W-FYD with a surgical robot could represent a viable solution for a precise localization of myomas during intraoperative procedures, enabling a less invasive and more located surgical intervention. However, an effective translation of this solution in real scenarios is not straightforward and comes with important challenges, such as the development of suitable sensing strategies for the teleoperated robot, for retrieving tissue stiffness information.

REFERENCES

- [1] E. A. Stewart, S. K. Laughlin-Tommaso, W. H. Catherino, S. Lalitkumar, D. Gupta, and B. Vollenhoven, "Uterine fibroids," *Nature reviews Disease primers*, vol. 2, no. 1, pp. 1–18, 2016.
- [2] E. E. Wallach, V. C. Buttram, and R. C. Reiter, "Uterine leiomyomata: etiology, symptomatology, and management," *Fertility and Sterility*, vol. 36, no. 4, pp. 433–445, oct 1981.
- [3] D. Doria, S. Fani, A. Giannini, T. Simoncini, and M. Bianchi, "Enhancing the localization of uterine leiomyomas through cutaneous softness rendering for robot-assisted surgical palpation application," *IEEE Transactions on Haptics*, 2021.
- [4] A. S. Moon, J. Garofalo, P. Koirala, M.-L. T. Vu, and L. Chuang, "Robotic surgery in gynecology," *Surgical Clinics*, vol. 100, no. 2, pp. 445–460, 2020.
- [5] S. Fani, S. Ciotti, E. Battaglia, A. Moscatelli, and M. Bianchi, "Wfyd: A wearable fabric-based display for haptic multi-cue delivery and tactile augmented reality," *IEEE transactions on haptics*, vol. 11, no. 2, pp. 304–316, 2017.
- [6] M. A. Srinivasan and R. H. LaMotte, "Tactual discrimination of softness," *J. of Neurophysiology*, vol. 73, no. 1, pp. 88–101, jan 1995.
- [7] S. Fani, S. Ciotti, E. Battaglia, A. Moscatelli, and M. Bianchi, "W-FYD: A wearable fabric-based display for haptic multi-cue delivery and tactile augmented reality," *IEEE ToH*, vol. 11, no. 2, pp. 304–316, apr 2018.

Towards a Handheld Robotic Instrument for Minimally Invasive Neurosurgery

Emmanouil Dimitrakakis¹, Holly Aylmore², Lukas Lindenroth¹, George Dwyer¹,
Hani J. Marcus^{1,3}, and Danail Stoyanov¹

¹Wellcome/EPSCRC Centre for Surgical and Interventional Sciences (WEISS), University College London (UCL), London, UK

²Queen Square Institute of Neurology, University College London (UCL), London, UK

³National Hospital for Neurology and Neurosurgery, London, UK

INTRODUCTION

Due to its delicate subject matter and challenging operations, neurosurgery has always been in need for adapting new techniques and technologies. A procedure that could widely benefit from robotic technology is the Endoscopic Endonasal Transsphenoidal Surgery (EETS). The EETS approach is a minimally invasive neurosurgical technique that is performed via an anterior sphenoidotomy and aims at the removal of sellar and parasellar lesions with the use of an endoscope and standard rigid instruments [1].

In recent years, there has been an increased interest in the Expanded Endoscopic Endonasal Approach (EEEA) that expands the EETS areas of interest [2]. Although a promising alternative to transcranial approaches that require craniotomies and brain retraction, the EEEA comes with its limitations. In [3], a number of surgeons were asked about these technical challenges, with 74% of them identifying the limited surgical manipulation that the standard non-articulated instruments offer as the biggest challenge of this procedure.

In this paper, current development of an ergonomically designed handheld robotic instrument for the EEEA is presented. The instrument employs a three degrees-of-freedom (DoF) robotic end-effector that provides access to targets on the surface of the brain that were previously unattainable with rigid tools. It also incorporates a rotating joystick-body mechanism that can be placed at the optimal position for the surgeon's postural ergonomics, aiming to increase the efficacy of keyhole neurosurgery without burdening the surgeon with physiological problems.

MATERIALS AND METHODS

A. Development of the robotic end-effector

The robotic end-effector aimed to be located at the distal tip of this instrument is a 3-DoF, tendon-driven, spherical joint manipulator, with a diameter of 4mm and 17.5mm length. A preliminary prototype of this miniature manipulator was fabricated and evaluated for its extended workspace and force capabilities in [4]. Based on the encouraging results from that preliminary design, the end-effector showcased in Fig. 1 was developed. This prototype was 3D printed both in resin and metal, and is currently undergoing experimental evaluation.

B. Development of the ergonomic handle

1) *Handle designs*: Long-term use of tools that have not been ergonomically designed can cause conditions

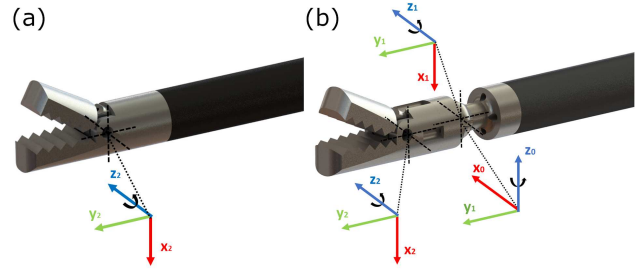


Fig. 1: (a) DoF of a conventional tool, and (b). DoF of the miniature end-effector.

such as carpal-tunnel syndrome [5]. Thus, appropriate ergonomic tool design is essential. To avoid such complications, the handle incorporated in the robotic instrument should be ergonomically designed, and ideally have a fast adoption rate, so that it can be easily integrated in the surgical workflow. To cater to a large set of ergonomic literature guidelines, two different handle concept prototypes were developed shown in Fig. 2.

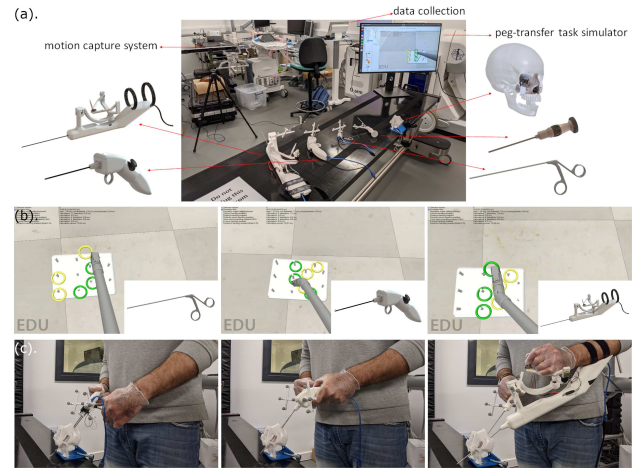


Fig. 2: (a) Experimental setup with renderings matched to their respective prototypes. (b). The simulated environment when a researcher is carrying out the peg-transfer task, and (c). The researcher holding the prototypes when carrying out the peg-transfer task.

The first handle is a forearm-mounted handle that maps the surgeon's wrist degrees-of-freedom to that of the robotic end-effector [6]. This handle alleviates the surgeon of any wrist-fatigue and its intuitive movement mapping

makes it easily adoptable. The second handle is a joystick-and-trigger handle with a rotating body that places the joystick to the position most comfortable for the surgeon, at a thumb abduction position. That position differs depending on hand-size and right- or left-handedness, making the use of a rotating body a convenient solution.

2) *Randomised crossover user-study*: The handles presented in the previous section were incorporated into a custom-designed virtual surgical task simulator and were assessed for their performance and ergonomics when compared with a standard neurosurgical grasper, to identify the superior handle design. The virtual task was performed by 9 novices with all 3 devices as part of a randomised crossover user-study. Their performance and ergonomics were evaluated both subjectively by themselves and objectively by a validated observational checklist [7].

In terms of performance and efficacy, the two robotic handles clearly outperformed the standard neurosurgical instrument. Between the two, the concept prototypes had very similar behaviours. As far as their learning-curves were concerned, the former was superior, whereas the latter appeared to be more time-efficient after training. Finally, during their ergonomic assessment, the rotating joystick-body handle proved to be the safest device to use for an extended amount of time according to a validated ergonomic measure.

C. Handheld robotic instrument adaptation

The superior ergonomic results of the rotating joystick-body handle, combined with the increased time and effort it would take to switch between forearm-mounted instruments during the endonasal approach, indicate that the more favourable handle amongst the two, is the former. Thus, current development focuses on the incorporation of electronics and the previously presented robotic end-effector into a handle with a rotating joystick-body, to form a fully functional robotic instrument prototype. A rendering of this prototype is shown in Fig. 3.

To incorporate electronics, as well as to create a robust coupling between the handle body and the robotic end-effector, the concept prototype presented in Fig. 2(b). needed redesign. A translational, rather than a rotational, joystick is used to make the device more compact, whereas the rotating joystick-body is now secured in place with a latch mechanism. The handle houses 3 miniature motors with their encoder adapters and motor controllers (FAULHABER, Schönaich, Germany).

It was fabricated with polylactic acid (PLA) plastic using the desktop 3D printer Ultimaker S5 (Ultimaker, Geldermalsen, Netherlands). Fig. 3 showcases the prototype handle body of the instrument as well as the end-effector fabricated in metal. With the development of the handle body of the instrument now finalised, the focus has shifted on the end-effector body.

CONCLUSIONS AND DISCUSSION

In this work, the current development of a novel handheld robotic instrument for minimally invasive neurosurgery was presented. The handle employs a miniature

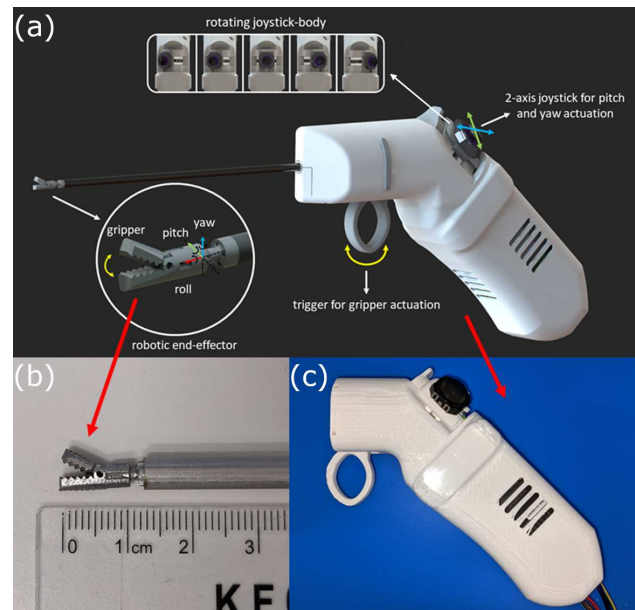


Fig. 3: Rendering of the robotic instrument, the metal end-effector, and a prototype of the actuating handle.

robotic end-effector and an ergonomically designed handle, aiming to increase the efficacy of keyhole neurosurgical approaches while also improving the procedural ergonomics associated with these complex procedures.

In future work, the development of the robotised prototype will be finalised, and its performance will be evaluated in laboratory and clinical experiments. After the efficacy of the instrument has been validated, the device will be re-evaluated for its human factors aiming to have developed a complete solution in terms of balance between performance and improved ergonomics.

REFERENCES

- [1] P. Cappabianca, L. M. Cavallo, and E. de Divitiis, "Endoscopic endonasal transsphenoidal surgery," *Neurosurgery*, vol. 55, no. 4, pp. 933–941, 2004.
- [2] A. R. Dehdashti, A. Ganna, I. Witterick, and F. Gentili, "Expanded endoscopic endonasal approach for anterior cranial base and suprasellar lesions: indications and limitations," *Neurosurgery*, vol. 64, no. 4, pp. 677–689, 2009.
- [3] H. J. Marcus, T. P. Cundy, A. Hughes-Hallett, G.-Z. Yang, A. Darzi, and D. Nandi, "Endoscopic and keyhole endoscope-assisted neurosurgical approaches: a qualitative survey on technical challenges and technological solutions," *British journal of neurosurgery*, vol. 28, no. 5, pp. 606–610, 2014.
- [4] E. Dimitrakakis, G. Dwyer, L. Lindenroth, P. Giataganas, N. L. Dorward, H. J. Marcus, and D. Stoyanov, "A spherical joint robotic end-effector for the expanded endoscopic endonasal approach," *Journal of Medical Robotics Research*, vol. 5, no. 03n04, p. 2150002, 2020.
- [5] P. Stoklasek, A. Mizera, M. Manas, and D. Manas, "Improvement of handle grip using reverse engineering, cae and rapid prototyping," in *MATEC Web of Conferences*, vol. 76. EDP Sciences, 2016, p. 02029.
- [6] E. Dimitrakakis, L. Lindenroth, G. Dwyer, H. Aylmore, N. L. Dorward, H. J. Marcus, and D. Stoyanov, "An intuitive surgical handle design for robotic neurosurgery," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2021.
- [7] E. Dimitrakakis, H. Aylmore, L. Lindenroth, G. Dwyer, J. Carmichael, D. Z. Khan, N. L. Dorward, H. J. Marcus, and D. Stoyanov, "Robotic handle prototypes for endoscopic endonasal skull base surgery: Pre-clinical randomised controlled trial of performance and ergonomics," *Annals of Biomedical Engineering*, pp. 1–15, 2022.

A Temporal Learning Approach to Inpainting Endoscopic Specularities and Its effect on Image Correspondence

Rema Daher, Francisco Vasconcelos, and Danail Stoyanov

Computer Science Department, Surgical Robot Vision Group, University College London
This work was supported by WEISS at UCL(203145Z/16/Z) and H2020 FET(GA863146)

INTRODUCTION

Computer vision has been utilized to analyze minimally invasive surgery videos and aid with polyp detection, tool localization, and organ 3D modelling tasks. However, irregular light patterns such as saturation, specular highlights, or extreme contrasts occlude texture and hinder these tasks. In this work, specular highlights were removed and the occluded data was reconstructed. To do that, an unsupervised temporal generative adversarial network (GAN) was used to inpaint specular highlights spatially and temporally. Due to the absence of a dataset with ground truth occluded textures, the network was trained on the in-vivo gastric endoscopy dataset (Hyper-Kvasir [1]) with specular highlight masks that were automatically created and processed to act as pseudo ground truths. Ablation studies and direct comparison with other methods were used to show the improved results of our system. In addition, the results on various datasets show the generalizability of our network on different environments and procedures as shown in Fig. 1. Finally, experiments also show the positive effect of inpainting on other computer vision tasks under the umbrella of 3D reconstruction and localization in endoscopy including feature matching, optical flow prediction, and disparity estimation.

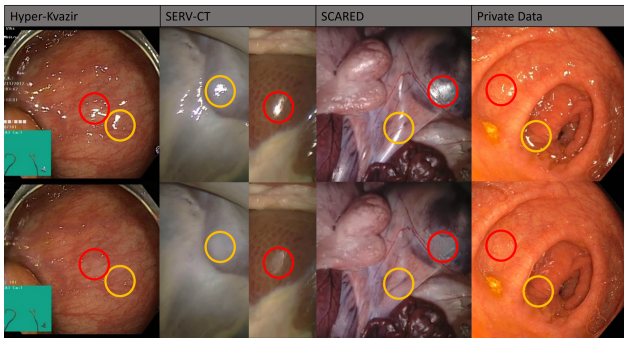


Fig. 1: Some motivational results of the specular highlight inpainting system on various datasets.

MATERIALS AND METHODS

The network architecture for inpainting is based on the Spatio-Temporal Transformer Network (STTN) proposed by [2]. STTN is a GAN mainly made up of spatial-temporal transformers that take as input feature vectors generated using a 2D convolutional encoder. The vector

outputs from the transformers are fed into a decoder to generate the inpainted images.

Some modification were performed to transform STTN from the task of object removal in diverse scenes to specularity removal in endoscopic videos.

First, their model is trained on inpainting diverse videos using random masks with frame ensured continuity. For the case at hand, a pseudo ground truth was created for training. To create this pseudo ground truth, specularities were first segmented from Hyper-Kvasir dataset using the chromatic characteristics of specular highlights [3]. After segmentation, the outputted masks are processed to cover specularity-free regions by using position translation and overlap clipping. Now that these masks cover up visible or unoccluded texture, this texture acts as pseudo ground truth to the inpainting output of these translated masks.

The pseudo ground truth was used for training the model; however, this model was first initialized by another model trained on temporally continuous random masks with endoscopic videos. The model was pretrained because having randomly located masks with various sizes can help the model learn missing textures of various regions of the gastrointestinal tract, whereas specularity masks can be limited to small sizes with certain textures for example. With these modifications, the proposed system becomes as is depicted in Fig. 2.

RESULTS AND DISCUSSION

Several experiments were performed to assess our system's inpainting ability. First, two models were analyzed, $Model_{S,C}$, which was trained from scratch using pseudo ground truth and $Model_{T,C}$, which was also trained on the pseudo ground truth, but it was initialized with another model trained on spatially consistent random shaped masks. Quantitative and qualitative analysis showed that $Model_{T,C}$ outperformed $Model_{S,C}$ in inpainting. Visually, with transfer learning, more details are captured as opposed to training from scratch. This was also evaluated quantitatively using Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) metrics ($Model_{S,C}$: PSNR=28.896, MSE=120.613. $Model_{T,C}$: PSNR=29.542, MSE=104.719)

Other learning based approaches that solve this problem do not have a temporal component [4], [5] and since these systems are not open source, an ablation study is used to show the importance of the temporal component in this application. To do that, the model was tested on one

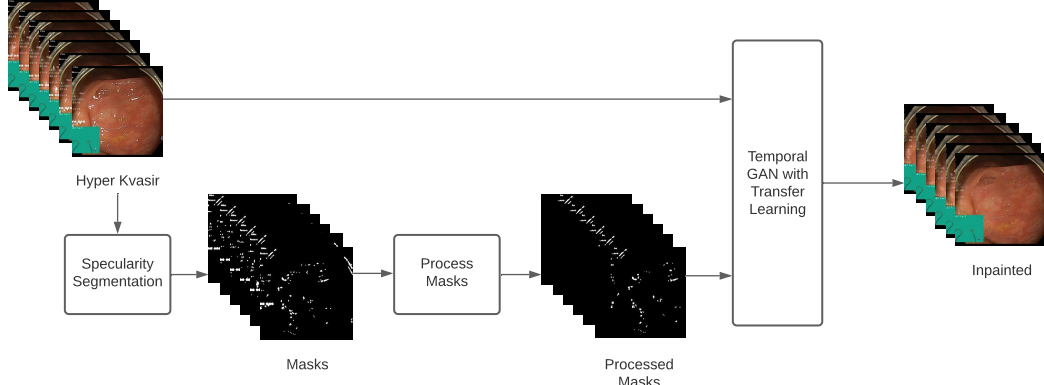


Fig. 2: The flowchart of the proposed system.

frame at a time, which made it act as a model with no temporal component, $Model_{T,C,NT}$. The model was also compared to a non-learning-based method represented as $Model_{Trad}$ and to the original STTN model $Model_{STTN}$. Visually, the proposed $Model_{T,C}$ showed the most detailed inpainting, which was also shown through quantitative analysis using PSNR and MSE metrics (Table I).

Models	$\blacktriangle PSNR_{mean}$	$\blacktriangledown MSE_{mean}$
$Model_{T,C}$	29.542	104.719
$Model_{STTN}$	28.683	119.541
$Model_{Trad}$	19.909	895.222
$Model_{T,C,NT}$	29.284	112.717

TABLE I: The $PSNR_{mean}$ and MSE_{mean} values for $Model_{STTN}$, $Model_{T,C}$, $Model_{T,C,NT}$, and $Model_{Trad}$.

$Model_{T,C}$ was tested on private data from a different hospital and on in-vivo datasets (SERV-CT [6], SCARED [7]). This showed the model’s generalizability (Fig. 1).

To analyse the effect of inpainting specular highlights in endoscopy on other computer vision tasks, disparity, feature matching, and optical flow are analysed. The results showed improvement in disparity with inpainting through visual and quantitative analysis on the basis of bad3% error, root mean square error (RMS), and the endpoint error (EPE) metrics.

As for feature matching, visually, the original frame pair matches count is higher than those generated from the inpainted frame pair. This indicates that without inpainting specular highlights, lower quality feature matches are generated. Visual analysis on optical flow estimation showed that the results were significantly better with inpainting giving smoother and more homogeneous optical flows.

For the quantitative analysis of both feature matching and optical flow, no direct ground truth was present. That is why, camera pose estimation was performed using the matches that were generated from feature matching and optical flow separately with the results shown in Table II. For feature matching, quantitative results were improved with inpainting. On the other hand, the optical flow results did not show any significant change; in fact, the results were even degraded slightly with inpainting.

Statistics	Feature Matching	Optical Flow
RTE ($^{\circ}$)	8.31 (% 11.72)	0.67 (% 1.73)
RRE ($^{\circ}$)	2.0 (% 30.12)	-0.05 (-% 1.59)
Inliers (pixels)	22.68 (% 25.17)	-2432 (-% 4.82)

TABLE II: Mean of the difference between the feature-/optical-flow-based pose estimation results of original versus inpainted sequences in terms of Relative Translation and Rotation Errors (RTE, RRE) and RANSAC Inliers.

CONCLUSIONS

In this work, an unsupervised inpainting system was proposed to remove specular highlights from endoscopic videos. A pseudo ground truth was created and used along with transfer learning to train a temporal GAN. Experiments showed improvement on previous methods by making use of direct comparison and ablation studies. Our inpainting system also showed generalizability on different datasets and a positive effect on endoscopic feature matching, disparity, and optical flow prediction.

REFERENCES

- [1] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen *et al.*, “Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy,” *Scientific Data*, vol. 7, no. 1, pp. 1–14, 2020.
- [2] Y. Zeng, J. Fu, and H. Chao, “Learning joint spatial-temporal transformations for video inpainting,” in *European Conference on Computer Vision*. Springer, 2020, pp. 528–543.
- [3] O. El Meslouhi, M. Kardouchi, H. Allali, T. Gadi, and Y. A. Benkaddour, “Automatic detection and inpainting of specular reflections for colposcopic images,” *Central European Journal of Computer Science*, vol. 1, no. 3, pp. 341–354, 2011.
- [4] S. Ali, F. Zhou, A. Bailey, B. Braden, J. E. East, X. Lu, and J. Rittscher, “A deep learning framework for quality assessment and restoration in video endoscopy,” *Medical Image Analysis*, vol. 68, p. 101900, 2021.
- [5] P.-R. Siavelis, N. Lamprinou, and E. Z. Psarakis, “An improved gan semantic image inpainting,” in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2020, pp. 443–454.
- [6] P. Edwards, D. Psychogios, S. Speidel, L. Maier-Hein, and D. Stoyanov, “Serv-ct: A disparity dataset from ct for validation of endoscopic 3d reconstruction,” *arXiv preprint arXiv:2012.11779*, 2020.
- [7] M. Allan, J. Mcleod, C. Wang, J. C. Rosenthal, Z. Hu, N. Gard, P. Eisert, K. X. Fu, T. Zeffiro, W. Xia *et al.*, “Stereo correspondence and reconstruction of endoscopic data challenge,” *arXiv preprint arXiv:2101.01133*, 2021.

Automatic Robotic Scanning for real-time 3D Ultrasound Reconstruction in Spine Surgery

Ruixuan Li, Ayoub Davoodi, Kenan Niu, Yuyu Cai, and Emmanuel Vander Poorten

Robot-Assisted Surgery Group, department of mechanical engineering, KU Leuven, Belgium

INTRODUCTION

Ultrasound (US) is a widely used non-radiative medical imaging modality considered for spine surgery. Through 3D reconstruction, US could help surgeons visualize the internal anatomy. Benefiting from intelligent robotic control, US-based robot-assistance could be exploited to perform autonomous scanning without leading to increased discomfort of the patient. Besides, robotic US scans could potentially offer better quality US images by keeping tight surface contact between the US probe and the skin.

During the past years, several robotic US systems have been developed and applied for spine surgery [1], [2]. Victorova *et al.* implemented a robotic US system with hybrid control for spine reconstruction. The proposed system regulated the applied force and enhanced the imaging quality [1]. However, it only focused on flat phantom with continuously surface. The post processing of US images was time-consuming. Zhang *et al.* also implemented an US scanning system with an adaptive approach to visualize spine [2]. The poses of US probe were manipulated while the contact force ranged from 14 to 16 N. Nevertheless, the US images were processed with complex image segmentation and volume rendering procedure. The aforementioned robotic US systems require post processing for 3D reconstruction and visualization. Therefore, it is necessary to develop a real-time reconstruction framework with robust image processing and control strategy for automatic robotic US scanning.

This paper provides a real-time US reconstruction with automatic robotic scanning. With hybrid control, the US probe keeps good surface contact during scanning. Meanwhile, the ultrasound images are simultaneously segmented and accurately reconstructed.

MATERIALS AND METHODS

The employed robotic US system consists of a lightweight robot (KUKA Robot LWR, Augsburg, Germany) and a 7.5 MHz US probe with US device (Sonosite, FUJIFILM, USA). During testing, the US images were recorded by a frame grabber (Epiphan, Palo Alto, USA). A 6 DOF F/T sensor (Nano25, ATI Industrial Automation, USA) was mounted with a custom designed probe housing at the robot end effector. A spine phantom (Model 034, CIRS, USA) was employed for experimental validation. In addition, a PC workstation (Intel i7, CPU @2.6 GHz, 64G RAM) with Nvidia P2000 GPU was used for data acquisition and processing. Figure 1 illustrates the workflow and the experimental setup.

The developed framework contains an automatic robotic scanning part and a real-time 3D US reconstruction. Before scanning, 15 predefined points were manually selected on the phantom surface with admittance control. The recorded points are used to generate an "S" shape scanning trajectory for automatic scanning. To perform the scanning, the z axis of the US probe was aligned to the normal vector of the surface which was computed from the predefined trajectory. Subsequently, automatic scanning was conducted with hybrid control while the force along probe z axis was kept constant. The target force was regulated on 3 N to guarantee the quality of US images and keep patient safe. To ensure real-time control, OROCOS (Open Robot Control Middleware) and eTaSL (expressiongraph-based Task Specification Language) were utilized.

The 3D US reconstruction was conducted in real-time during scanning. US spatial calibration was required to assure the reconstruction accuracy. Before scanning, the calibration was implemented with a custom designed sphere phantom [3]. Based on the calibration, the bone contours in the US images could be transformed to robot coordinates with the corresponding end-effector poses. A deep learning network, U-Net, was developed for realizing automatic image segmentation. For training, 500 images in 480×480 pixels were collected and manually labelled. Those images were augmented by mirroring and shifting on vertical and horizontal axes separately. Then, the model was trained with the 2000 augmented images. The trained model was evaluated with 50 images. A precision of 0.82 and accuracy of 0.98 was found. Subsequently, the model was integrated with the real-time reconstruction. The spine contours in the US images were automatically segmented and generated as point clouds into the robot coordinate. The outliers of the reconstructed point clouds were removed by computing the distance from their neighbors compared to the average with a 0.5 threshold level. Finally, the Visualization Toolkit (VTK) was used to visualize the generated point clouds in real-time by retrieving the data stream.

To evaluate the 3D reconstruction, the automatic scanning was repeated three times with the same scanning trajectory. Then, the three reconstructions were assessed by computing the distance between the three reconstructed point clouds (from source to target point clouds). This error indicates the scanning precision and actual spatial displacement between the reconstructed point clouds.

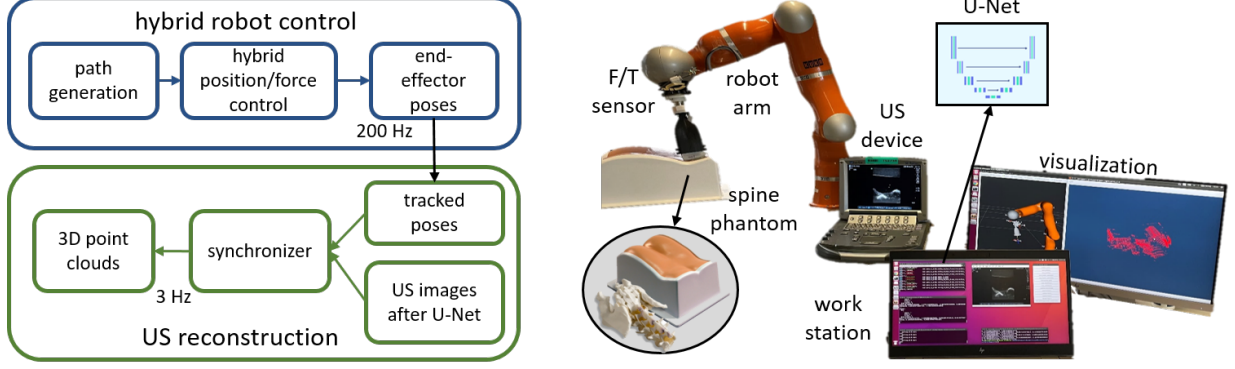


Fig. 1: (Left) Workflow of automatic robotic 3D US reconstruction. (Right) Overview of experimental setups.

RESULTS

The obtained force measurements are shown in Fig. 2 A). The force along the z axis is kept constant at 3 N during scanning. The mean value of measurements is 2.95 N while the standard deviation is 0.21. The mean force measurements along x and y axis are 0.05 N and 0.09 N respectively. The average scanning time is 273 seconds for the three repeated experiments.

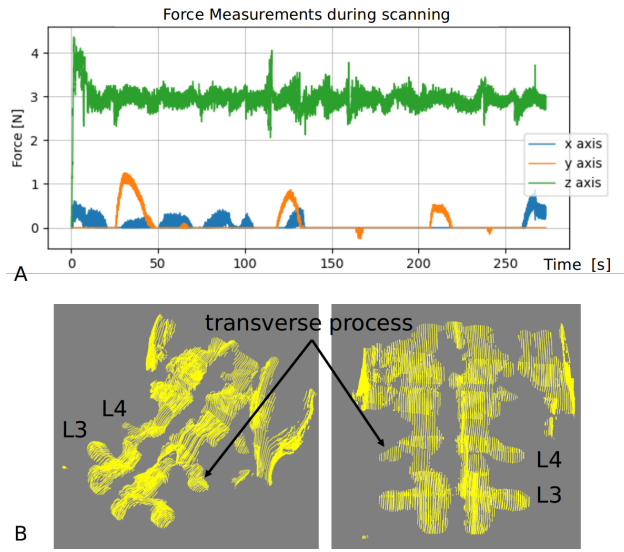


Fig. 2: A) An example of force measurements. B) An example of automatic reconstructed 3D point clouds.

The US images were segmented and processed in real-time during scanning. With the U-Net, the point clouds were generated and visualized at 3 Hz. The reconstructed 3D point clouds were displayed in Fig. 2 B). This result shows the geometric structures and features of the spine. The shape of the transverse process and the pedicle are visibly reconstructed. Table I shows the repeatability of the three 3D reconstructions. The mean errors of the reconstructed point clouds ranged from 0.53 to 2.14 mm.

CONCLUSIONS AND DISCUSSION

This paper implements a real-time robotic US reconstruction approach for spine surgery. With hybrid control,

TABLE I: Results of the repeatability assessment

source - target	RMSE [mm]	Mean [mm]	Std. Dev.
1-2	0.73	0.53	0.94
1-3	0.91	0.83	1.82
2-1	0.66	0.43	0.91
2-3	1.45	2.11	2.89
3-1	1.46	2.14	2.70
3-2	1.41	1.99	2.64

the proposed method constantly keeps the scanning force within a safe range to ensure good surface contact with patient's skin. The proposed scanning framework is not affected by the curvature of phantom while following the predefined scanning trajectory. For future research, it is interesting to focus on trajectory optimization to reduce the scanning time while keeping good reconstruction accuracy.

Besides, the real-time DL-based image segmentation improves the reconstruction efficiency and accuracy. The image reconstruction is implemented with scanning simultaneously. Thus, the post reconstruction in previous research [4], up to 8 seconds, is not required anymore. Furthermore, the 3D reconstructions are also accurately repeated with an RMSE lower than 1.46 mm. The reconstructed point clouds illustrate the geometric features of anatomy and spatial relation between each vertebra.

ACKNOWLEDGEMENT

This project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No.101016985 and Flemish Research Foundation (FWO) under grant agreement NO.G0A1420N and NO.1S36322N.

REFERENCES

- [1] M. Victorova, D. Navarro-Alarcon, and Y.-P. Zheng, "3d ultrasound imaging of scoliosis with force-sensitive robotic scanning," 2019, pp. 262–265.
- [2] J. Zhang *et al.*, "Self-adaptive ultrasound scanning system for imaging human spine," *IEEE Transactions on Industrial Electronics.*, vol. 69, no. 1, pp. 570–581, 2021.
- [3] R. Li, K. Niu, and E. Vander Poorten, "A framework for fast automatic robot ultrasound calibration," in *International Symposium on Medical Robotics*, 2021.
- [4] Q. Huang, B. Wu, J. Lan, and X. Li, "Fully automatic three-dimensional ultrasound imaging based on conventional b-scan," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 2, p. 426–436, 2018.

Detector-Free Dense Feature Matching for Fetoscopic Mosaicking*

Sophia Bano¹, Francisco Vasconcelos¹, Anna L. David², Jan Deprest³, and Danail Stoyanov¹

¹Wellcome/EPSCRC Centre for Interventional and Surgical Sciences & Department of Computer Science, University College London, UK

²Fetal Medicine Unit, University College London Hospital, UK

³Department of Development and Regeneration, University Hospital Leuven, Belgium

INTRODUCTION

Twin-to-twin transfusion syndrome (TTTS) is a rare fetal anomaly that affects the twins sharing a monochronic placenta. It is caused by abnormal placental vascular anastomoses on the placenta, leading to uneven flow of blood between the two fetuses [1]. Fetoscopic Laser Photocoagulation (FLP) is used to treat TTTS, however, this procedure is hindered because of difficulty in visualizing the intraoperative surgical environment due to limited surgical field-of-view (FoV), unusual placenta position, limited manoeuvrability of the fetoscope and poor visibility due to fluid turbidity and occlusions (Fig. 1. This adds to the surgeon’s cognitive load and may result in increased procedural time and missed treatment, leading to persistent TTTS. Fetoscopic video mosaicking can create a virtual expanded FoV image of the fetoscopic intraoperative environment, which may support the surgeons in localizing the vascular anastomoses during the FLP procedure.

Classical video mosaicking techniques perform hand-crafted feature detection and description (i.e. SIFT, SURF, ORB, etc) followed by feature matching and homography estimation for image stitching. However, these methods perform poorly on the in vivo fetoscopic videos due to low resolution, poor visibility, honeycomb or blur effect, floating particles and texture paucity or repetitive texture. Deep learning-based sequential mosaicking [2] method overcomes the limitation of feature-based mosaicking methods but results in drifting error when stitching non-planar views. A recent intensity-based image registration [3] method relies on placental vessel segmentation maps for registration. This facilitate in overcoming some visibility challenges, however, this method fails when the predicted segmentation map is inaccurate or inconsistent across frames or in views with thin or no vessels.

In the paper, we propose the use of transformer-based detector-free local feature matching (LoFTR) method [4] as a dense feature matching technique for creating reliable mosaics with minimal drifting error. Using the publicly available dataset [3], we experimentally show the robustness of the proposed method over the state-of-the-art vessel-based method.

*This work was supported by the Wellcome/EPSCRC Centre for Interventional & Surgical Sciences (WEISS) (203145Z/16/Z), EPSRC (EP/P027938/1, EP/R004080/1, NS/A000027/1, EP/P012841/1); and the Royal Academy of Engineering Chair in Emerging Technologies Scheme.

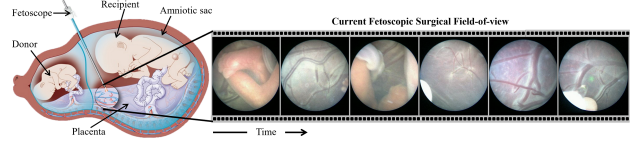


Fig. 1: During FLP, a fetoscope, having limited field-of-view, is inserted into the amniotic cavity and is used to localize and ablate the vascular anastomoses sites.

MATERIALS AND METHODS

Detector-Free Feature Representation: The recently proposed LoFTR [4] method first establish pixel-wise dense matches at a coarse level and later refine the good matches at a fine level. Given two consecutive frames F^t and F^{t+1} , a standard convolutional neural network architecture is used to extract dense features at coarse, $\tilde{F}^t, \tilde{F}^{t+1}$ (at $1/8^{th}$ of input resolution), and fine, \hat{F}^t, \hat{F}^{t+1} (at $1/2^{th}$ of input resolution), levels from both frames. The coarse local features, $\tilde{F}^t, \tilde{F}^{t+1}$, becomes the input to the LoFTR module. LoFTR uses transformer with positional encoding, and self and cross-attention layers to transform \tilde{F}_1, \tilde{F}_2 into position and context dependent local features, denoted as $\tilde{F}_{tr}^t, \tilde{F}_{tr}^{t+1}$, that can be matched easily.

Feature Matching: The coarse level matches \mathcal{M}_c between $\tilde{F}_{tr}^t, \tilde{F}_{tr}^{t+1}$ are established by using a differential matching layer, which gives a confidence matrix \mathcal{P}_j . The matches in \mathcal{P}_j with confidence higher than a predefined threshold and that also satisfies the mutual nearest neighbour criteria are selected as \mathcal{M}_c . Finally, coarse (\mathcal{M}_c) to fine (\mathcal{M}_f) matches are obtained by taking local window size from fine-level features, \hat{F}^t, \hat{F}^{t+1} , at each coarse match positions, applying a LoFTR module to obtained the fine transformed representation and correlating them. For more detail, please refer to [4], in which it is shown that LoFTR produces high-quality matches even in regions having low-textures and are affected by motion blur or repetitive patterns; making it an ideal matching module for fetoscopic mosaicking.

Registration and Mosaicking: A circular mask covering only the fetoscopic FoV is used to remove matches from the unwanted blank regions. The registration between F^t and F^{t+1} is approximated as an affine transformation [3] using RANdom SAMple Consensus (RANSAC) method. The obtained transformation is refined by using only

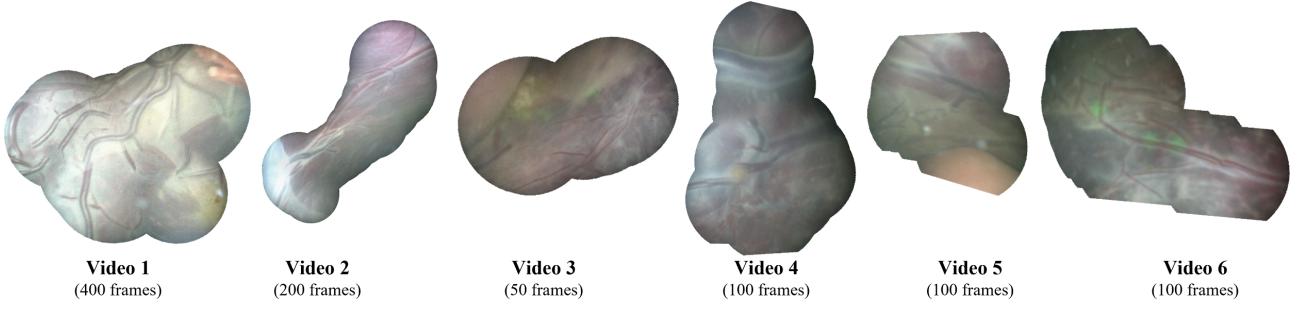


Fig. 2: Visualization of the generated mosaic using the proposed LoFTR-based method for the 6 in vivo clips.

the inliers with the Levenberg-Marquardt method that further reduces the reprojection error. Left-hand matrix multiplication is applied to the pairwise transformations to obtain the relative transformations of all frames in a video with respect to the first frame [2], following by image blending to generate an expanded FoV image.

RESULTS AND DISCUSSION

For the experimental analysis, we used the publicly available fetoscopy placenta dataset [3] that contains 6 in vivo fetoscopy video clips from 6 different TTTS procedures. The LoFTR matching model, pretrained on the ScanNet dataset [4], is used for obtaining the fine-level matches between two consecutive frames. Since groundtruth transformations are not available in in vivo fetoscopic videos, we use the 5-frame structural similarity index measure (SSIM) [3] for the quantitative evaluation of the resulting mosaics. The proposed LoFTR-based method is also compared with the state-of-the-art vessel-based [3] method (see Fig. 3). The qualitative analysis on all 6 video clips is also performed (see Fig. 2).

From Fig. 3, we can observe that the proposed LoFTR-based method performed significantly better on all video clips (except video 1) resulting in significantly low interquartile range and high median 5-frame SSIM when compared to the vessel-based (MICCAI2020) [3] method. Video 1 contains heavy amniotic fluid particles dynamically floating in the in vivo environment, which affects the performs of the LoFTR resulting in inaccurate transformation estimation. In a vessel-based method, such particles are already filtered during vessel segmentation. Video 2-6 have dynamically changing non-planar views, with some videos having low illumination and some frames having either no or very thin vessels. This negatively influences the vessel-based method, resulting in increase drifting error. LoFTR-based method, on the other hand, showed robustness even in regions having poor textures (very thin or no vessels) and low illumination. This is also evident from the qualitative results (see Fig. 2 which can be compared with the qualitative results presented in [3]. In the case of Video 1, discontinuities are visible in the generated mosaic. In the case of Video 2-6, the generated mosaics are reliable and accurate without any visible discontinuities, which is also inline with the observations drawn from the quantitative results.

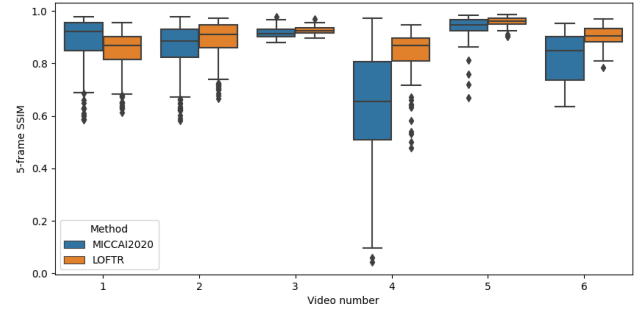


Fig. 3: Quantitative comparison of the proposed LoFTR-based method with the vessel-based (MICCAI2020) method [3] using the 5-frame SSIM metric.

CONCLUSIONS

We propose a fetoscopic video mosaicking method that benefited from the detector-free feature matching with transformers (LoFTR) [4] method, resulting in generating reliable virtual expanded field-of-view image of the intraoperative fetoscopic environment. Using the publicly available fetoscopy placenta dataset [3], we experimentally showed that the proposed LoFTR-based method outperformed the state-of-the-art vessel-based fetoscopic mosaicking method. The proposed method is robust even in low-textured and low illumination non-planar views, which shows the potential of facilitating the surgeons during the TTTS procedure. Future work involves validating the proposed method on the larger FetReg [5] dataset.

REFERENCES

- [1] A. Baschat, R. H. Chmait, J. Deprest, E. Gratacós, K. Hecher, E. Kontopoulos, R. Quintero, D. W. Skupski, D. V. Valsky, Y. Ville *et al.*, “Twin-to-twin transfusion syndrome (TTTS),” *Journal of Perinatal Medicine*, vol. 39, no. 2, pp. 107–112, 2011.
- [2] S. Bano, F. Vasconcelos, M. T. Amo, G. Dwyer, C. Gruijthuijsen, J. Deprest, S. Ourselin, E. Vander Poorten, T. Vercauteren, and D. Stoyanov, “Deep Sequential Mosaicking of Fetoscopic Videos,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019.
- [3] S. Bano, F. Vasconcelos, L. M. Shepherd, E. Vander Poorten, T. Vercauteren, S. Ourselin, A. L. David, J. Deprest, and D. Stoyanov, “Deep Placental Vessel Segmentation for Fetoscopic Mosaicking,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020.
- [4] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, “LoFTR: Detector-free Local Feature Matching with Transformers,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [5] S. Bano, A. Casella, F. Vasconcelos, S. Moccia, G. Attilakos *et al.*, “FetReg: Placental Vessel Segmentation and Registration in Fetoscopy Challenge Dataset,” *arXiv:2106.05923*, 2021.

INTRODUCTION

Awake surgery consists into neurosurgery while the patient is awake, i.e., under conscious sedation. Possible objectives of awake surgery are tumor removal, resection of brain vascular malformations, and deep brain stimulation (e.g., treatment of Parkinson disease and epilepsy), otherwise inoperable due to the proximity of the interested areas to those of the brain that control vision, language, or body movements. The main advantage of awake surgery over unconscious sedation is the precise localization of brain functions, in order to preserve as much as possible the patient's functions. During surgery, the neurosurgeon stimulates the area around the tumor with small electrodes (i.e., direct electrical stimulation) to locate precisely the functional areas of the brain that must be avoided for the surgical incision. The patient, who receives a sedation that lets him/her to be conscious, is able to perform specific tasks demanded by the neuropsychologist. These tasks are chosen in the pre-surgery phase, according to the brain localization of the target and taking into account the patient's pre-surgical capabilities (i.e., level of education, presence of mental impairment, etc.) to verify the presence of movement disorders (e.g., slowing or freezing movements during the test), aphasia (e.g., sudden language impairment), dysexecutive syndrome (e.g., increased delay in reasoning and answering), etc. In the surgical practice, the patient's assessment is visually performed by the neuropsychologist who evaluates whether the patient retains his/her capabilities or not, and eventually guides the surgeon through a minimally damaging cut to reach the target. Such evaluation is widely subjective and the neuropsychologist must catch any variation during the tests, ensuring, throughout the procedure, that the patient does not feel any pain.

To support neuropsychologist and to overcome limitations posed by perspective evaluations, in the related work [1] a tablet has been employed during the awake surgery, but no data has been stored and no criteria for decision support are investigated. In [2], a platform to deliver language and cognitive tasks to patients undergoing awake surgery has been tested in order to provide a portable system for streamlining cognitive tasks and record standardized patient response data. Serious games have been proposed to assess the capabilities of patients, such as [3], where a tablet has been employed for cognitive capabilities, or [4], where a popular game has been reworked and developed, using Leap Motion Controller (LMC), for stroke rehabilitation. In [5], the authors have employed a LMC to quantify bradykinesia in parkinsonian patients under deep brain stimulation. Previously, the authors of [6] have proposed an automatic classification tool, based on LMC, to assess the motor dysfunction in patients with Parkinson disease.

OBJECTIVE

Inspired by literature achievements in similar scenarios, the objective of this study is to propose an intraoperative integrated system to support, in their evaluation, the neuropsychologists during the interaction with the awake patient, including cognitive and motor tasks, with a specific focus on assessment of hand function and fine motor skills. The cognitive evaluation is supported by the use of a tablet, while the motor monitoring makes use of a LMC. These two tools are selected for their real time outcomes, which is crucial considering the needs of promptly notifying whether the surgeon is touching relevant brain structures.

MATERIALS AND METHODS

Two main devices are taken into account: a tablet and a LMC. The tablet serves as an interface to propose the tests, collects the answers automatically, performs an unbiased evaluation of the patient skills, and provides visual support to the neuropsychologist. We note that the tests are currently performed using both sheets and electronic devices. To achieve the objectives, a custom application (app) is developed which includes a set of tasks, and each one is conceived to evaluate one or more skills. Part of the tasks merely consist into digitalization of previous ones, while some other tasks are agreed upon the neuropsychologist to exploit the features of the device. In any case, which tests are to be adopted for each patient is a responsibility of the neuropsychologist. Starting from the patient's input, specific features are extracted for each task. A possible task is presented in Figure 1, where the patient is asked to select a specific tile, indicated in the topmost text line.



Figure 1: Task example.

For example, relevant features are latency of response to stimuli (e.g., delay in touching a specified object, indicated by a visual message) and precision of finger positioning (e.g., missing the right point due to impaired movements). The features are analyzed in real time and made available to the neuropsychologist through intuitive data visualization. In the testing phase, the neuropsychologist should signal whenever any anomaly occurs, in order to improve the automatic detection of anomalies at a deployment stage. We believe it is impractical to employ acoustic signals or messages, as the operating room is noisy and the patient's voice is attenuated by the oxygen mask. Pre-surgical skill assessment are considered as a baseline for tablet-mediated intra-surgical evaluation. Moreover, as to avoid possible bias during intra-surgical evaluation, these assessments should be performed mimicking the patient's position during surgery.

The LMC, instead, acquires the hand images, determines palm and finger positions, and tracks the motion of each joint in real time. The objective is to track the execution of motor tasks that involves the hand using a markerless technique, i.e., the patient does not need to wear any device or glove. Wearing a device, instead, could represent a limitation in the operating room, because several sensors are already in use (e.g., pulse oximeter) and the patient is encumbered by intravenous feeding. The LMC is a small device (80x30x11.3 mm) that employs infrared stereo cameras as tracking sensors. It must be placed next to the patient, because its interaction zone depth is 10 to 60 cm from the camera. The LMC must be connected to a personal computer to perform data acquisition, elaboration, and memorization. It can track the hand gesture at high speed (approximately 120 Hz) and it can discern 27 distinct hand elements, including joints and bones. Moreover, their motion can be approximately tracked even when they are temporarily concealed. Thanks to the software development kit, the position of each hand element can be exported and analyzed to generate motion features, such as smoothness of the motion, presence of sudden stops, and motion speed. Preliminary results (Figure2) related to monitoring thumb and index position show that freezing can be clearly detected.

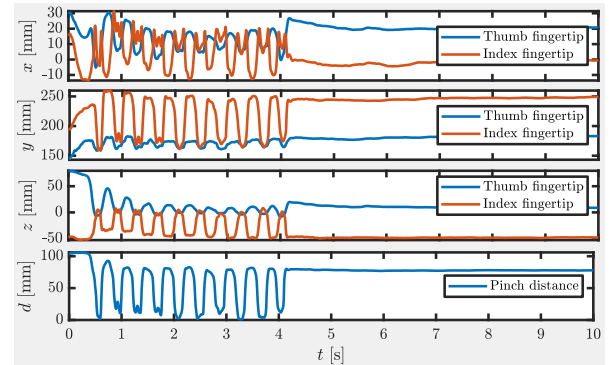


Figure 2: Position of the fingertips while performing a simple task (repetitive hand pinching). The task is stopped after 4 seconds to simulate freezing.

DISCUSSION AND CONCLUSIONS

To the best of authors' knowledge, it is the first proposal of a comprehensive tool to digitalize tests and evaluations given by the neuropsychologist during awake surgery and the first tentative in the literature to provide a decision support system for evaluating cognitive and motor skills of patients under awake surgery. Smartawake is a project proposal that includes pre-surgical skill evaluation to automatically compensate the patient's previous skills and the first data collection tool that integrates data from multiple sources (e.g., tablet and LMC) in the field of awake surgery. Such structured data may also be of scientific interest in order to find correlations between data acquired under the Smartawake project, annotations in the electronic medical record, and events annotated by the neuropsychologist during the experiments, which may be undiscovered yet.

References

- [1] M. A. Morrison, F. Tam, M. M. Garavaglia, L. Golestanirad, G. Hare, M. D. Cusimano, T. A. Schweizer, S. Das, and S. J. Graham, *A novel tablet computer platform for advanced language mapping during awake craniotomy procedures*. Journal of neurosurgery, vol. 124 4, pp. 938–44, 2016.
- [2] N. U. F. Hameed, Z. Zhao, J. Zhang, L. Bu, Y. Zhou, L. Jin, H. Bai, W. Li, J. Tang, J. Lu, J. Wu, and Y. Mao, *A Novel Intraoperative Brain Mapping Integrated Task-Presentation Platform*. Operative Neurosurgery, vol. 20, no. 5, pp. 477–483, 02 2021.
- [3] T. Tong, M. Chignell, M. C. Tierney, and J. Lee, *A serious game for clinical assessment of cognitive status: validation study*. JMIR serious games, vol. 4, no. 1, p. e5006, 2016.
- [4] M. Khademi, H. Mousavi Hondori, A. McKenzie, L. Dodakian, C. V. Lopes, and S. C. Cramer, *Free-hand interaction with leap motion controller for stroke rehabilitation*. CHI'14 Extended Abstracts on Human Factors in Computing Systems, 2014, pp. 1663–1668.
- [5] J. Wu, N. Yu, Y. Yu, H. Li, F. Wu, Y. Yang, J. Lin, J. Han, and S. Liang, *Disturbance-observer-based control and related methods—An overview*. Parkinson's Disease, vol. 2021, 2021.
- [6] A. H. Butt, E. Rovini, C. Dolciotti, G. De Petris, P. Bongioanni, M. Carboncini, and F. Cavallo, *Objective and automatic classification of parkinson disease with leap motion controller*. Biomedical engineering online, vol. 17, no. 1, pp. 1–21, 2018.

Recovering 3D information of human soft tissue using stereo endoscopic images

Ziyang Chen¹, Davide Alberti¹, Aldo Marzullo², Giancarlo Ferrigno¹, and Elena De Momi¹

¹Department of Electronics, Information and Bioengineering, Politecnico di Milano

²Department of Mathematics and Computer Science, University of Calabria

INTRODUCTION

Robot assisted minimally invasive surgery (RAMIS) has been widely applied in various clinical treatment, and da Vinci surgical system is the most typical representative because of its special advantages, like hand-eye coordination and 3D vision. However, the movement of the stereo endoscope inside human body is limited. Augmented reality (AR) is considered to be integrated into RAMIS, since it can provide more visualization. Recovering 3D information of surgical scene directly affects the performance of AR. Many researchers have implemented 3D reconstruction based on stereo images, such as the semi-global block matching (SGBM) [1]. It has high disparity search efficiency for common stereo images, while the effect in medical field remains to be explored. Hence, an enhanced semi-global block matching approach with preprocessing (P-SGBM) is developed to recover the 3D information of endoscopic images in this paper.

METHODOLOGY

Stereo vision based 3D reconstruction can be divided into several parts, including stereo camera calibration, image rectification, stereo correspondence and depth mapping, as shown in Fig.1. In particular, stereo correspondence and depth mapping are crucial for recovering the space information, and they are implemented by means of the following steps,

1) Image Preprocessing To enhance the model performance and robustness, image preprocessing is a necessary way before the disparity generation. Here, Gaussian filtering and image equalization are combined to perform the image preprocessing, since it improves the quality on medical images because of removing noise and enlarging global contrast.

2) Matching cost computation The left and right pixels in the scanning line of processed images are matched horizontally, and BT cost [2] is adopted to measure their dissimilarity based on the linear intensity interpolation,

$$\hat{C}(p_k, q_k, I_L, I_R) = \min_{q_k - \frac{1}{2} \leq q \leq q_k + \frac{1}{2}} |I_L(p_k) - \tilde{I}_R(q)| \quad (1)$$

Where p_k, q_k denote the pixels to be measured in stereo images. I_L is the intensity function of left scanline, while \tilde{I}_R is the interpolated function of right scanline. The matching cost with disparity value d_k can be defined as,

$$C(p_k, d_k) = \min \{ \hat{C}(p_k, q_k, I_L, I_R), \hat{C}(q_k, p_k, I_R, I_L) \} \quad (2)$$

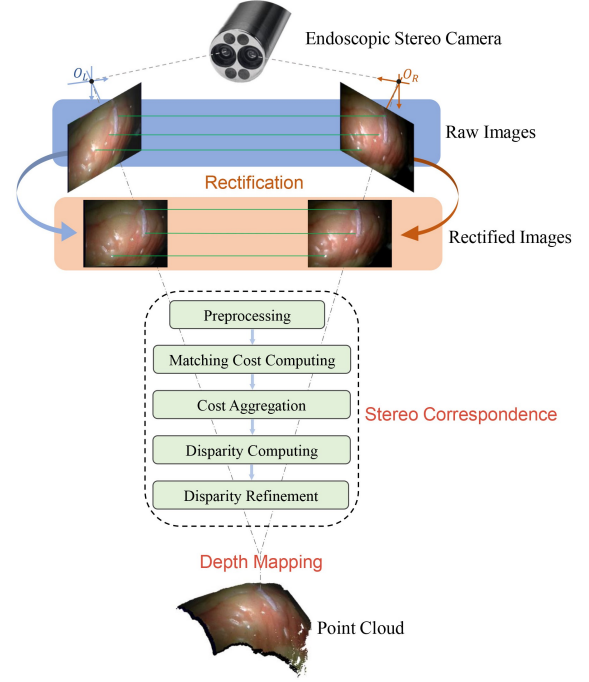


Fig. 1: The reconstruction process using stereo images.

3) Cost aggregation Considering the possibility of pixel mismatching, the smoothness of neighboring disparity is also regarded as a constraint. The cost S'_r is defined recursively as [1],

$$S'_r(p_k, d_k) = C(p_k, d_k) + \min(S'_r(p_k - r_l, d_k), S'_r(p_k - r_l, d_k \pm 1) + L_1, \min_{n \in [d_1, d_2]} S'_r(p_k - r_l, n) + L_2) - \min_{i \in [d_1, d_2]} S'_r(p_k - r_l, i) \quad (3)$$

Where $p_k - r_l$ denotes the left pixel near p_k , and L_1, L_2 are the different penalty factors. n represents the remaining values within the specific disparity range, while i is potential values with the same range. Then, the cost values in different directions are summed to calculate the final matching cost S_r . Empirically, we set L_1, L_2 as 144, 576 respectively, d_1 is -32 and d_2 is 96.

4) Disparity computation The final disparity map consists of the optimal disparity value corresponding to each pixel. Winner Takes All is adopted to extract the disparity value d_k by minimizing the aggregate cost S_r . Fig. 2 presents the searching process of optimal disparity.

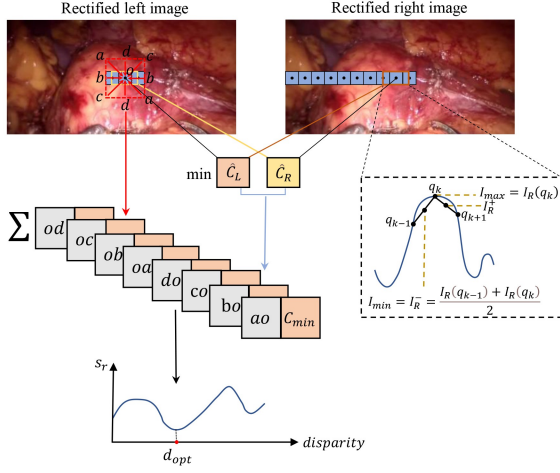


Fig. 2: Schematic diagram of disparity generation.

5) Disparity refinement By following [1], three optimization methods are utilized to improve the quality: (i) Uniqueness test is implemented to delete error values to avoid falling into local optimal solution; (ii) Subpixel interpolation aims to enhance the smoothness of disparity map; (iii) Left-right consistency check can eliminate disparity errors caused by possible object occlusion.

6) Depth Matching Following the camera projection transformation, the depth value z is denoted as,

$$Z = f * b / (p_k - q_k) \quad (4)$$

Where f is the camera focus, and b represents the baseline horizontal distance of the stereo camera. Lastly, global point cloud traversal is applied to remove potential singular points and black background points.

EXPERIMENTAL EVALUATION

To evaluate the practical performance of the proposed approach, we used the laparoscopic images from EndoAbs dataset [3]. Also, three different metrics [4] were adopted to evaluate the reconstruction performance,

1. *Accuracy*. The average Euclidean distance between the reconstructed point cloud and the ground truth.

2. *Computing Time*. Time from image preprocessing to point cloud generation.

3. *The percentage of matched pixels*. Ratio of reconstruction points of interest to the total pixels in the image.

Fig. 3 presents the preprocessed images of different organs and the final generated point clouds. Then, the quantitative evaluation is completed to verify the model performance, and the corresponding box plot is shown in Fig. 4. It can be seen that P-SGBM can have an apparent improvement in accuracy and save little computing time, although the percentage of matched pixels decreases slightly. In the meanwhile, the Wilcoxon rank-sum test ($p < 0.05$) is performed, and it shows that there are significant differences in Accuracy and Computing Time.

CONCLUSION

This paper proposes a 3D reconstruction approach named P-SGBM, and the experimental result shows that it

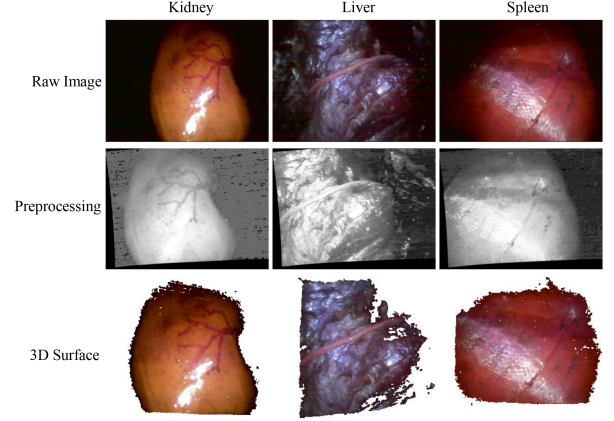


Fig. 3: Qualitative evaluation on Kidney, Liver and Spleen.

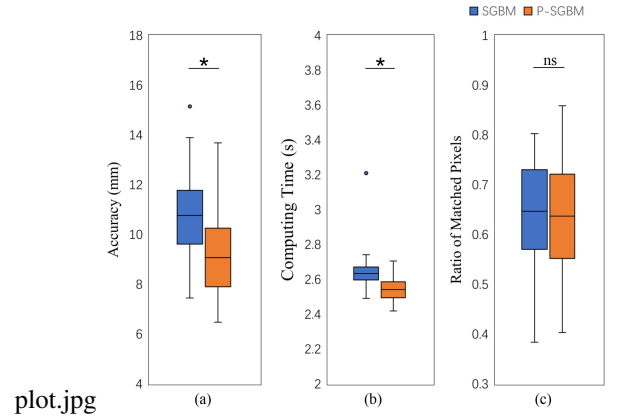


Fig. 4: Box plot representation of the quantitative performance with SGBM [1] and P-SGBM. (a)-(c) presents the result of accuracy, computing time and the percentage of matched pixels sequently.

can improve the performance of accuracy and computing time compared with the traditional one. In the future, deep learning based approaches [5] will be explored to develop a real-time implementation.

REFERENCES

- [1] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2007.
- [2] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 269–293, 1999.
- [3] V. Penza, A. S. Ciullo, S. Moccia, L. S. Mattos, and E. De Momi, "Endoabs dataset: endoscopic abdominal stereo image dataset for benchmarking 3d stereo reconstruction algorithms," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 5, p. e1926, 2018.
- [4] V. Penza, J. Ortiz, L. S. Mattos, A. Forgione, and E. De Momi, "Dense soft tissue 3d reconstruction refined with super-pixel segmentation for robotic abdominal surgery," *International journal of computer assisted radiology and surgery*, vol. 11, no. 2, pp. 197–206, 2016.
- [5] G. Yang, J. Manela, M. Happold, and D. Ramanan, "Hierarchical deep stereo matching on high-resolution images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5515–5524.

Autonomous Pick and Place using RL: Demonstrations Optimization

Claudia D’Ettorre¹, Agostino Stilli¹, Danail Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, UK
c.dettorre@ucl.ac.uk

INTRODUCTION

Automation of sub-tasks during robotic surgical operations still represents a challenging goal, especially for more elaborate tasks, such as pick and place, where different actions need to be executed sequentially in order to complete a task [1]. Among others, learning-based approaches are used to generate solutions that can solve tasks in different environments and conditions without the need of a human to tailor the algorithm to specific solutions once conditions change [2]. In the context of Robotic-Assisted Minimally Invasive Surgery (RAMIS), this work aims at contributing to the development of a flexible simulated environment to train the da Vinci surgical robot (Intuitive Surgical, Inc., Sunnyvale, CA) to autonomously execute pick and place tasks. To demonstrate the capabilities of the proposed simulation environment, we present a case study focused on the use of the PAF rail system for intraoperative organ repositioning [3]. The pick and place environment is developed inside the *rlman* framework [4], that interfaces the simulation environment in a gym-like fashion, with state of the art Reinforcement Learning (RL) libraries [5]. RL algorithms rely on an agent that interacts with the environment hundreds of thousands to millions of times, to learn the task. Becomes fundamental to have a good simulation environment to allow the agent to learn the specific task properly. We conduct four sets of experiments to show the importance of demonstration for correctly solving pick and place tasks.

MATERIALS AND METHODS

We consider a Markov Decision Process as a framework for picking optimal actions to maximize rewards in an environment E over discrete timesteps. An agent in a state x_t takes an action a_t , receiving a reward r for every timestep t , and E evolves to state x_{t+1} . In RL, the aim of the agent is to learn a policy $a_t = \pi(x_t)$ to maximize expected returns. $R_t = \sum_{i=t}^T \gamma^{(i-t)} r_i$ denotes the return, where T is the horizon that the agent optimizes over and γ is a discount factor for future rewards. The agent’s objective is to maximize expected return from the start distribution $J = \mathbb{E}_{r_i, s_i \sim E, a_i \sim \pi} [R_0]$.

For the experiments in this work we adopted Deep Deterministic Policy Gradients (DDPG) [7]. DDPG is a model-free off-policy reinforcement learning algorithm for continuous control which can utilise large function approximators such as neural networks. To handle varying task instances and parametrized goals, we use Hindsight Experience Replay (HER) [8]. Using HER, even fail episodes can be transformed by the agent into

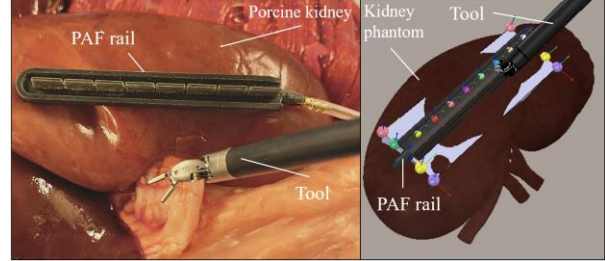


Figure 1 Case study surgical scene using the PAF rail system and the Large Needle Driver (LDN) tool reproduced using ex-vivo porcine kidney (left) and representation of the same scene in the *rlman* environment (right).

successful ones, provided that the state it experiences during the episode was the actual goal.

Besides the optimality of the aforementioned algorithms, it becomes very challenging for a RL agent to solve multi-step tasks, such as pick and place, without any previous knowledge of the task [9]. Usually for RL algorithms is challenging to discover the high reward space, especially when training an agent from scratch in sparse reward settings. To overcome these exploration challenges and ease the training, there is the need to integrate demonstrations into the learning process, collecting samples to be used as behavioral cloning [10].

EXPERIMENTS

Given the recent effort of the scientific community in developing different environments for training surgical subtasks [6], [11], it becomes crucial to understand the importance of demonstration needed in order to correctly learn a specific task, especially if there is the need for acquiring new demonstrations for every new task that the algorithm attempt to solve. Moreover, collecting demonstrations can be a challenging task based on the strategy adopted, while also being extremely time consuming. We decided to test four different demonstration datasets, each characterized by a different number of episodes, respectively: 10, 25, 50 and 100 episodes demonstrated.

We use the success rates during testing as evaluation metric for the results of the experiments [11]. The results for training agents in *rlman* are compared to *SurRoL* benchmarks. In both frameworks, the agents are trained using the same algorithms. We replicate the results on our machine only for the Gauze Retrieve (GR) task which is comparable to the pick and place task in *rlman*. Figure 2 shows a snapshot for each environment. We selected the GR environment as comparison based on the definition of the state space, action space and reward.

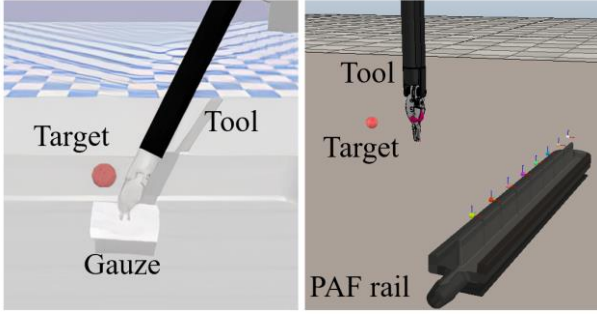


Figure 2 On the left: Gauze Retrieve environment from *SuRoL* and on the right the pick and place from *rlman*.

For both tasks, the goal is to sequentially pick the object, respectively the surgical gauze and the PAF rail, and drive it towards the target position.

In *rlman*, separate Docker containers are used to collect the agent experiences in parallel environments, maintaining the same shared replay buffer. Figure 3 shows the results of the run experiments. The first two columns represent the success rate over the epochs for the GR task, run in the *SuRoL* environment. The last two columns are the pick and place results from *rlman*. Each row is a different experiment run with a different number of demonstrations, as highlighted by the label on the left-hand side of the image.

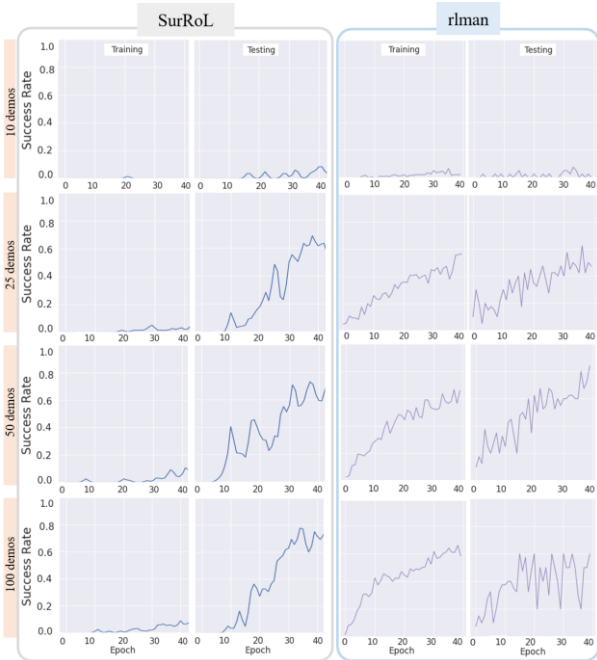


Figure 3 Results comparing the success rates during training and testing for the Gauze Retrieve task in *SuRoL* and the pick and place in *rlman*.

CONCLUSION AND DISCUSSION

We carried out four sets of experiments to understand the importance of number of demonstrations when training a pick and place task. The task has been analyzed and validated using the PAF rail system as a test case. The experiments compared the performances of the *rlman* framework and *SuRoL* benchmark. As shown in Fig.3, the agent is not able to learn in none of the environments

Proceedings of the 11th Joint Workshop on
New Technologies for Computer/Robot Assisted Surgery (CRAS 2022)

when using only 10 demonstrated episodes. Better performances are obtained when the number of episodes increases to 25 and 50. For 100 episodes demonstration, the performances start decreasing again, reflected by the lower success rates. In this case, the large number of episodes available as previous knowledge to the agent, is constraining the agent knowledge around the demonstrated trajectories, substantially reducing its exploration capability. This results in worst learning performances.

Although the tasks present the same formulation, the discrepancy encountered in the results might be due to differences in the definition of the environments and the associated variables. Results show that, although between 25 and 50 episodes are demonstrated, there are differences in the learning performances, these are not substantial. Further experiments, in the future work, will aim at defining the best trade-off between number of episodes recorded and agent performances in order to decrease the recording time. Moreover, it will be interesting to investigate how the quality of the demonstrations can affect the training. In this work, all the episodes recorded for demonstrations were successful. For future work we are going to analyse if failed demonstrations can impact the training and in what measure.

REFERENCES

- [1] C. D’Ettorre *et al.*, “Accelerating Surgical Robotics Research: A Review of 10 Years With the da Vinci Research Kit,” *IEEE Robot. Autom. Mag.*, 2021.
- [2] F. Richter, R. K. Orosco, and M. C. Yip, “Open-Sourced Reinforcement Learning Environments for Surgical Robotics,” 2019.
- [3] A. Stilli, E. Dimitrakakis, C. D’Ettorre, M. Tran, and D. Stoyanov, “Pneumatically Attachable Flexible Rails for Track-Guided Ultrasound Scanning in Robotic-Assisted Partial Nephrectomy—A Preliminary Design Study,” *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, 2019.
- [4] C. D’Ettorre, S. Zirino, N. N. Dei, A. Stilli, E. De Momi, and D. Stoyanov, “Learning Intraoperative Organ Manipulation with Context-based Reinforcement Learning,” in *The 13th International Conference on Information Processing in Computer-Assisted Interventions*, 2022.
- [5] P. Dhariwal *et al.*, “Openai Gym,” 2016.
- [6] M. C. Y. Florian Richter, Ryan K. Orosco, “Open-Sourced Reinforcement Learning Environments for Surgical Robotics.”
- [7] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning,” in *International Conference on Learning Representations (ICLR)*, 2016.
- [8] M. Andrychowicz *et al.*, “Hindsight Experience Replay,” *ArXiv*, 2018.
- [9] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, and S. Levine, “How to train your robot with deep reinforcement learning: lessons we have learned,” *Int. J. Rob. Res.*, vol. 40, 2021.
- [10] A. Nair, B. McGrew, M. Andrychowicz, W. Zaremba, and P. Abbeel, “Overcoming Exploration in Reinforcement Learning with Demonstrations,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [11] J. Xu, B. Li, B. Lu, Y.-H. Liu, Q. Dou, and P.-A. Heng, “SurRoL: An Open-source Reinforcement Learning Centered and dVRK Compatible Platform for Surgical Robot Learning,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

Deep Learning-based Plane Pose Regression: a First Step Towards Training and Guidance in Freehand Obstetric Ultrasound

Chiara Di Vece^{*1}, Brian Dromey^{1,2}, Francisco Vasconcelos¹, Anna L. David^{2,3},
Donald Peebles^{2,3}, and Danail Stoyanov¹

¹Wellcome/EPSCRC Centre for International and Surgical Sciences (WEISS), University College London, London, UK

³Elizabeth Garrett Anderson Institute for Women's Health, University College London, London, UK

⁴NIHR University College London Hospitals Biomedical Research Centre, University College London, London, UK

^{*}chiara.divece.20@ucl.ac.uk

INTRODUCTION

Obstetrics ultrasound (US) is a non-invasive, real-time and cost-effective diagnostic tool for monitoring mother and fetus throughout gestation. The learner's ability to mentally build a three-dimensional (3D) map of the fetus from a two-dimensional (2D) US image represents a major challenge to skill acquisition. Pre-defined anatomical planes, known as standard planes (SPs), retain a significant clinical relevance. Their correct acquisition, essential in the second-trimester fetal anatomic survey to investigate the morphological characteristics of the fetus, requires extensive experience due to the complexity of SP definition, their high intra- and low inter-class variation.

Previous work proposed automating the extraction of SPs from data acquired with a simplified protocol rather than assisting operators in acquiring typical freehand 2D SPs. These methods, based on traditional machine learning (ML) approaches [1] or convolutional neural network (CNN) [2], are mostly confined to selection or classification of SPs. Besides, the quality of the obtained SP cannot be compared with the one achieved with freehand scanning. Automatic approaches for the localisation of planes in 3D volumes are based either on supervised learning (SL) [3] or reinforcement learning (RL) [4]. The first ones, even though effective, can mainly localize one single plane at a time or are tailored to just one organ. Also, they attempt to learn a mapping from high-dimensional volumetric data to low-dimensional abstract features directly, making the training difficult. In the second case, most of the available solutions employing deep reinforcement learning (DRL) are based on robotic navigation and are used at the point of care. To the best of our knowledge, there are no available training systems to guide the novice sonographers in freehand obstetric US.

Currently, training and guidance in obstetric US are very focused on SPs recognition. The autonomous probe navigation towards SPs remains a highly challenging task due to the need to interpret variable and complex images and their spatial relationship. Our work aims to develop a real-time training platform to guide inexperienced sonographers in acquiring proper obstetric US images that could be potentially deployed for existing US machines.

MATERIALS AND METHODS

The development of our training system has been divided into five main blocks (Figure 1).

1. Unity simulator for volume reconstruction and synthetic images acquisition: The fetal volume can be sliced with arbitrarily oriented planes. The user can visualize them and annotate the standard ones. Besides, the environment can be used for the automated generation of supervised data using pre-acquired 3D US volumes.

2. Plane pose regression system: We use a regression CNN to estimate the six-dimensional (6D) pose of arbitrarily oriented US planes with respect to the center of the brain volume using a continuous rotation representation [5]. Our method is purely image-based and, therefore, does not require tracking sensors. Besides, it differs from traditional slice-to-volume registration methods since it does not require a previously acquired volume of the same subject being scanned. Instead, we predict poses relative to a generalized brain center, where training and test data belong to different subjects. The network was trained on synthetic images acquired from phantom US volumes and fine-tuned on real scans. Training data was generated by slicing US volumes into imaging planes in Unity at random coordinates and more densely around the transventricular (TV) SP manually annotated in Unity.

Network architecture. We used ResNet-18 as a backbone for feature extraction. We modified the network by re-initialising the fully connected layer and adding a regression head to directly output the rotation and translation representations. The inputs are the US image I (128×128) obtained by slicing the volume and its 6D pose $\theta_{GT} = (t_x, t_y, t_z, \alpha_x, \alpha_y, \alpha_z)$. The CNN predicts the 6D pose θ_{Pred} . Specifically, the network first outputs a vector of nine parameters $\theta_{Out} = (t_1, t_2, t_3, r_1, \dots, r_6)$; the first three are used for the translation and the last six for the rotation, then used internally by our CNN to reconstruct the rotation matrix \mathbf{R}' in the forward pass.

Loss function. For both translation and rotation we used as loss the mean squared error (MSE) between predicted $(\mathbf{t}', \mathbf{R}')$ and ground truth (\mathbf{t}, \mathbf{R}) values: $\mathcal{L}_{Tot} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{R}' - \mathbf{R}\|_2 + \lambda \frac{1}{N} \sum_{i=1}^N \|\mathbf{t}' - \mathbf{t}\|_2$, where N denotes the total number of images I within one training epoch, and λ the hyperparameter balancing the two losses.

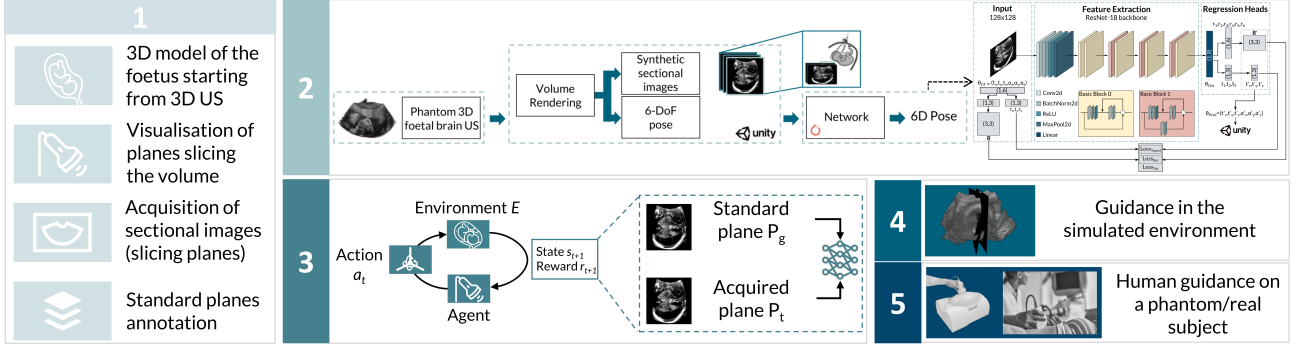


Fig. 1: Overall structure of the vision-based DRL training system for obstetric US acquisitions

Experiments. Our framework is implemented in *PyTorch* and trained using a single Tesla® V100-DGXS-32GB GPU of an NVIDIA® DGX station. The network was trained for 50 epochs with batch size of $K = 100$ using Adam optimiser, learning rate of 0.0001 and weighting factor $\lambda = 0.01$. We performed two experiments.

- **Experiment 1:** (1.1) Training (4 scans, 75088 images) and testing (2 scans, 37544 images) on phantom data; initialisation with weights from ImageNet; (1.2) Training and testing on real data; initialisation with weights from the case 1.1. The model is trained on two scans of one fetus (38754 images) and tested on two scans of a second fetus (38754 images). The test sets are divided into two subgroups: random planes (*Test A*) and planes around the TV SP (*Test B*). To evaluate the translation results, we employed the Euclidean distance between the two planes (*mm*). For rotation, we display errors as the geodesic distance to ground truth ($^{\circ}$) defined as $Error_{Rot} = \arccos((\mathbf{R}_{00}'' + \mathbf{R}_{11}'' + \mathbf{R}_{22}'' - 1)/2)$, where $\mathbf{R}'' = \mathbf{R}'^{-1}$.
- **Experiment 2:** We fed sectional images of the TV SPs into the network to estimate their pose as a sanity test. We plotted back the two planes within the volume into Unity to visually evaluate the distance between the annotated SPs and the predicted ones (2.1, 2.2).

3. DRL framework for planes alignment: The process reminds of the sonographers' behaviour as they continuously manipulate (*action*) the US probe (*agent*) to scan the maternal abdomen (*environment*) while visualising the intermediate planes P_t on the screen (*state*) until the SP P_g is acquired (*reward* and *terminal state*).

4. Guidance in simulation: Once the direction to align the plane P_t to the standard one P_g is found, the users could be guided through a 6-DoF haptic device by making them *feel* a force pointing towards the correct direction.

5. Guidance on the phantom/patient: The system could be developed as an interface and tested in a clinical setting, with validation studies with novice operators.

RESULTS

For phantom data, the median errors are 0.90 mm/1.17° and 0.44 mm/1.21° for random planes and planes close to the TV one. For real data, using a different fetus than training, these errors are 0.58 mm/0.74° and

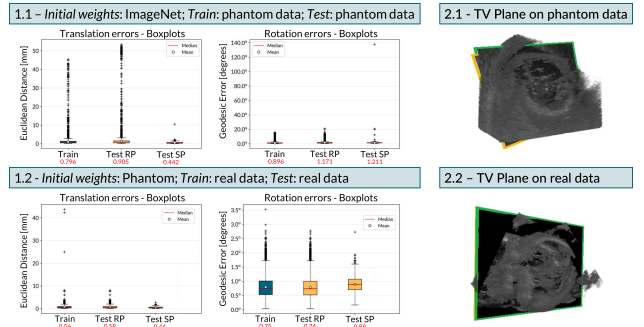


Fig. 2: Left: Translation and rotation error distributions in phantom (1.1) and real (1.2) US data for planes acquired at random coordinates (Test A) and around the annotated TV SP (Test B). Right: Visual evaluation of TV SP prediction on phantom (2.1) and real (2.2) US data

0.44 mm/0.89°. The average inference time is 2.97 ms per plane. Figures 2 reports the translation and rotation error distributions and the sanity test results.

CONCLUSIONS AND DISCUSSION

Our regression CNN can reliably localize US planes within the fetal brain in phantom data and successfully generalizes pose regression to an unseen fetal brain without the need for real ground truth data in real-time or 3D volume scans of the patient beforehand. Future development will expand the prediction to volumes of the whole fetus and assess its potential for vision-based, freehand US assisted navigation when acquiring fetal SPs.

REFERENCES

- [1] D. Ni *et al.*, “Standard Plane Localization in Ultrasound by Radial Component Model and Selective Search,” *Ultrasound in Medicine and Biology*, vol. 40, no. 11, pp. 2728–2742, 2014.
- [2] C. Baumgartner *et al.*, “SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, 2017.
- [3] A. Schmidt-Richberg *et al.*, “Offset regression networks for view plane estimation in 3D fetal ultrasound,” in *Medical Imaging*, 2019.
- [4] K. Li *et al.*, “Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning,” *IEEE International Conference on Robotics and Automation*, pp. 8302–8308, 2021.
- [5] Y. Zhou *et al.*, “On the continuity of rotation representations in neural networks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5738–5746.

OCT as tool for laser ablation monitoring applied to cholesteatoma

Q. Boyer^{*1}, D. V. A. Nguyen^{*1}, P. Rougeot¹, O. Gaiffe², B. Tamadazte³, J. Szewczyk³,
L. Tavernier², Y. Nguyen⁴, and K. Rabenoroso¹

¹FEMTO-ST, Université Bourgogne Franche-Comté, CNRS, Besançon, France

²Univ. Hospital of Besançon, Univ. Bourgogne Franche-Comté, France

³Institute for Intelligent Systems and Robotics (ISIR), Sorbonne University, CNRS, France.

⁴Sorbonne University/AP-HP, GHU Pitié-Salpêtrière, DMU ChIR, ORL Department, Unit of Audiology, France

I. INTRODUCTION

Cholesteatoma is an accumulation of exfoliated keratin from squamous epithelium that invades the middle ear, erodes the bony structures, and causes hearing loss as well as other serious complications. The only treatment of this disease is surgical intervention. As the middle ear cavity is small and contains hidden recesses, the complete removal of the disease is still challenging and the recurrence rate can reach 25% for conventional methods. An advanced treatment for this disease is laser surgery that has been proven to remove efficiently the residual cholesteatoma, thus reducing the recurrence rate [1]. In the μ RoCS project¹, we aim to propose a dexterous continuum robot with embedded laser instrument for exhaustive cholesteatoma removal. As part of this project, this study focuses on measuring the vaporized volume of cholesteatoma during the laser ablation process depending on the laser parameters. This volume will be obtained using OCT scanning and the proposed image processing. The result has an important role to vaporize the right amount of cholesteatoma tissue as insufficient removal would result in cholesteatoma recurrence and excessive removal would damage the healthy structures nearby.

II. MATERIALS AND METHODS

A. Experimental setup

An LBO fibered laser (Velas-8G, 532 nm, 8 W) was employed to vaporize cholesteatoma samples collected from Besançon Hospital² (see Fig. 1). The optical fiber has a core diameter of 400 μ m and a numerical aperture of 0.22. The studied parameters are the laser power (6 W and 8 W) and the exposure time (100 ms to 250 ms). The distance between the fiber tip and the tissue is important due to the divergence of the exiting laser beam. It should be as short as possible so the size of the laser spot is minimal. A small laser spot allows for precise ablation while limiting the unwanted heating of surrounding tissues. The fiber tip

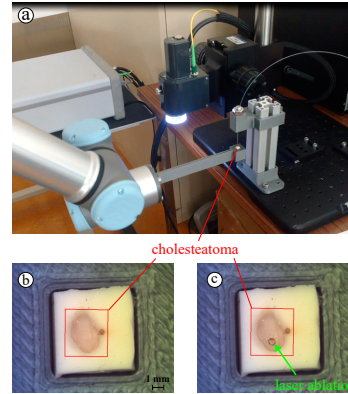


Fig. 1: (a) Experimental Setup. (b) Cholesteatoma before and (c) after a laser ablation.

is manually positioned in contact with the sample before each laser pulse. The distance is therefore considered constant throughout the study. Then, the tested laser pulse was shot to vaporize a small amount of cholesteatoma tissue. To estimate this vaporized volume, 3d-images (C-scan) of the sample were acquired using a TELESTO-II OCT with an LSM03 probe. The voxel dimensions were 25 μ m along the X and Y axes and 3.5 μ m along the Z axis (optical axis). Eventually, a UR3 collaborative robot arm was used to sequentially position the sample under the laser instrument and the OCT.

B. Cholesteatoma segmentation

The main phases of the cholesteatoma segmentation are presented in Fig. 2. First, a threshold is applied to roughly separate the volumes of matter (cholesteatoma and support) from the background (Fig. 2b). In order to remove the support, we extract from the obtained image two arbitrary cross-sections parallel to the (0xz) plane that only contain the support element, not the tissue (e.g., at the boundaries of the image). The outlines of these cross-sections are then obtained using the Canny edge detector (*edge* function in Matlab). As the support has a prismatic shape, its surface can be approximated by performing a linear interpolation between the outlines detected in the previous step. This can then be used to subtract the

^{*}Q. Boyer and D. V. A. Nguyen contributed equally to this work.

¹<https://anr.fr/Project-ANR-17-CE19-0005>

²The samples used are human surgical wastes collected under ethical approval.

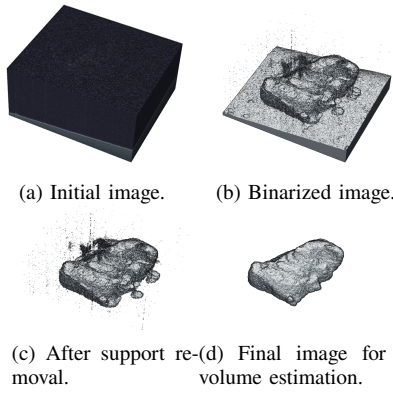


Fig. 2: Steps of cholesteatoma OCT image processing.

support from the binary image obtained after thresholding. At this stage, the image may still contains noises: either false negatives (holes in the middle of the matter) or false positives (matter in the background) (Fig. 2c). The *imfill* function in Matlab is used to remove the false negatives. It detects and fills the holes characterized by background pixels that cannot be accessed from the boundary pixels of the image through the face connectivities (6-connected mode). The false positives are filtered using the opening morphological operation (*imopen* function in Matlab) with a structuring element based on a $50\mu\text{m}$ -radius ball. The solid component corresponding to the cholesteatoma is then extracted using the *bwselect3* function of Matlab (Fig. 2d). Its volume can be simply estimated by counting the voxels and multiplying the result by the voxel element volume. The computation time for the segmentation is about 5 s for a volume of $171 \times 193 \times 768$ voxels using a 1.6 GHz Intel Core i5.

C. Vaporized volume of cholesteatoma

The vaporized volume can be calculated as the difference of volume in the images before and after the laser ablation. The experiments were conducted in the hours following the cholesteatoma removal. Positioning the sample under the fiber tip takes around 1 minute and the scanning time is 20 seconds so each iteration takes a few minutes. The time delay between the operation and the experiments should be kept as low as possible because time alters the optical properties of the cholesteatoma [2].

III. RESULTS

The experimental results of the vaporized volumes with respect to variant time exposures of the laser pulse for two different laser powers, 6 W and 8 W, are presented in Figure 3. In this experiment, the OCT can measure an ablation volume of 0.03 mm^3 (its resolution is even about $2.2 \times 10^{-6}\text{ mm}^3$ for each voxel volume). One can confirm that the vaporized volume increases linearly with the exposure time. Moreover, the steady-state ablation rate increases when the laser power increases. This experimental result is consistent with the computational model in the literature [3]. As the tissue needs to be heated from its initial temperature (20°C) to the water vaporization

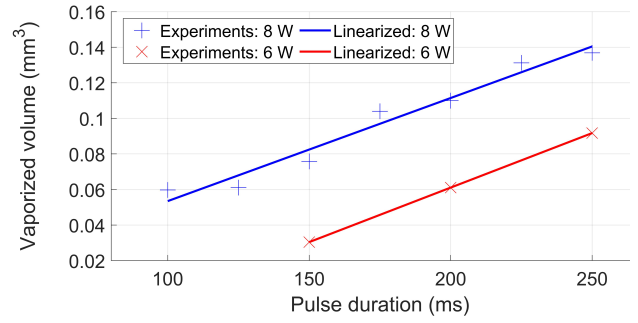


Fig. 3: Vaporized volume as a function of exposure time.

temperature (100°C), there exists a minimum value of the pulse duration so that cholesteatoma can be vaporized with a given laser power. This limited value can be estimated using linear extrapolation from the vaporized volume - pulse duration line (e.g., 6 W and 105 ms).

IV. CONCLUSIONS AND DISCUSSION

This study provide an effective method for cholesteatoma laser ablation monitoring using OCT and the proposed image processing. Based on the obtained quantitative results, the surgeon can select the proper laser parameters (power and time exposure) as well as the required number of shots for any specific volume of residual cholesteatoma during the laser ablation process. More experiments can be conducted to obtain a complete description of the relationship between the cholesteatoma vaporized volume and the laser parameters.

As an ex-vivo study, there are probably several differences in the results comparing to the in-vivo practice such as: the presence of blood in the tissue, the higher temperature in human body (37°C), or the tissue shrinkage due to natural evaporation of water in ex-vivo case. The best solutions to these limitations are making the experimental environment close to the middle ear condition and avoiding intraoperative bleeding.

Future work will focus on the integration of the OCT probe into our hybrid concentric tube robot [4] to perform cholesteatoma detection and laser ablation monitoring. The integrated OCT probe will also be used to control the distance between the fiber tip and the cholesteatoma during the ablation and study its impact on the ablation process.

REFERENCES

- [1] G. le Nobel and A. James, "Recommendations for Potassium-Titanyl-Phosphate Laser in the Treatment of Cholesteatoma," *The Journal of International Advanced Otolaryngology*, vol. 12, Nov. 2016.
- [2] E. Wisotzky, P. Arens, S. Dommerich, A. Hilsmann, P. Eisert, and F. Uecker, "Determination of optical properties of cholesteatoma in the spectral range of 250 to 800 nm," *Biomedical Optics Express*, vol. 11, pp. 1489–1500, 02 2020.
- [3] H. El-khalil, A. Alshare, G. Shafirstein, and J. Bischof, "A three-dimensional transient computational study of 532-nm laser thermal ablation in a geometrical model representing prostate tissue," *International Journal of Hyperthermia*, vol. 35, pp. 1–10, Oct. 2018.
- [4] D.-V.-A. Nguyen, C. Girerd, Q. Boyer, P. Rougeot, O. Lehmann, L. Tavernier, J. Szewczyk, and K. Rabenorosoa, "A hybrid concentric tube robot for cholesteatoma laser surgery," *IEEE Robotics and Automation Letters*, pp. 1–1, 2021, Early access, DOI: 10.1109/LRA.2021.3128685.

Augmented reality-assisted markerless hip centre measurement for total knee arthroplasty

Xue Hu and Ferdinando Rodriguez y Baena

Mechatronics in Medicine Lab, Imperial College London

xue.hu17@imperial.ac.uk

INTRODUCTION

The hip rotation centre (HRC) is an important anatomical reference point for total knee arthroplasty (TKA) [1]. The correctly measured HRC is vital to geometrically align a knee implant to the mechanical leg axis [2]. Conventionally, in many commercial orthopaedic surgical systems (e.g., Navio, Smith&Nephew) [3], an infrared (IR) marker is pinned in the distal femur and tracked by a stationary optical tracker. The surgeon manually pivots the limb while the pelvis is kept unmoved. The rotary trace is displayed on a monitor in real-time to guide the surgeon to cover a necessarily wide spread of points (Figure 1). Such a procedure, however, is invasive to the patient due to the marker incision, hard to interpret due to the nonintuitive guidance, and unclear in terms of accuracy since the computed hip location cannot be checked *in vivo* [4].

In this work, we present a head-mounted augmented reality (AR)-assisted markerless method for HRC measurement during TKA. A markerless segmentation and registration method [5] is adopted to track an assumed landmark at the femur end. No marker preparation, insertion or registration is required, resulting in a less invasive and tedious surgical workflow. We design an AR visual feedback system to provide an intuitive perception of where the collected landmarks locate in the physical world and how well the computed hip centre aligns with the observed anatomy. Through phantom trials, our proposed HRC measurement achieves a more straightforward workflow together with reliable accuracy.

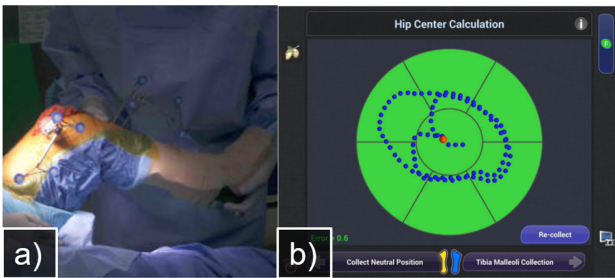


Fig. 1: User interface to guide the hip centre measurement in Navio surgical system. Figure adopted from [3].

MATERIALS AND METHODS

We demonstrate the proposed workflow on a phantom leg with the Microsoft HoloLens 2 (Figure 2). A commercial RGBD camera, the RealSense D415 (Intel Corp), is anchored to the HoloLens. The relative pose ${}^H_D T$ between the camera D and HoloLens front-facing camera H is calibrated offline through a quick stereo pose calculation procedure [6]. We adopted a previous published and validated markerless segmentation and registration algorithm [5] to track the partially exposed femur surface in the camera frame D : the femur points are first automatically segmented from RGBD captures by a pretrained neural network, then registered to a pre-scanned model for the real-time pose $F^{(D)}(t)$. The origin of the femur coordinate system ($p = [0, 0, 0, 1]^T$) was used as the “landmark” to avoid any bias due to rotational tracking errors. As shown in Figure 3, during limb rotation, the tracked landmark position is seamlessly sent to the HoloLens via Transmission Control Protocol over Internet Protocol (TCP/IP) and rendered as a red ball in real-time according to:

$$p^{(w)}(t) = {}^w_H T(t) \times {}^H_D T \times F^{(D)}(t) \times p \quad (1)$$

where ${}^w_H T(t)$ is given by the HoloLens self-localisation capability.

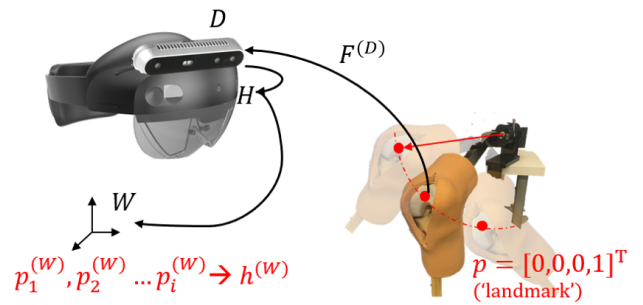


Fig. 2: Transformation of the markerlessly measured hip centre to HoloLens world for AR rendering.

If the user is satisfied with the tracked location, he/she can perform an “air tap” to collect the current landmark. The sphere will turn green and remain on display. Meanwhile, the confirmed $p^{(w)}$ is sent back to the PC and stored there for hip centre computation. Such landmark collection is repeated at least ten times, resulting in a set

of confirmed green balls distributed in the space around the hip centre (Figure 3). The PC is then used to compute the HRC by means of a 3D version of the hyper-accurate algebraic sphere fitting algorithm [7], [8]. The computed hip centre $h^{(w)}$ is finally sent back to the HoloLens for AR rendering. As more points are collected, the latest computed hip centre is displayed as a ball in a solid yellow colour, and the previous estimates are kept on display in fading yellow colours (Figure 3). When additional landmarks do not shift the current trail of hip centres, convergence is reached. The user can confirm the final hip location by voice command.

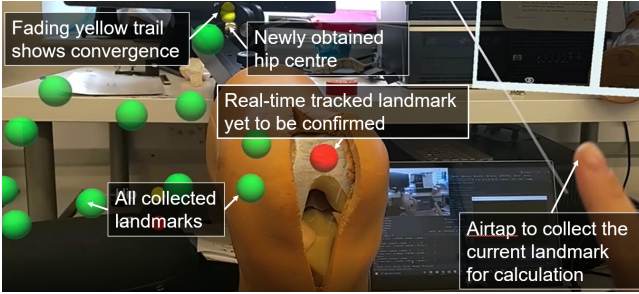


Fig. 3: AR interactive feedback to guide landmark collection for hip centre measurement.

EVALUATION AND RESULTS

As shown in Figure 4, we used a highly accurate optical tracking system A (FusionTrack 500, Atracsys LLC) as the ground truth to evaluate the HRC measurement accuracy. An IR reflective marker M_f was pinned into a phantom leg. The optical tracker A tracked the movement of M_f during leg rotation, from which the ground truth hip centre $h_{gt}^{(A)}$ was computed. To transform the computed location back into the HoloLens world frame, another marker M_d was attached to D . The relative transformation ${}_{M_d}^D T$ was calibrated offline using the method suggested in [8]. The ground truth hip location can be expressed by:

$$h_{gt}^{(w)}(t) = {}_H^W T \times {}_D^H T \times {}_{M_d}^D T \times {}_{M_f}^{M_d} T(t) \times h_{gt}^{(A)} \quad (2)$$

The error is defined as the positional difference between $h_{gt}^{(w)}(t)$ and $h^{(w)}(t)$. As shown in Figure 5, the average 3D error is [1.21, -8.45, 8.80] mm. According to [8], such an error in hip centre estimate would result in an overall registration accuracy that would be acceptable for clinical deployment.

CONCLUSIONS AND DISCUSSION

This work presents an AR-assisted markerless system that allows interactive HRC measurement avoiding the need for additional invasive markers. Proof-of-concept performance was tested on a phantom leg by simulating HRC measurement during an open TKA. The AR display helps the surgeon be aware of how the past collection distributes in space so that he/she can intuitively direct the limb to evenly cover a wide spread for reliable HRC calculation. Furthermore, the computed hip location can be easily checked by comparing it to the naturally observed limb. Our experiment shows promising accuracy

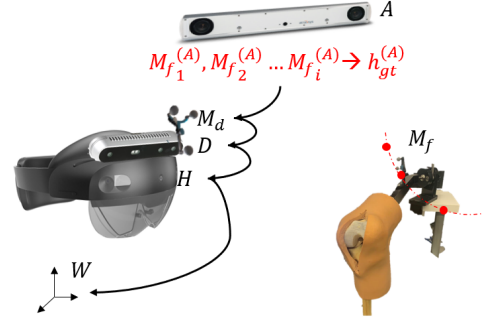


Fig. 4: Transformation of ground truth hip centre obtained by optical tracking for accuracy evaluation.

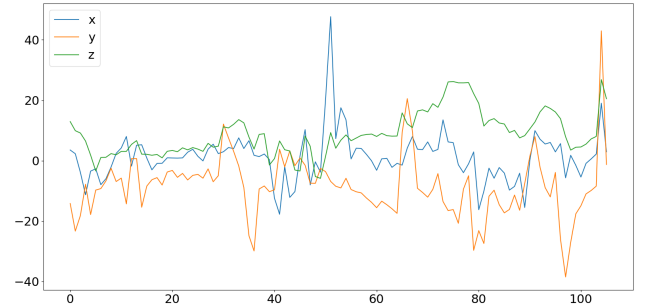


Fig. 5: AR interactive feedback to guide landmark collection for hip centre measurement.

with a much more straightforward and immersive workflow. In the future, we will integrate the suggested HRC measurement into a complete knee surgical procedure to demonstrate the benefit brought by an AR display and markerless tracking on surgical workflow and patient safety.

REFERENCES

- [1] H. Bombaci, B. Simsek, M. Soyarslan, and M. M. Yildirim, "Determination of the hip rotation centre from landmarks in pelvic radiograph," *Acta orthopaedica et traumatologica turcica*, vol. 51, no. 6, pp. 470–473, 2017.
- [2] X. Huang, X. Wei, and C. Wang, "Accuracy assessment for hip rotation centre in total knee replacement surgical navigation," in *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2008, pp. 514–517.
- [3] Surgical technique for use with the journey ii bcs and journey ii cr. [Online]. Available: https://www.smith-nephew.com/global/assets/pdf/products/surgical/navio_tka_st_manual_500081revb.pdf
- [4] G. Matziolis, D. Kroker, S. Tohtz, U. Weiss, and C. Perka, "Accuracy of determination of the hip centre in navigated total knee arthroplasty," *Zeitschrift für Orthopädie und ihre Grenzgebiete*, vol. 144, no. 4, pp. 362–366, 2006.
- [5] H. Liu and F. R. Y. Baena, "Automatic markerless registration and tracking of the bone for computer-assisted orthopaedic surgery," *IEEE Access*, vol. 8, pp. 42 010–42 020, 2020.
- [6] X. Hu, F. R. y Baena, and F. Cutolo, "Head-mounted augmented reality platform for markerless orthopaedic navigation," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [7] A. Al-Sharadqah, N. Chernov *et al.*, "Error analysis for circle fitting algorithms," *Electronic Journal of Statistics*, vol. 3, pp. 886–911, 2009.
- [8] X. Hu, H. Liu, and F. R. Y. Baena, "Markerless navigation system for orthopaedic knee surgery: A proof of concept study," *IEEE Access*, vol. 9, pp. 64 708–64 718, 2021.

WAYNet: Weakly-Supervised Multi-Task Anastomoses Detection in Fetoscopy

Alessandro Casella^{1,2}, Gaia Romana De Paolis², Elena De Momi², Dario Paladini³,
Sara Moccia⁴, and Leonardo S. Mattos¹

¹*Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy*

²*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*

³*Department of Fetal and Perinatal Medicine, Istituto “Giannina Gaslini”, Genoa, Italy*

⁴*The BioRobotics Institute and Department of Excellence in Robotics & AI, Scuola Superiore Sant’Anna, Pisa, Italy*

INTRODUCTION

Twin-to-Twin Transfusion Syndrome (TTTS) is a rare pathology that may affect monochorionic twin pregnancies. TTTS depends on the unbalanced blood transfer from one twin (the donor) to the other (the recipient) through abnormal placental vascular anastomoses. Currently, the treatment for TTTS consists of the photo-ablation of abnormal anastomoses in fetoscopic laser surgery [1]. Residual anastomoses still represent a major complication [2] and their identification is not a trivial task. Visual challenges such as small field of view, amniotic fluid turbidity, low-resolution imaging, and unfavourable views are due to the position of the insertion site for the tools. To support surgeons, researchers are working on vessel and placenta segmentation [3], [4]. Recently, [5] presented the first multi-centre large-scale dataset to improve the current state-of-the-art in segmentation and registration in fetoscopy. However, to date, there is no work in the literature on anastomosis detection. There is also no available datasets for this task.

This work aims to develop a deep-learning-based framework for anastomosis detection in intra-operative fetoscopic videos. Considering the challenges of labelling anastomoses, we propose a weakly-supervised strategy by training a multi-task convolutional neural network (CNN) for (i) segmenting vessels in the fetoscopy frame and (ii) classifying frames as containing anastomoses or not. Relying on class activation mapping (CAM), anastomosis detection is then accomplished.

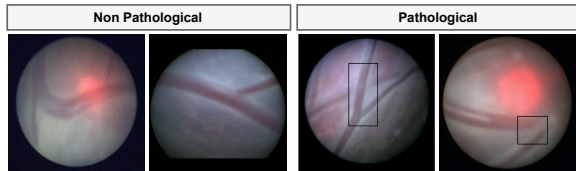


Fig. 1: A sample of the anastomoses dataset. Black bounding boxes highlight the pathological anastomosis.

MATERIALS AND METHODS

In this work, we want to exploit the representation learning capability of CNNs for weakly supervised anastomoses detection. CNNs can extract meaningful features from data, and CAM can be used for investigating the CNN

visual encoding process and localising anastomoses. As shown in Fig. 2, the proposed framework is composed by a backbone and three branches. The backbone comprises (i) a dense-feature encoder and (ii) a Feature Pyramid Network (FPN) decoder. The decoder is connected to a Y-shape head consisting of two branches, for vessel segmentation and frame classification. The localisation branch extracts activation maps from the last three segmentation blocks of the encoder to perform anastomoses detection with CAM.

The encoder is based on the DenseNet121 architecture. Dense connectivity in DenseNet is implemented to improve gradients’ flow among layers, avoiding the problem of gradient vanishing in deeper layers and improve network training efficiency. The decoder is designed as FPN to leverage feature hierarchy and semantics learning at different scales. All the feature maps from each pyramid level in the FPN are processed by a segmentation block to produce rough segmentation maps, one at each pyramid scale. The segmentation block contains a 3×3 convolution followed by *Group Normalisation* and *ReLU* activation. The partial segmentation maps are summed up and processed by the Y-shape head. The segmentation branch in the Y-shape head consists of 1×1 convolution for channel reduction, bilinear upsampling to recover the original size, and sigmoid activation. The classification branch is made of a Global Average Pooling (GAP) layer followed by three fully connected layers. We add one dropout layer to each of the first two layers to regularise the training process [6]. In the localisation branch, we compute the weighted sum of all Gradient-weighted Class Activation Map (GradCAM) for the segmentation blocks in the decoder, which represents the feature map response at each scale. The overall activation map is then rescaled using MinMax scaling, and a threshold is applied to extract regions of interest that may contain anastomoses. We experimentally found that 0.75 is a reasonable threshold value.

A. Experimental protocol

To develop and test our framework, we built a dataset by merging a publicly available dataset¹ presented in [4]

¹<https://www.ucl.ac.uk/interventional-surgical-sciences/fetoscopy-placenta-data>

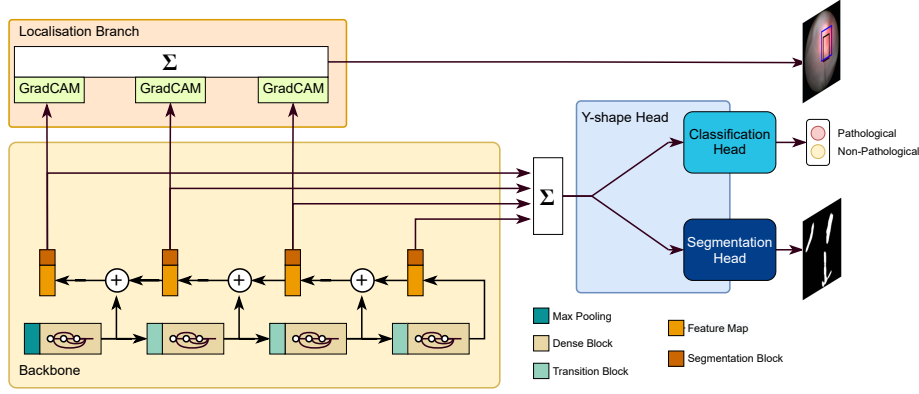


Fig. 2: Overview of the proposed WAYNet framework.

TABLE I: Results of the ablation studies for the backbone network. In the baseline the Segmentation Head is missing.

		Classification			Segmentation	Detection
		Acc	Rec	Prec	IoU	IoU
Baseline	ResNet50	0.6200	0.0570	0.7500	-	0.0424
	ResNet50	0.6827	0.5377	0.6064	0.5770	0.0062
Fine-Tuning	Dense	0.7380	0.5849	0.6966	0.5426	0.0217
	ResNet50	0.9889	0.9811	0.9905	0.5976	0.0570
Multi-Task	Proposed	0.9668	0.9906	0.9292	0.5314	0.2206

that we extended with additional data collected at *Istituto Giannina Gaslini, Italy* for a total of 18 TTTS procedures. We manually extracted 1476 frames and asked an expert clinician to annotate vessels and the presence of pathological anastomoses in each frame. 27.6% of frames includes pathological anastomosis. A sample of our dataset is shown in Fig. 1. For testing the detection performance, we further annotated 83 frames from 3 additional patients with bounding boxes. The overall framework was trained end-to-end using a combination (L_{Overall}) of two loss functions. L_{Overall} is defined as: $L_{\text{Overall}} = L_S + L_C$, where L_S is the binary cross entropy used for the segmentation task and L_C is the weighted cross entropy used for the classification task. We used a weighted cross entropy to account for class imbalance.

We evaluated the performance of the proposed framework in terms of accuracy (Acc), recall (Rec), and precision ($Prec$) for the classification task. As for the segmentation and detection task, intersection over union (IoU) was utilized. A Wilcoxon test assessed the statistical significance of the results. In the ablation study, we investigated several configurations and training strategies for our framework. **Baseline** consists of a classification network based on ResNet50. In **fine-tuning** configurations (ResNet50 and Dense-FPN), we trained the backbone on vessel segmentation and then on the classification task, while in **Multi-Task** tests (ResNet50 and Proposed), we trained the backbone on both tasks end-to-end.

RESULTS

As shown in Table I, the proposed framework achieved good performance in classification ($Acc = 0.9668$, $Rec = 0.9906$, $Prec = 0.9292$) and segmentation ($IoU = 0.5314$). For anastomoses detection we achieved an $IoU = 0.2206$. Our framework performed comparably with the other

tested approaches for the classification and segmentation task, while it showed by far the best performance for the anastomosis detection task.

CONCLUSIONS AND DISCUSSION

This paper presented a first step towards the automatic localisation of anastomoses in TTTS surgical videos. Our framework showed promising results and outperformed tested approaches for weakly-supervised anastomoses detection. The vessel segmentation and anastomoses classification performance of our framework is reliable, with few misclassifications. Sometimes pathological and non-pathological anastomoses may look very similar, affecting classification performance, especially with such small datasets. Contrastive learning approaches might tackle this challenge. However, the high imbalance between classes could be an issue for those techniques. Collecting more data and data stratification are reasonable improvements to our experimental protocol. Despite achieving promising results, there is still work to do to tackle the complexity of the task.

REFERENCES

- [1] Roberts et al. Interventions for the treatment of twin-twin transfusion syndrome. *Cochrane Database of Systematic Reviews*, 2014(1), 2014.
- [2] Lopriore, et al. Residual Anastomoses After Fetoscopic Laser Surgery in Twin-to-Twin Transfusion Syndrome: Frequency, Associated Risks and Outcome. *Placenta*, 28(2-3):204–208, 2 2007.
- [3] Casella, et al. A shape-constraint adversarial framework with instance-normalized spatio-temporal features for inter-fetal membrane segmentation. *Medical Image Analysis*, 70:102008, 2021.
- [4] Bano, et al. Deep placental vessel segmentation for fetoscopic mosaicking. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 763–773. Springer, 2020.
- [5] Bano et al. FetReg: Placental Vessel Segmentation and Registration in Fetoscopy Challenge Dataset. *ArXiv*, 2021.
- [6] Lee, et al. Robust tumor localization with pyramid grad-cam. *ArXiv*, abs/1805.11393, 2018.

An optimization method for surgical targeting planning

Chiara Tenga , Andrea Roberti , and Paolo Fiorini

Department of Computer Science, University of Verona, Verona, Italy
{name.surname}@univr.it

INTRODUCTION

The research interest in Robotic-assisted Minimally Invasive Surgery (R-MIS) is shifting from teleoperated devices to autonomous robots and to the development of autonomous support systems for the execution of repetitive surgical steps, such as suturing, ablation, needle biopsy and microscopic image scanning [1]. Needle biopsy [2] is currently the most reliable and widely used technique to confirm the suspicion of prostate cancer (PCa), the second most common cancer worldwide and the cancer most common in men [3]. Early and reliable detection of PCa has a huge impact on the successful treatment of high-risk patients and on avoiding overtreatment in low-risk patients. In [4] we presented PROST, a robotic device for prostate biopsy, with the aim to introduce some autonomy during surgery in order to assist surgeons. In this paper, we propose a planning procedure to optimize the biopsy in terms of both patient recovery and prostate coverage.

The proposed method applies a voxel based spatial subsampling to the prostate volume and uses graph theory to select the lowest number of insertion points that reach the targets to be analyzed. In this way we optimize the choice of the entry points and minimize the duration of the intervention. However, the needle insertion is manual so that the surgeon has complete control of the surgical task.

MATERIALS AND METHODS

We propose a planning procedure which consists of two phases:

- We start with a voxel based spatial subsampling of the prostate to obtain targets ($t_{i...n} \in T$) throughout the volume, as shown in Fig. 1. The targets are calculated as the centroids of the voxels;
- We divide the entry ($p_{i...m} \in P$) and the target ($t_{i...n} \in T$) points in a weighted bipartite graph, as shown in Fig. 2.

Given the number m of entry points, our goal is to try to cover the highest number of targets with a single entry point. The entry points are chosen on the plane of the perineum, and their position is taken by the kinematics of the robot. Once we have the position of the entry points and target points we construct the weighted bipartite graph. When the algorithm starts, it searches for a node with cardinality $C = K - l$, where K is the number of targets and l are the cycles performed. If one node is found then that entry point reaches all targets and is the

solution. If it does not exist, it searches for a node with $K - (l + 1)$ cardinality, if it exists, this node is part of the solution and we check which target does not touch and find any entry point that reaches it. If we don't find it, we have to search through all the combinations of nodes that touch each target, focusing on cardinality. The solution is represented through the *incidence matrix*. The elements of the matrix, which therefore represent the weights, are computed using the optimization parameters for the type of intervention (e.g. biopsy, cryoablation, vertebroplasty, kyphoplasty, etc).

One of the most important optimization parameter, which consequently has a greater weight, is the robot kinematic. We check if the target is in the robot workspace and if is reachable by the instrument/needle. Another optimization parameter is the presence of obstacles, such tissues or veins, to avoid in order to do not compromise the intervention. Also the execution time: a shorter intervention has multiple benefits like lower costs. In case of biopsy, we must consider the stiffness of the tissues we pass through to reach the target avoiding needle deflections, also taking into account the length of the needle defines the depth of the targets that can be reached.

After the incidence matrix has been constructed, we choose as first entry point the one with the highest cardinality, sorting on the targets with a lower weight in the matrix. This means we reach the easier targets first.

EXPERIMENTAL SETUP

The experimental setup consists of the PROST robotic system prototype [4] which integrates the US Ultrasonix system which allows us to obtain the volume of the prostate.

To validate our algorithm we used a synthetic phantom representing a prostate. The phantom was also acquired in computed tomography (CT) so that a 3D volumetric medical image can be available for image fusion with those acquired in real-time. The semi-automatic segmentation approach provided by 3D Slicer [5] is exploited to extract the contour of the prostate.

Table I shows an example of the incidence matrix created with our algorithm during a biopsy procedure with 3 entry points. The weights of the matrix were calculated using kinematics and execution time as optimization parameters. In this case, we voxelise the volume in 14 blocks, we extract the centroid of each block representing 14 targets ($t_{1..14}$). In the incidence matrix we have column

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}
p_1	64.29	0	56.43	64.70	63.47	0	55.29	53.05	54.64	65.47	0	59.10	56.94	53.62
p_2	64.74	0	56.97	0	65.37	0	57.23	54.99	56.37	0	0	59.64	58.77	55.32
p_3	64.03	0	55.82	64.10	63.10	0	54.27	51.80	53.49	65.06	0	57.38	56.17	52.99

TABLE I: Incidence matrix of the needle biopsy experiment.

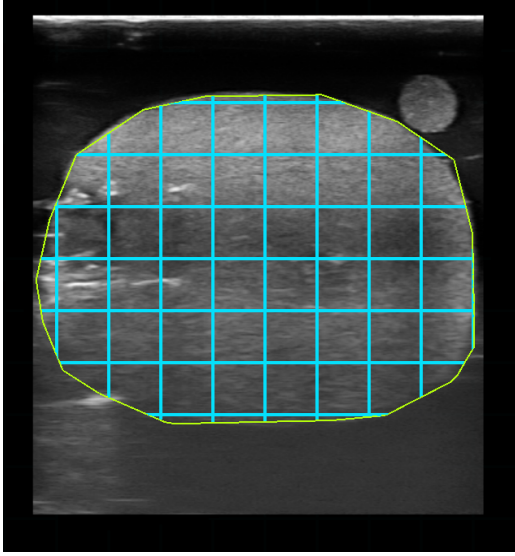


Fig. 1: Example of 2D projection of the voxel representation in a single slice.

of zeros when a target is not reachable from any entry point (t_2, t_6, t_{11}).

In this experiment we have both p_1 and p_3 with the same cardinality ($C = 11$), but we choose as solution p_3 since the sum of the weights of the incidence matrix for that entry point is less than other entry points. Once we have the entry point we sort the target list to obtain the final planning.

$$PLAN(p_3) = (t_8, t_{14}, t_9, t_7, t_3, t_{13}, t_{12}, t_5, t_1, t_4, t_{10})$$

CONCLUSIONS AND DISCUSSION

In this work, we proposed a planning procedure which optimize the choice of entry points in reaching as many targets as possible. We tested our algorithm in simulation and on a synthetic phantom for a needle biopsy procedure. The use of a lower number of insertion points can reduce trauma to the patient without sacrificing the accuracy of a template biopsy; it also leads to a quicker execution. We aim to increase the accessibility of high-quality prostate biopsy thorough compatibility with outpatient settings, quick setup and fast intervention time independently of the expertise level of the operating physician. The use in outpatients' clinics will potentially allow time and cost reduction of the procedure in comparison with template biopsy, which is currently carried out in the operating room.

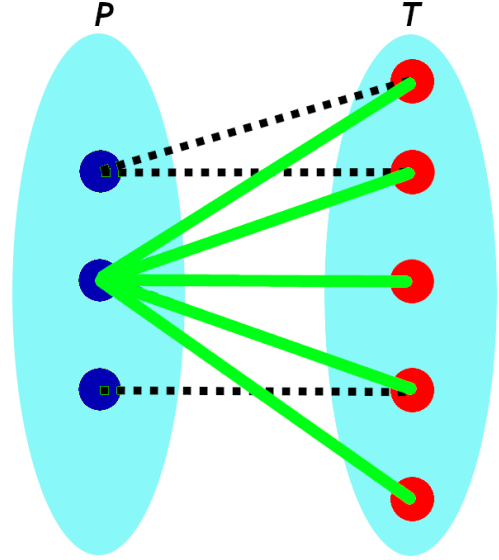


Fig. 2: Bipartite graph: entry points $p_{i...m}$ on the left (blue) and target points $t_{i...n}$ on the right (red). The green lines represent the solution, therefore the connection between p and t with less weight.

REFERENCES

- [1] E. Oleari, A. Leporini, D. Trojaniello, A. Sanna, U. Capitano, F. Dehó, A. Larcher, F. Montorsi, A. Salonia, F. Setti *et al.*, "Enhancing surgical process modeling for artificial intelligence development in robotics: The saras case study for minimally invasive procedures," in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*. IEEE, 2019, pp. 1–6.
- [2] S. Verma, P. L. Choyke, S. C. Eberhardt, A. Oto, C. M. Tempany, B. Turkbey, and A. B. Rosenkrantz, "The current state of mr imaging-targeted biopsy techniques for detection of prostate cancer," *Radiology*, vol. 285, no. 2, pp. 343–356, 2017.
- [3] C. H. Pernar, E. M. Ebot, K. M. Wilson, and L. A. Mucci, "The epidemiology of prostate cancer," *Cold Spring Harbor perspectives in medicine*, vol. 8, no. 12, p. a030361, 2018.
- [4] B. Maris, C. Tenga, R. Vicario, L. Palladino, N. Murr, M. De Piccoli, A. Calanca, S. Puliatti, S. Micali, A. Tafuri *et al.*, "Toward autonomous robotic prostate biopsy: a pilot study," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 8, pp. 1393–1401, 2021.
- [5] S. Pieper, M. Halle, and R. Kikinis, "3d slicer," in *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*. IEEE, 2004, pp. 632–635.

Surgical knot training in ophthalmic surgery: Skill assessment with eye-tracking

D. Anastasiou¹, L. Raja^{2,3}, G.M. Saleh^{1,2}, D. Stoyanov¹, E. Mazomenos¹

¹Wellcome / EPSRC Centre for Interventional and Surgical Sciences, UCL, UK

²NIHR Biomedical Research Centre for Ophthalmology, Institute of Ophthalmology, UCL, UK

³Imperial College Healthcare NHS Trust, UK
dimitrios.anastasiou.21@ucl.ac.uk

INTRODUCTION

Suturing is a fundamental task in ophthalmic surgery. Focused training is necessary to master technical (tissue handling, knot tying) and cognitive (appropriate selection of instruments, forward planning) skills and develop a high level of hand-eye coordination required for ophthalmic microsurgical procedures. Formulating novel objective measures of operational performance will be beneficial for training in ophthalmic microsurgery.

Capturing eye movements and points of focus while performing surgical tasks can provide meaningful information to assess the operator's technical and cognitive skills and overall performance. The locations of gaze focus and spatial distribution of fixations embed valuable information for assessing the use of instruments, the sequence and quality in executing subtasks, and possibly the level of hand-eye coordination the operator demonstrates. This study explores eye-tracking for developing performance metrics for suturing tasks in ophthalmic surgery and preliminary analysis focuses on the total duration of executing a surgical suture and its subtasks. It also introduces the spatial distribution of fixations as a feature to characterize the level of surgical expertise.

Eye-tracking has been used as a tool for skill analysis in a variety of different surgical applications. Copogna et al. [1], compared anesthesiologists of different expertise levels performing an epidural block. Attentional heat-maps and gaze plots showed different gaze dispersion between the groups. Causer et al., showed that quiet eye training significantly improved learning of surgical knot tying compared to a traditional technical approach [2]. In [3], expert and novice neurosurgeons performing under a surgical microscope were examined, concluding that experts spend more time fixating on the region of interest before performing an action. Lee et al., used eye-tracking data to identify gaze patterns and blind spots in a real-time EGD [4].

While efforts have been made to analyze gaze patterns [3],[4], it has yet to be developed a metric that can be used to statistically compare the spatial distributions of gaze focus points, a rather useful tool to evaluate the skill level of groups with different expertise.

MATERIALS AND METHODS

This study was conducted in the Royal College of Ophthalmologists (London, UK) during the training

course "Introduction to Ophthalmic Surgery". It includes two one-hour long training sessions for practicing surgical knots. Prior to practicing, participants were educated through lectures, videos, and live demonstrations. In the first session, attendees were instructed to practice the surgical knot on a standard suture training board, where on average they completed five knots and continued with more demonstrations before the second practice session.

Equipment

For capturing eye movements, the Tobii Pro Glasses 2 (Stockholm, SWE) was used and deployed as shown in Fig. 1(a). Tobii glasses use near-infrared illumination to create reflection patterns on the cornea and pupil of the subject's eye, while image sensors are used to capture images of the eyes with the reflection patterns. Image-processing and a physiological 3D model of the eye are then used to estimate the point of gaze within the field of view (FoV) provided by a camera at the front side of the glasses. Time-stamped gaze focus points (pixel coordinates) are synchronized with the FoV video recording and overlaid (red circle) as shown in Fig. 1(b).

Participants

We recruited 20 volunteers, 9 *experts* – faculty members delivering the course and 11 *novices* – trainees attending the course, in the early stage of their residency training. Of the 11 novices, we selected 6 from the first session and 5 from the second one, to compare novices that had more time to practice. All participants were asked to execute a single surgical knot on the training board, whilst wearing the eye-tracking glasses. After obtaining basic familiarity with the task, we recorded one video from each participant, and calibration took place before every recording.

Dataset and Annotation

A surgical knot consists of two major subtasks, as shown in Fig. 2. These are needle passing and knot tying, where typically surgeons tie three knots to secure the suture. By reviewing the videos, we annotated the start and end moments of each subtask, as shown in Fig. 2.



Figure 1. (a) Experimental setup, (b) Snapshot of a video showing the wearer's field of view along with gaze focus point.

Inspired from the literature [1],[2], we then calculated i) the total duration of the entire suturing task, which commences with the needle passing and ends with suture cutting; ii) the duration of the needle passing subtask, as well as iii) the duration of tying the first knot. These time-based metrics were further used for statistical analysis.

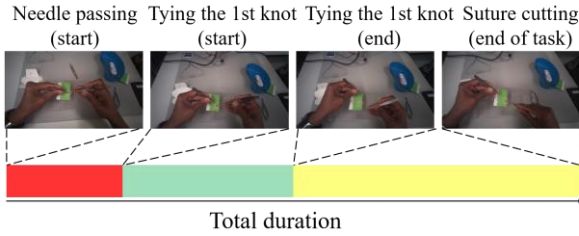


Figure 2. Annotation of a surgical suture task from our dataset.

In order to aggregate gaze focus points in a single image, we should assume that the subject's FoV doesn't change. During knot tying head movements were limited, in contrast with the needle passing phase, hence we assume that the FoV remained mostly the same. This allowed us to investigate the spatial distribution of gaze samples and fixations during that phase. Gaze samples are classified as part of a fixation, if the velocity of directional shifts of the eye is below 30 visual degrees per second. The distribution of fixations is illustrated as a heatmap plot (see Fig. 3) in which warmer colors signify a larger concentration of fixations observed on that point. From Fig. 3, it is evident that differences in the spatial distribution of fixations between the expertise groups exist. Experts' fixations are clustered together in smaller areas than novices' ones. To quantify the spatial distribution of fixations during the knot tying phase., we introduce *fixation sparsity*, as the ratio of the total number of color-valued pixels (locations that we obtain fixations) in the heatmap divided by the image resolution, 1080×1920 .

RESULTS

The four variables, total duration of the task, needle passing duration, first knot tying duration, and fixation sparsity were analyzed using the Mann-Whitney U-test with $p = 0.05$ significance level (two-tailed). Results are summarized in Table 1, where one can observe significant differences between the groups of experts and novices. No significant differences ($p > 0.05$) between novices from the first session and the second one were found.

Table 1. Mann-Whitney U-test results with $p = 0.05$ significance level (two-tailed).

Dependent Variable	Median/Min/Max		p-value
	Experts ($n_e = 9$)	Novices ($n_n = 11$)	
Total duration	70.12/46.36/83.32s	98.08/55.6/191.2s	0.001
Needle passing duration	19.04/35/18.96s	30.12/15.36/79s	0.04
1st Knot tying duration	12.68/5.6/18.96s	16.44/13.7/40.24s	0.006
Fixation sparsity	0.0362/0.0215/0.0486	0.0516/0.0216/0.0697	0.008

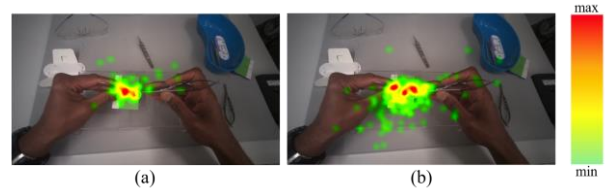


Figure 3. Example of the spatial distribution of fixations during knot tying (a) expert surgeon, (b) novice surgeon.

CONCLUSION AND DISCUSSION

In this work, we compared expert and novice surgeons based on their duration in executing a surgical knot as well as its different subtasks. Overall, expert surgeons were found to be significantly faster than their novice counterparts. Our analysis on the spatial distribution of fixations concluded that experts focus their gaze almost entirely on areas of interest (e.g., suture, tools) when tying the knots, resulting in the distribution, being concentrated in the area around the suture (see Fig. 3a). On the contrary, novices' distribution is more dispersed, with fixations occurring across a larger area (see Fig. 3b). A possible explanation is that novices need to visually verify that they are using the appropriate instruments. We also found that when tightening the knot, novices focus their gaze on the tools pulling the thread, rather than on the suture. Analyzing the fixation distribution can indicate areas of high/low interest, and patterns associated with levels of expertise, contributing towards novel metrics for skill assessment in ophthalmic surgery. These interesting observations can potentially lead to objective metrics for assessing the level of hand-eye coordination. Future work will focus on developing novel eye-tracking features and examining their correlation with standardized manual assessment, as well as on algorithms for predicting surgical skills.

ACKNOWLEDGEMENTS

This work was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences at UCL (203145Z/16/Z, NS/A000050/1); Dimitrios Anastasiou is supported by an EPSRC DTP and ISAD award (EP/R513143/1, EP/T517793/1).

REFERENCES

- [1] Capogna E., et al., "Novice and Expert Anesthesiologists' Eye-Tracking Metrics During Simulated Epidural Block: A Preliminary, Brief Observational Report", *Local Reg*; 2020; 2020(13): 105-109.
- [2] Causer J., et al., "Quiet eye training improves surgical knot tying more than traditional technical training. A randomized, controlled study", *Am J Surg*; 2014; 208(2): 171-177.
- [3] Chainey J., et al., "Eye-Hand Coordination of Neurosurgeons: Evidence of Action-Related Fixation in Microsuturing", *World Neurosurg.*; 2021; 155:196-202
- [4] Lee A., et al., "Identification of gaze pattern and blind spots by upper gastrointestinal endoscopy using an eye-tracking technique", *Surg. Endosc*; 2021.

Fast and Automatic Optical 3D Cameras Calibration for Contactless Surface Registration in Computer Assisted Surgery

Lénaïc Cuau¹, Marie de Boutray^{1,2}, João Cavalcanti Santos¹, Philippe Poignet¹, and Nabil Zemiti¹

¹LIRMM, University of Montpellier, CNRS, Montpellier, France

²Department of Maxillofacial Surgery, Gui de Chauliac University Hospital, Montpellier, France

INTRODUCTION

Computer Assisted Surgery (CAS) has expended over the years and is now commonly used in orthopedic surgery. It was shown that CAS improves functional outcomes compared to conventional non-assisted surgeries [1]. However, to implement such assistance, a precise registration between the patient's preoperative CT-scan data and the patient's intraoperative data in the operating room (OR) is required.

Artificial markers are the gold standard for this registration but imply a lot of drawbacks such as patient discomfort, hospital flow complications, etc. Surface-based registration methods are preferred and are usually implemented in the OR. The goal is to register the preoperative patient's bone surface data extracted from a CT-scan volume with the corresponding intraoperative surface data digitized with a navigated tracked stylus.

However the use of tracked stylus for surface digitization is not always accurate due to its calibration precision and tissue deformation upon contact. Moreover, for large surface acquisition, it is time consuming and user-dependent.

Camera-based solutions (laser range scanner, time-of-flight cameras, structured-light cameras, etc.) for patient's bone surface digitization have become quite common nowadays in CAS since they are faster, contactless and have less sterilisation constraints than tracked stylus [2]. They allow to acquire thousands of points in a few seconds without any contact. For example, these advantages inspired the application of structured light [2] in spinal surgery. However, precise camera calibration, namely "Hand-Eye" (H-E) calibration, is necessary to localize the camera's acquired data in the physical OR space. This calibration consists in finding the transformation between the camera and the robot end-effector that handles it or between the camera and a tracked marker fixed to it: the problem remains the same.

For 3D cameras, it has been shown in [3], [4] that exploiting the depth information gives better H-E calibration results than exploiting only the RGB information. In this context, different methods based on the use of camera depth information have been proposed in the literature. Khan *et al.* [3] proposed an iterative method based on the palpation, with a probe handled by a robot, of a complex calibrated wavy 3D object to

digitize and register its surface in the robot frame. The surface of this 3D object is also digitized with the optical 3D camera. The desired H-E calibration has been thus obtained by registering the obtained two sets of points. Nevertheless, the obtained results depend on the probe calibration precision, the quality of palpation and the quality of 3D-printing of the calibrated object. Yang *et al.* [4] proposed to use a simpler 3D object (namely a sphere) for the H-E calibration. In this method, thanks to the depth information, the 3D object has been digitized and its center has been computed using a RANdom SAMple Consensus (RANSAC) algorithm. This allows to estimate the translation between the camera and the 3D object and compute the H-E transformation thanks to a non linear optimization. Their method gave non-satisfying results (2.864 mm) in accordance with surgical application.

In our study, a fast and automatic method adapted to optical 3D cameras and based on a 3D printed phantom is presented. This phantom is separated from the camera to reduce printing errors and deformation. It was thought to be as simple as possible to ease printing and registration. The proposed method is easy to implement and does not require any experience from the user. It also gives precision compatible with surgical application.

MATERIALS AND METHODS

The proposed camera calibration method is represented in Fig.1 where bT_a is the transformation from the coordinate system a to coordinate system b, and W , P , C , M_p , and M_c , represent respectively the frames attached to the world (localizer), the phantom, the camera, the marker on the phantom, and the marker on the camera. The marker was fixed on the camera with a 3D-printed support adapted to the camera and to the marker. Our objective is to compute the transformation ${}^CT_{M_c}$ and thus perform the 3D camera H-E calibration. The proposed approach is based on the digitization and registration of a tracked 3D-printed phantom equipped with a marker. The tracked phantom was designed as one compact single piece to be as simple as possible to be produced: a cube with a marker fused to it to minimize deformations and cubic shapes on its faces to facilitate their digitization and allow precise registration. The requested H-E transformation can thus be computed with the following equation:

$${}^CT_{M_c} = {}^CT_P {}^PT_{M_p} {}^{M_p}T_W {}^WT_{M_c}$$

In accordance with Fig.1, the transformations of this equation can be obtained separately:

- ${}^C T_P$ is computed with an Iterative Closest Point (ICP) registration of the phantom CAD (Computer Aided Design) model expressed in P with the phantom point cloud generated by the camera expressed in C .
- ${}^P T_{M_p}$ is known by CAD model of the calibration phantom.
- ${}^M_p T_W$ and ${}^W T_{M_c}$ are given by the optical tracking system.

This method has the advantage to be simple (it only requires the design and printing of the phantom), fast and does not necessitate any experience from the user. It is easy to use as the user intervention is requested only for ICP initialization: he/she has only to click on three characteristic corresponding points on the phantom CAD model and on the camera's point cloud to perform a first coarse registration. Those points are usually the cube corners that are easy to find.

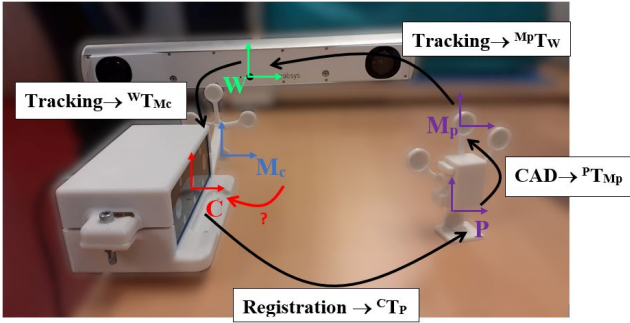


Fig. 1: Set up for camera calibration using a 3D-printed phantom and a localizer

RESULTS

To evaluate the camera calibration precision, another tracked 3D-printed phantom was designed and used as a validation geometry. This validation phantom consists in a marker and a cube designed as one single piece with three validation points on its faces. Each validation point corresponds to the center coordinate of a holes (a half-sphere of $\varnothing 1$ mm) engraved at the surface of the phantom.

The position of each hole in the camera frame, noted $\{{}^C P_i\}_{i=1:3}$, is obtained using the CloudCompare software (www.cloudcompare.org) by a mouse click on the center of hole as it is clearly visible in the screen. The position of each hole in the marker frame, noted $\{{}^M_p P_i\}_{i=1:3}$, is known by CAD model design. The ${}^M_p P_i$ points are reprojected in the camera coordinate system through the localizer information and the found H-E transformation using ${}^C T_{M_c} {}^M_c T_W {}^W T_{M_p} {}^M_p P_i$.

To evaluate the calibration precision, we thus compute the euclidean distance between the reprojected and observed holes:

$$\|{}^C P_i - {}^C T_{M_c} {}^M_c T_W {}^W T_{M_p} {}^M_p P_i\|$$

We performed 25 calibrations followed by a systematic validation step after each calibration process. The

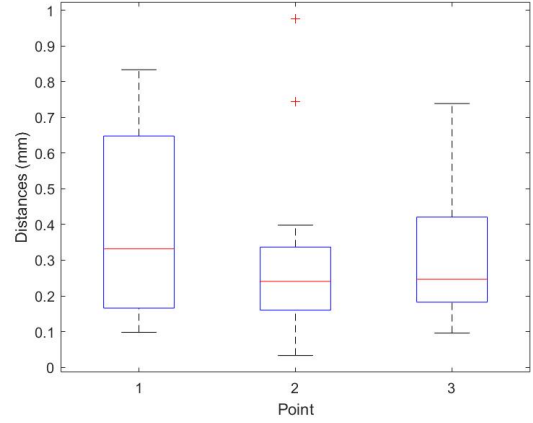


Fig. 2: Distances between detected and registered points on 25 calibrations

global mean distance between the observed and reprojected points was 0.324 mm with a standard deviation of 0.206 mm. More detailed results for each validation point are given on Fig.2. Distances between the observed and reprojected points were 0.388 ± 0.234 mm, 0.279 ± 0.199 mm, 0.303 ± 0.169 mm for point 1, 2 and 3 respectively. It is important to notice that all the values are inferior to 1 mm. Moreover, the calibration method is fast: the whole procedure, including point cloud generation, user clicks for ICP initialization, and ICP registration, takes on average 37.5 ± 2.8 s.

CONCLUSIONS AND DISCUSSION

Using a 3D-printed phantom for 3D camera calibration proved successful to reduce calibration time and ease the process while being precise. The proposed approach gives precision compatible with surgical applications (less than 1 mm). As a future work, even if it is fast, the process could be sped up by more automatization. Indeed, the manual ICP initialization by point picking could be replaced by Principal Component Analysis (PCA) or other coarse automatic registration method. Besides, the validation procedure could also be improved to make it independent from the user because manual clicking could introduce a bias in the error evaluation. Finally a quantitative comparative study with other methods should be done.

REFERENCES

- [1] W. Fan, Y. Dai, and G. Giordano, "CAOS TKA Provides Improved Functional Outcomes Compared to Conventional TKA," pp. 74–69. [Online]. Available: <https://easychair.org/publications/paper/QNfl>
- [2] Jakubovic and Raphael, "High Speed, High Density Intraoperative 3D Optical Topographical Imaging with Efficient Registration to MRI and CT for Craniospinal Surgical Navigation," *Scientific Reports*, vol. 8, no. 1, p. 14894, Dec. 2018. [Online]. Available: <http://www.nature.com/articles/s41598-018-32424-z>
- [3] S. Kahn, D. Haumann, and V. Willert, "Hand-eye calibration with a depth camera: 2D or 3D?" p. 10.
- [4] L. Yang, Q. Cao, M. Lin, H. Zhang, and Z. Ma, "Robotic hand-eye calibration with depth camera: A sphere model approach," in *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*. Auckland: IEEE, Apr. 2018, pp. 104–110. [Online]. Available: <https://ieeexplore.ieee.org/document/8384652/>

Tool motion estimation from internal pressure variation in a parallel, fluid-driven endoluminal soft robot

Lukas Lindenroth¹, Jeref Merlin¹, Sophia Bano¹, Joseph G. Manjaly², and Danail Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London

²University College London Hospitals Biomedical Research Centre, National Institute for Health Research

INTRODUCTION

Soft robots provide the inherent advantage of being able to conform with their environment, which renders them suitable for endoluminal interventions. For delicate surgical scenarios, however, this compliance can pose challenges, particularly when there is a human in the loop interacting with the robot.

In our previous work we have presented a fluid-driven endoluminal soft robot which shows the potential to aid needle guidance through the ear canal when delivering intratympanic injections [1]. Whilst the robot is able to effectively suppress small motions induced in the needle at its proximal end by the operating clinician during insertion, larger needle motions could deflect the robot and thus move it away from the desired insertion point. A robust method for estimating when needle motion occurs or is imminent is therefore required.

Although needle motion, which induces motion in the soft robot, could be estimated with the on-board camera, the rate of detection is limited by camera framerate and the subsequent detection algorithm. Moreover, needle motion has to have occurred to be detectable through vision, therefore indirect measures predicting motion before it manifests in the scene could be advantageous.

In this work we find an alternative approach by predicting needle motion through pressure variation in the driving fluid of the soft robot. We demonstrate that external pressures applied to the ballooning fluid actuators of the soft robot can be deduced and that it is possible to infer needle motion by relating the applied forces to the motion of a significant landmark point in the image.

MATERIALS AND METHODS

The soft robot used in this study is actuated by ballooning membrane actuators, which are employed to center the robot within the surrounding lumen and through differential actuation steer the robot. From our previous work, the prototype has been refined to achieve a greater range of motion by reducing the distance between the two actuation stages and increased actuator diameters. An overview of the application and the current prototype are shown in Fig. 1a) and b) respectively.

The robot is actuated by a hydraulic pump system. It is comprised of stepper motor-driven syringe pumps for which control commands are generated by a motion controller (Trinamic TCM-6214). High-level control of the system is achieved by a laptop running ROS. Pressure

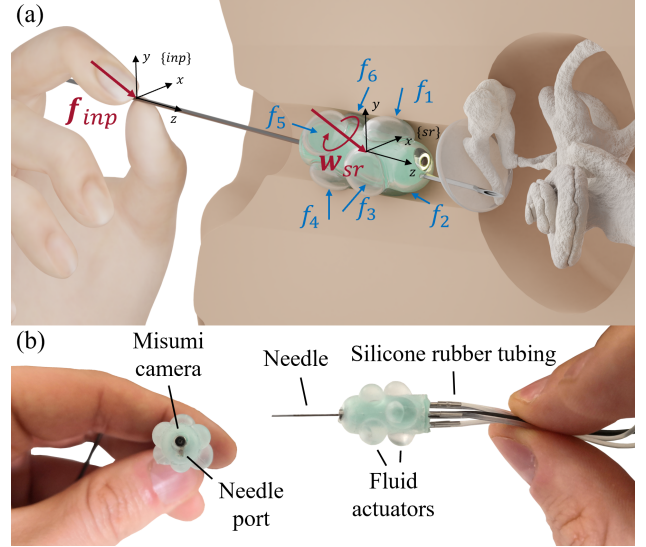


Fig. 1: Rendering (a) and prototype (b) of a fluidic-driven soft robot for needle guidance through the ear canal

sensors (MPRLS, Honeywell, United States) are added to the fluid circuit through custom 3d-printed connectors. Pressure signals from the sensors are compiled by a microcontroller (Teensy 4.0, PJRC, United States) at a rate of 100Hz. The robot is equipped with a camera (MD-V1001L-91X, Misumi Electronics Corporation, Taiwan). The target insertion point is inferred at a rate of approx. 15Hz [1].

Prior to puncturing the tympanic membrane, the distal motion of the needle is solely based on the forces and moments applied to the soft robot. As we've shown in our previous work [2], in the reference frame of the robot, the quasi-static relationship between external force and hydrostatic pressure variation can be written as

$$\mathbf{w}_{ext} = \mathbf{H} \cdot \mathbf{P}_{ext} \cdot \hat{\mathbf{a}} = \mathbf{H} \cdot (\mathbf{P}_{tot} - \mathbf{P}_{int}) \cdot \hat{\mathbf{a}} \quad (1)$$

Where $\mathbf{w}_{ext} = [\mathbf{f}_{ext}, \mathbf{t}_{ext}]^T$ is the external wrench applied to the soft robot in its frame of reference $\{sr\}$, \mathbf{P}_{ext} the pressure variation in the membrane actuators caused by the external force, \mathbf{P}_{tot} the measured pressure in the respective fluid channels and \mathbf{P}_{int} the internal pressure in the actuators caused by the introduced fluid volume. $\hat{\mathbf{a}}$ is the approximate contact area between inflated membrane actuator and surrounding lumen and \mathbf{H} the wrench transformation matrix mapping from individual actuator frames to robot frame such that

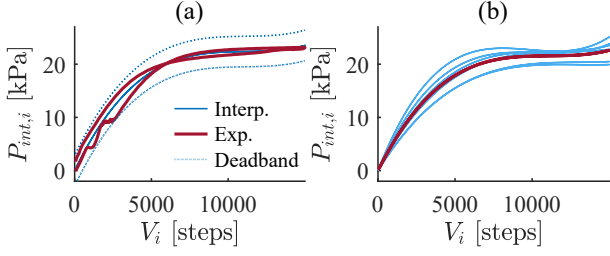


Fig. 2: Internal pressure induced by actuator inflation with the resulting interpolation (a) as well as the combined interpolation for all six actuators (b)

$$\mathbf{H} = \begin{bmatrix} 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \\ -1 & \frac{1}{2} & -\frac{1}{2} & 1 & -\frac{1}{2} & -\frac{1}{2} \\ b & -\frac{1}{2}b & -\frac{1}{2}b & b & -\frac{1}{2}b & -\frac{1}{2}b \\ 0 & \frac{\sqrt{3}}{2}b & -\frac{\sqrt{3}}{2}b & 0 & \frac{\sqrt{3}}{2}b & -\frac{\sqrt{3}}{2}b \end{bmatrix} \quad (2)$$

Where b is the distance between the stages of actuation and the center of the robot. To determine the internal pressure P_{int} , a calibration is performed by gradually inflating all actuators and measuring the pressure response. A 3rd order polynomial is used to interpolate the internal pressure as a function of the fluid volume. To account for unmodelled phenomena e.g. hysteresis, a deadband is introduced in which the external pressure variation is neglected. The deadband, which leads to a marginal offset force when a contact is applied, is defined as the maximum deviation between $P_{int,i}^{act} = f(V_i)$ and the acquired data. Estimation of external pressures under varying fluid volume is demonstrated by positioning the soft robot within a cylindrical tube and gradually inflating the actuators until contact occurs.

Motion $d\mathbf{X}_{sr}$ induced by the externally applied force \mathbf{w}_{ext} is estimated by taking into account the directional compliance \mathbf{C} of the soft robot such that

$$d\mathbf{X}_{sr} = \mathbf{C} \cdot \hat{\mathbf{a}} \cdot \mathbf{H} \cdot (\mathbf{P}_{tot} - \mathbf{P}_{int}) \quad (3)$$

This is validated by placing the soft robot in a custom ear canal phantom, pointing towards the desired needle insertion point which is inferred from the camera image as shown in [1]. The robot is manually displaced through an inserted needle and the pressure response is monitored. As rotation of the robot has a more significant impact on the change in location of the insertion point, we show the impact of only torque on the motion such that $d\mathbf{X}_{target} = \mathbf{u} \cdot d\mathbf{X}_{sr}$, where \mathbf{u} maps from local pose changes to the movement of the tracked insertion point on the tympanic membrane. The factors \mathbf{u} , \mathbf{C} and $\hat{\mathbf{a}}$ are calibrated in the given experiment using a number of sample points.

RESULTS

The calibration curve with its 3rd order interpolation as well as the resulting deadband are shown for a single actuator in (Fig. 2a)) with resulting interpolations (Fig. 2b)). The calibrated system is tested by placing the robot in the center of a surrounding lumen and inflating all six actuators simultaneously (Fig. 3). It can be seen that

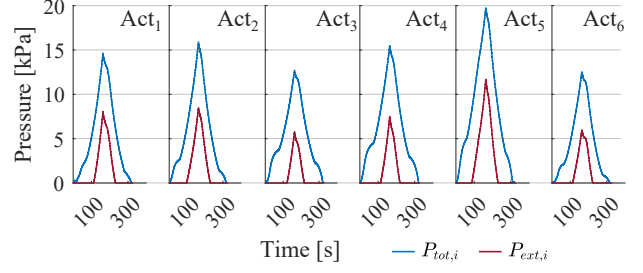


Fig. 3: Actuator inflation sequence in a cylindrical lumen

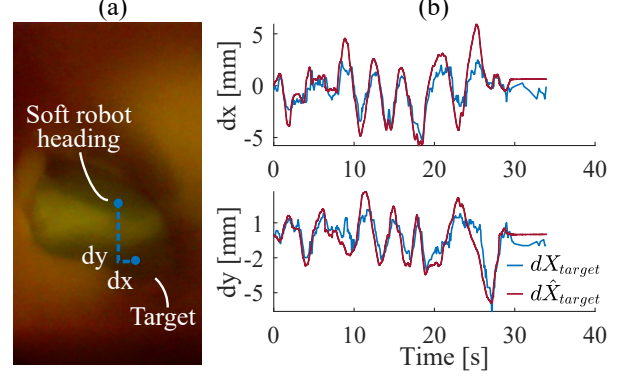


Fig. 4: Camera target view (a) and measured $d\mathbf{X}_{target}$ and estimated motion $d\hat{\mathbf{X}}_{target}$ of the target insertion point (b)

despite continuous fluid flow into and out of the actuators external pressure is only detected when an external contact is present. The motion predicted from the estimated torques (Fig. 4) tracks the measured target movement well and errors are consistent across the investigated sequence.

CONCLUSIONS AND DISCUSSION

This work has demonstrated that external contact pressure applied to fluid-driven, ballooning membrane actuators can be measured intrinsically from fluid pressure variation while varying actuator inflation volumes. We have shown that these external pressures can be mapped to forces and torques applied to the soft robot which in return can be related to the robot motion. Whilst models still have to be derived to accurately relate the applied pressures to forces under varying contact area of the ballooning actuators and to describe pose changes of the robot as a function of these forces, our preliminary results indicate that inferring motion from pressure variation is possible for such type of endoluminal soft robot. For the application of intratympanic injections, this methodology could be used as a feedback for the steadiness of the needle or as a means for active compensation of needle motion through the robot.

REFERENCES

- [1] L. Lindenroth, S. Bano, A. Stilli, J. G. Manjaly, and D. Stoyanov, "A fluidic soft robot for needle guidance and motion compensation in intratympanic steroid injections," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 871–878, 2021.
- [2] L. Lindenroth, D. Stoyanov, K. Rhode, and H. Liu, "Towards intrinsic force sensing and control in parallel soft robots," *Arxiv preprint*, vol. submit/4036231, 2021.
- [3] L. Lindenroth, C. Duriez, J. Back, K. Rhode, and H. Liu, "Intrinsic force sensing capabilities in compliant robots comprising hydraulic actuation," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 2923–2928.

Acoustic or Visual? A User Study on the Effectiveness of Sensor-based Feedback to Train Tissue Manipulation in Robotic Surgery

A. Mariani¹, D. Galeazzi², S. Tognarelli¹, A. Menciassi¹

¹*The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy*

²*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy*
andrea.mariani@santannapisa.it

INTRODUCTION

Feedback has been long recognized as a crucial factor in surgical education since it provides the trainee with the proper motivation and guidance towards proficiency [1]. Indeed, the use of error-based feedback during training has been proven to produce more accurate performance on subsequent trials [2]. In the recent years, simulation has come to the forefront for the development and refinement of surgical practical skills [3]. Among the advantages that simulation offers, there is the possibility to perform skill assessment objectively and provide feedback accordingly. Simulators can be either virtual or physical (mannequins). In the former, quantitative skill assessment is enabled by the easy extraction of an ample number of parameters and metrics from the virtual environment [4]. In the latter, the integration of sensors is used to sense metrics that are computer-based processed for skill assessment [5].

In a recent project of ours, we designed and fabricated an anatomy-based simulator of pulmonary vein [6]. This simulator was integrated with soft sensors to detect the strain applied to the vein during its manipulation. In fact, excessive strain applied to delicate tissues as blood vessels may cause dramatic intraoperative events, such as major bleeding [7]. This aspect is even more fundamental with the spreading of robotic surgery, which requires the surgeons to learn the new control dynamics of the surgical instruments [8] and, usually, how to deal with the loss of haptic feedback [9]. In [6], we preliminary tested the effectiveness of a concurrent combination of acoustic and visual feedback based on the sensed strain. This feedback was proven to minimize the strain while performing vein manipulation by robotic control of the surgical instruments. In this work, we aim at comparing the effectiveness of acoustic versus visual feedback.

MATERIALS AND METHODS

Experimental setup – The experiments were performed using the da Vinci Research Kit, a research platform built upon the components of a first-generation da Vinci Surgical System [10]. Both the surgical instruments mounted on the patient-side robotic manipulators were single fenestrated forceps. The pulmonary vein simulator was integrated with a clamping system to pre-tension the vein as in physiological conditions and it was placed on a silicon-based board. In order to provide acoustic feedback, we took advantage of a buzzer (placed close to the surgeon console). In order to provide visual feedback,

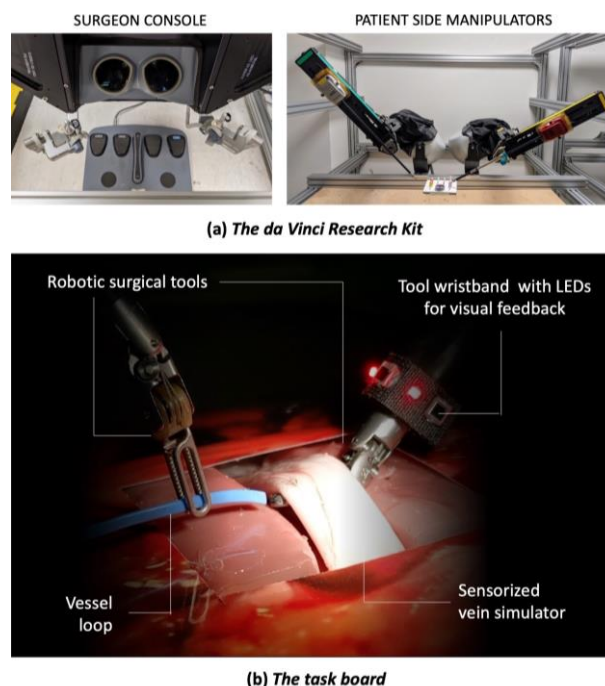


Figure 1. a) The key components of the da Vinci Research Kit as a teleoperation system for robotic surgery; b) The task board including the sensorized simulator of pulmonary vein. The figure shows the tool wristband with active visual feedback.

we used a tool wristband [11] mounted on the right instrument. Fig. 1 shows the experimental setup.

Acquisition protocol – Twenty non-medical participants (all right-handed and with no experience in robotic teleoperation) participated to the study. All the subjects gave informed consent in accordance with the declaration of Helsinki. All the participants performed two tasks: 1) *Task 1* – vein raising using the right instrument till feedback triggering and immediate dropping of the vein; 2) *Task 2* – passage of a vessel loop under the vein by using the left instrument, while raising the vein by using the right instrument and minimizing its strain by relying on the feedback. These tasks resembled the initial part of blood vessel stapling (common surgical step in several procedures). *Task 1* aimed to assess feedback effectiveness while concentrating on the feedback itself. *Task 2* (more complex than *Task 1*) aimed to assess feedback effectiveness while focusing on the task. A video showing a repetition of the tasks (that was also shown to each user before the experiment) is available at <https://bit.ly/32dWSKB>. Each participant performed 2 repetitions of each task (*Task 1* and *Task 2*): one with acoustic feedback and one with visual feedback. The

order of these 2 repetitions was randomized (to mitigate possible learning effects on the results). The feedback was binary (active/inactive): it was active (1kHz tone for acoustic; red LEDs on for visual) all the time that the vein strain was over a pre-defined threshold (i.e., $\epsilon_{\text{thre}}=0.25$). This threshold did not aim to be a realistic safety threshold (like we did in [6]), yet it was tuned to facilitate feedback activation (the effect of which we aimed to assess) in *Task 1* and *Task 2*.

Evaluation metrics and data analysis – Two objective metrics were then computed for each task repetition: 1) the time spent with active feedback (i.e., with strain over the abovementioned pre-defined threshold); 2) the total time to complete the task. Due to the small sample size, non-parametric statistical significance tests were used to compare the effect of the two kinds of feedback. The *Wilcoxon signed-rank* was used in MATLAB. The two metrics were selected as dependent variables, while the kind of feedback as independent factors. To also evaluate the subjective perception of feedback, all the participants were asked to express which kind of feedback they found most effective in *Task 1* and in *Task 2*.

RESULTS

All the participants successfully completed the four task repetitions. Fig. 2 summarizes the results. In terms of objective metrics, the median value of both the time over threshold and the total time to complete the task was lower in case of acoustic feedback with respect to visual feedback. However, this difference was statistically significant just in case of time over threshold in *Task 1* (0.72 s vs. 1.1 s, $p<0.001$). In terms of subjective preference of the users, the acoustic feedback was perceived as more effective both in *Task 1* (90% of the users' scores) and in *Task 2* (95%).

CONCLUSION AND DISCUSSION

In this study, we compared the effectiveness of a binary sensor-based acoustic versus visual feedback while performing vein manipulation by robotic control of the surgical instruments. The results highlighted a clear subjective preference of acoustic feedback by the users. Analyzing objective performance metrics, the major effectiveness of acoustic feedback was less distinct. Only when focusing on the feedback itself, the acoustic feedback was proven to significantly reduce the strain applied to the vessel. The total time to complete the task was similar, thus suggesting a similar intuitiveness of the feedbacks. In general, the lack of statistical significance could be also linked to the high variability in the data and relatively small sample size. These results can be useful in the refinement of our sensorized simulator as a training platform to optimize training on tissue manipulation in robotic surgery (in the funding framework of the 2022 Intuitive Surgical Technology Research Grants). While this user study promoted the acoustic feedback as warning signal when exceeding a certain threshold, future work could shed light on the usefulness of a continuous information on strain values (e.g., by tone variation or multiple light colors).

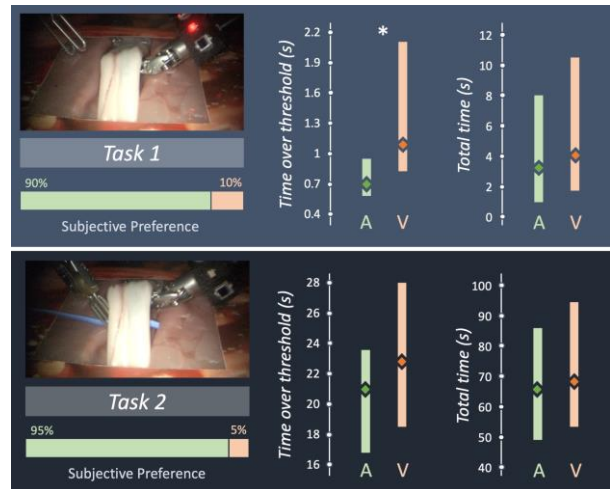


Figure 2. Results associated to *Task 1* (top) and *Task 2* (bottom). The color code refers to the kind of feedback: acoustic (A, green) and visual (V, orange). Diamonds refer to the median across users, while vertical bars stand for data variability (25th and 75th percentiles). White asterisks indicate a statistically significant difference between feedback types.

REFERENCES

- [1] Hatala, et al., "Feedback for simulation-based procedural skills training: A meta-analysis and critical narrative synthesis," *Adv. Health Sci. Educ. Theory Pract.*, vol. 19, no. 2, pp. 251–272, 2014.
- [2] Boyle, et al., "Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure." *Postgraduate medical journal*, 87(1030), pp.524-528, 2011.
- [3] Aydin, et al., "The role of simulation in urological training - A quantitative study of practice and opinions", *Surgeon*, 14: 301–7, 2016.
- [4] Connolly, et al., "Validation of a virtual reality-based robotic surgical skills curriculum," *Surg. Endoscopy*, vol. 28, no. 5, pp. 1691–1694, 2014.
- [5] Caccianiga, et al., "An evaluation of inanimate and virtual reality training for psychomotor skill development in robot-assisted minimally invasive surgery". *IEEE Transactions on Medical Robotics and Bionics*, 2(2), 118–129, 2020.
- [6] Galeazzi, et al., "A Physical Simulator Integrated with Soft Sensors for Mastering Tissue Manipulation in Robotic Surgery", (*submitted to*) 2022 IEEE 5th International Conference on Soft Robotics, 2022.
- [7] Meyerson, et al., "Needs assessment for an errors-based curriculum on thoracoscopic lobectomy", *The Annals of thoracic surgery*, 94(2), 368-373, 2012.
- [8] Nisky, et al., "Uncontrolled manifold analysis of arm joint angle variability during robotic teleoperation and freehand movement of surgeons and novices", *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2869–2881, 2014.
- [9] Okamura, "Haptic feedback in robot-assisted minimally invasive surgery", *Current Opinion in Urology*, vol. 19, no. 1, pp. 102-107, 2009.
- [10] Kazanzides, et al., "An open-source research kit for the da Vinci® Surgical System", *IEEE international conference on robotics and automation (ICRA)*, pp. 6434-6439, 2014.
- [11] Mariani, et al., "Prototyping a sensorized tool wristband for objective skill assessment and feedback during training in minimally invasive surgery", *International Symposium on Medical Robotics*, 2021.

Experimental Evaluation of the SoftSCREEN Capsule System in a Colon Phantom

Vanni Consumi¹, Lukas Lindenroth¹, Danail Stoyanov¹ and Agostino Stilli¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, UK

INTRODUCTION

Colorectal cancer (CRC) is one of the leading causes of cancer related death worldwide [1] and colonoscopy represents the gold standard procedure for its diagnosis. However, conventional colonoscopy, where the scope is maneuvered by pushing, pulling and rotating it through the anus, remains a complex procedure to be carried out even for highly skilled physicians, so that undergoing such treatment is typically cause of discomfort and anxiety for patients. Since the survival rate of CRC patients is greatly impacted by the stage at which the disease is detected, the adoption of mass-screening programs of the gastro-intestinal (GI) tract has the potential to significantly reduce the CRC death rate.

To overcome the drawbacks of standard colonoscopy, researchers are investigating robotic solutions to reduce both the complexity and the invasiveness of this procedure. Tethered robotic capsules with active navigation capabilities have attracted large interest in the research community: to reduce the discomfort experienced by the patient and increase the safety of the procedure [2] the propulsion force is applied locally rather than remotely as in the case of traditional colonoscopy. Furthermore, the GI tract presents various curves and the diameter of the lumen varies significantly throughout the organ, ranging from 25-40 mm in the anal canal to 60-70 mm in the ascending colon, reaching up to 80 mm when insufflated with CO₂. In addition, the walls of the colon are covered with a thin layer of mucus that further limits the friction force that can be achieved by a locally applied propulsion force, making navigation within the GI tract extremely challenging for small active devices.

Robotic capsules relying on track-based locomotion [3] have shown good navigation capability thanks to the large propulsion force associated with wide contact area that the tracks can establish on the colon surface. The use of micro patterned soft tracks, such as those proposed in [4], significantly increased the traction force the system can provide. To characterise and validate the navigation capabilities of robotic systems for colonoscopy, testing inside either ex vivo colon or colon phantoms [5] are typically carried out. However, controlling accurately features such wall-pressing force to enhance propulsion or the camera position for an optimal field of view, can be challenging aspects of the design of these robotic capsules.

In this work the design and the navigation capabilities of the latest integration of the SoftSCREEN system [6] are presented with testing conducted in a silicone-based colon phantom.

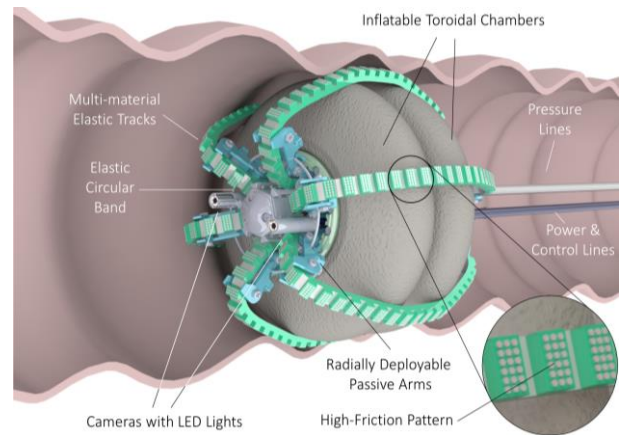


Figure 1 – CAD of the SoftScreen System navigating the colon in partially inflated state. The system is comprised of six elastic tracks, two inflatable toroidal chambers, two cameras and two elastic bands to recall the two sets of six radially deployable arms

MATERIALS AND METHODS

The current 2.5:1 scale system, prototyped at this scale to validate the navigation capability, is a track-based, tethered robot which exploits silicone materials to enable diameter adaptability, thanks to the inflation of two toroidal chambers, and locomotion, relying on multi-materials elastic tracks. The proposed design intends to be a low-cost and easy-to-control solution for colonoscopy. In Fig.1 a CAD of the SoftScreen system is depicted showing the main components, including the two front cameras for visual inspection. A rigid chassis encases a brushless motor (315170, Maxon Motor™, Sashseln, Switzerland), paired with a 256:1 ratio gearbox and a rotatory encoder, connected to a worm gear: the rotation of the latter causes the six elastic tracks to evert from the front or the rear of the capsule based on the direction of rotation of the motor while, externally, the tracks are in contact with the surrounding walls of the colon and provide the propulsion needed for navigation. The toroidal chambers, surrounding the chassis and passing through the tracks are inflated with air accordingly to the local diameter of the colon to change the overall shape of the robot. This action works as a wall-pressing mechanism to control the force normal to the walls and therefore the tangential friction force applied by the tracks. At the same time, the use of two inflatable toroidal chambers ensures a stable alignment of the central axis of the system with the lumen axis, therefore the cameras are set at an optimal angle for inspection. From the previous design [6], rotational arms

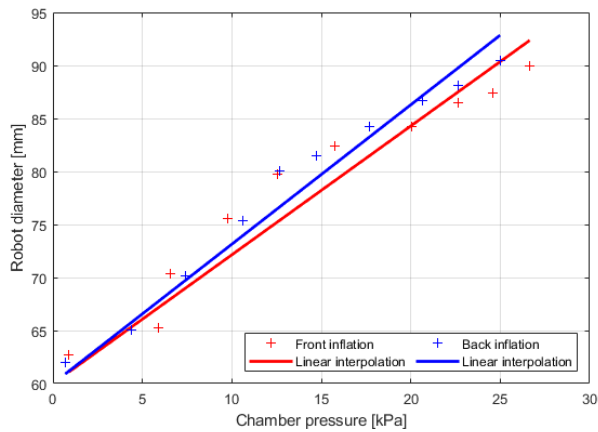


Figure 3 – System diameter change (measured at the most external point of the tracks) for front and back chambers. Deformation measured with visual markers in the case of free inflation of the chambers (no contact with external walls).

(see Fig. 1) have been also integrated in the exit and entry points of each track guide of the rigid chassis. Each set of six arms is kept closed and parallel to the internal guides in the non-inflated state by a circular elastic band but when the inflation of the chambers takes place, the arms are deployed radially. This compliant mechanism reduces the frictional loss between the tracks and chamber surfaces by reducing the contact pressure between the two that arises due to the longitudinal elastic deformation of the tracks. The elastic tracks are composed by three silicones of different Shore Hardness and all but the over moulded patterns on the top of the teeth are treated with an additive to reduce frictional losses. The base of the track is made of Dragon Skin™ 30 (Smooth-On Inc., Easton, PA, US) a 30A silicone capable of withstanding the large deformations caused by the inflation of the chambers. The teeth are made of Smooth Sil™ 60 instead, a 60A silicone that ensures reliable mechanical transmission between the worm gear and the tracks. Then, a micro-patterned layer made of Dragon Skin™ FX Pro, a 2A Shore silicone is over moulded on the top side of the teeth to increase the adhesive force that propels the system. In addition, an elastic sleeve made of spandex with low-friction surface has also been incorporated to create a smooth interface between the tracks and the chambers.

In terms of control, the user can independently control the inflation of each chamber with an external pressure regulator (VPPX, FESTO™, Esslingen am Neckar, Germany). The pressure feedback signals of the chambers are then processed via a 16-bit ADC (Adafruit™, New York, US) and displayed using Arduino UNO board. The inflation of the two chambers, fixed to flanges on the chassis 54 mm distant one each other, has been validated to vary the diameter of the robot from the non-inflated status of 60 up to 90 mm, as shown in Fig. 2, but deformation up to 120 mm can be achieved without failure. The SOFTScreen system navigating a silicone-based colon phantom of 80 mm of diameter is presented in Fig. 3. The scaled system has demonstrated good shape shifting capabilities within the phantom used, while being able of navigating the conduct.

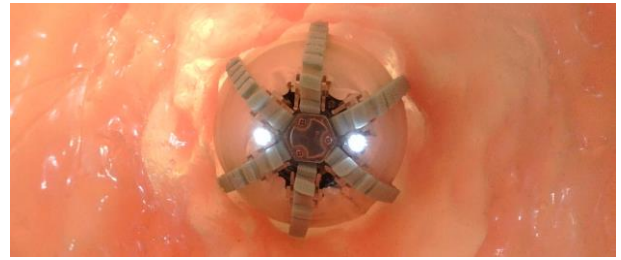


Figure 2 – The SOFTScreen capsule system inflating and navigating inside the lumen of the silicone colon phantom.

RESULTS

The silicone phantom is used to evaluate the navigation capabilities of the system in a realistic scenario. The SOFTScreen system navigated the phantom lumen adjusting its diameter to reach the walls as shown in Fig. 3. During the navigation test the phantom was held at the extremities by two circular supports to prevent the collapse of the silicone membrane and to simulate colon insufflation. During the navigation, inspection of the lumen was successfully conducted via the two front cameras which remained at an optimal angle during the navigation offering 360 degrees of the lumen as expected. The system was able to reach a navigation speed of 2.5 mm/s in the phantom, even though a peak speed of 5 mm/s has been reached in rigid pipes.

CONCLUSION AND DISCUSSION

The navigation of the colon represents a challenge for standard colonoscopic systems. A track-based propulsion mechanism paired with a wall-pressing actuator can overcome the limitations of traditional scopes and offer better solutions for navigating inside the colon. Not only has the SOFTScreen system with its active propulsion system and high compliance demonstrated to be able to adapt and navigate lumens of different sizes and surfaces, but its inherently compliance allows for local deformations of the profile to overcome obstacles and irregular surfaces which commonly occurs in colonoscopy. The validation of the presented larger scale robot has paved the way for the development and pre-clinical testing of the 1:1 scale system.

REFERENCES

- [1] J. Ferlay *et al.*, “GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013.” 2015.
- [2] L. Manfredi, “Endorobots for Colonoscopy: Design Challenges and Available Technologies,” *Front. Robot. AI*, vol. 8, no. July, pp. 1–10, 2021.
- [3] D. Kim, D. Lee, B. Kim, and B.-I. Lee, “A self-propelled robotic colonoscope using elastic caterpillars,” in *IEEE ISR 2013*, 2013, pp. 1–4.
- [4] G. A. Formosa, J. M. Prendergast, S. A. Edmundowicz, and M. E. Rentschler, “Novel Optimization-Based Design and Surgical Evaluation of a Treaded Robotic Capsule Colonoscope,” *IEEE Trans. Robot.*, 2019.
- [5] S. A. Coleman, M. Pakleppa, and A. Cuschieri, “Hydro-jet propelled colonoscopy: proof of concept in a phantom colon,” *Surg. Endosc.*, vol. 35, no. 2, pp. 989–995, 2021.
- [6] V. Consumi, L. Lindenroth, D. Stoyanov, and A. Stilli, “SoftSCREEN–Soft Shape-shifting Capsule Robot for Endoscopy based on Eversion Navigation,” in *Joint Workshop on New Technologies for Computer/Robot Assisted Surgery (CRAS 2020)*, 2020, vol. 1, pp. 62–63.

Unconstrained Rotation for Control of Concentric Tube Robots with Deep Reinforcement Learning

Keshav Iyengar¹, Sarah Spurgeon², and Danail Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Science, Department of Computer Science, University College London

²Department of Electronic and Electrical Engineering, University College London

INTRODUCTION

Concentric tube robots (CTRs) are a class of continuum robot that depend on the interactions between neighbouring, concentrically aligned tubes to produce the curvilinear shapes of the robot backbone [1]. The main application of these unique robots is that of minimally invasive surgery (MIS), where most of the developments for CTRs have been focused. Due to the confined workspaces and resulting extended learning times for surgeons in MIS, dexterous, compliant continuum robots such as CTRs have been under development in preference to the mechanically rigid and limited degrees-of-freedom (DOF) robots used in interventional medicine today. The precurved tubes in CTRs, which are sometimes referred to as active cannulas or catheters, are manufactured from super-elastic materials like Nickel-Titanium alloys with each tube nested concentrically. From the base, the individual tubes can be actuated through extension and rotation, which results in the bending and twisting of the backbone as well as access to the surgical site through the channel and robot tip. Clinically, CTRs are motivated for use in brain, cardiac, gastric surgery as well other procedures [2].

Due to tube interactions, modelling and control is challenging. Position control for CTRs has relied on model development, and although a balance between computation and accuracy has been reached in the literature [1], there remain issues such as performance in the presence of tube parameter discrepancies and the impact of unmod-

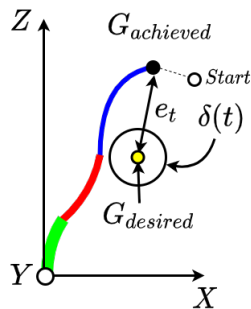


Fig. 1: State with start position and achieved goal, $G_{achieved}$ (black), desired goal, $G_{desired}$ (yellow), goal tolerance, $\delta(t)$. Outer tube (green), middle tube (red) and inner tube (blue).

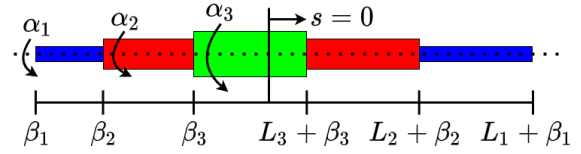


Fig. 2: Joint variables β and α of a 3 tube CTR. s is the arc-length or axis along the backbone.

elled physical phenomena such as friction and permanent plastic deformation. This motivates the development of an end-to-end model-free control framework for CTRs. One such model-free framework for control that is gaining popularity is reinforcement learning (RL), a paradigm of machine learning that necessitates an agent to output action that interacts with an environment. The environment then processes this action, and returns a new state and, depending on the task, a reward signal. The task we give the agent then is to control the end-effector Cartesian robot tip position by means of actions that represent changes in joint values. In Fig. 1 the components of the state are shown in relation to a illustrated CTR and further described in the next section.

In this work, we investigate how the rotational actuation affects final errors during evaluation of the learned policy. We find by avoiding constraining the rotational DOF of each tube, the agent can freely rotate to achieve goals as opposed to when constrained that result in more steps and larger error metrics.

MATERIALS AND METHODS

First, the Markov Decision Process (MDP), a definition required for RL algorithms is defined as follows.

- State (s_t) : States are defined as the concatenation of the trigonometric joint representation, Cartesian goal error and current goal tolerance. As shown in Fig. 2, rotation and extension of tube i (ordered innermost to outermost) are α_i and β_i . The trigonometric representation [3] of tube i is defined as:

$$\gamma_i = \{\gamma_{1,i}, \gamma_{2,i}, \gamma_{3,i}\} = \{\cos(\alpha_i), \sin(\alpha_i), \beta_i\} \quad (1)$$

In constrained rotation, α_i for each tube is constrained to be between -180° and $+180^\circ$ during each episode

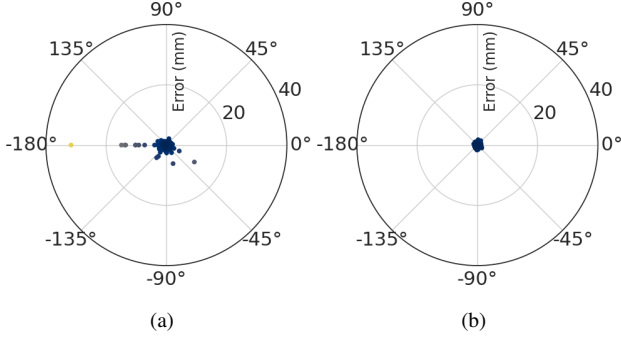


Fig. 3: Achieved goal errors for constrained (a) and free rotation (b) agents. Errors are shown in a polar plot with α_1 rotation values and associated final errors.

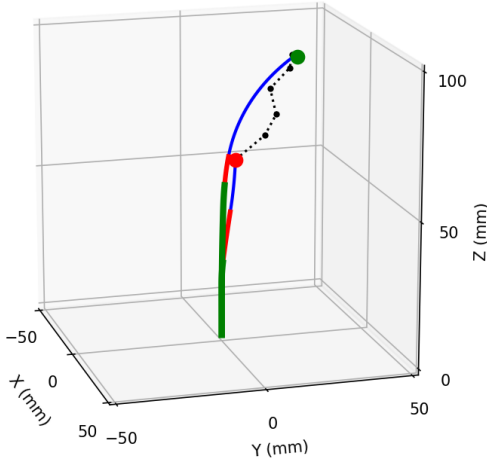


Fig. 4: Reaching a desired goal of (20, 20, 100) mm with a tip error of 0.98 mm. Starting position (red), desired goal (green).

for generation of new goals. However, these constraints are non-essential in the trigonometric representation during episode steps. The extension joint β_i can be retrieved directly and has constraints

$$0 \geq \beta_3 \geq \beta_2 \geq \beta_1 \quad (2)$$

$$0 \leq L_3 + \beta_3 \leq L_2 + \beta_2 \leq L_1 + \beta_1 \quad (3)$$

from the actuation side. Lastly, the current goal tolerance, $\delta(t)$, is included in the state where t is the current timestep. A decay curriculum function was used for 250,000 steps out of the total 500,000 training steps. The full state, s_t , can then be defined as:

$$s_t = \{\gamma_1, \gamma_2, \gamma_3, G_{achieved} - G_{desired}, \delta(t)\} \quad (4)$$

- Action (a_t) : Actions are defined as a change in rotation and extension joint positions.

$$a_t = \{\Delta\beta_1, \Delta\beta_2, \Delta\beta_3, \Delta\alpha_1, \Delta\alpha_2, \Delta\alpha_3\} \quad (5)$$

- Goals (G) : Goals are defined as Cartesian points within the workspace of the robot. The achieved

goal, $G_{achieved}$, is determined with the forward kinematics of the geometrically exact model [4] and is recomputed at each timestep as the joint configuration changes from the agents actions. The desired goal $G_{desired}$ updates at the start of every episode where a desired goal is found by sampling valid joint configurations and applying forward kinematics.

- Rewards (r_t) : The reward is a scalar value returned by the environment as feedback for the chosen action by the agent at the current timestep. The reward function used in this work is defined as:

$$r_t = \begin{cases} 0 & \text{if } e_t \leq \delta(t) \\ -1 & \text{otherwise} \end{cases} \quad (6)$$

where e_t is the Euclidean distance $\|G_{achieved} - G_{desired}\|$ and $\delta(t)$ is the goal-based curriculum function that determines the goal tolerance at timestep t . The workspace and various state and reward elements are illustrated in Fig. 1.

We train a policy for constraint and constraint-free rotation in simulation using the base hyperparameters and tube parameters as our previous work [5]. To evaluate each policy, 1000 evaluation episodes where the desired goal where randomized. Representing error metrics as mean error \pm standard deviation, The constraint-free agent had error of $0.69 \text{ mm} \pm 0.24 \text{ mm}$ with a success rate of 97.1% while constrained agent had error of $0.94 \text{ mm} \pm 1.44 \text{ mm}$ and with a success rate of 87.7%. In our previous work [5], where the rotation was constrained, evaluation showed error metrics of $1.29 \text{ mm} \pm 0.18 \text{ mm}$ and a success rate of 90.3%. Providing a goal of (20, 20, 100) mm, the solved joint values were $[-2.36, -2.03, -0.92]$ mm for β and $[-205.6^\circ, -108.2^\circ, -271.4^\circ]$ for α with a final tip error of 0.98 mm as seen in Fig. 4.

CONCLUSIONS AND DISCUSSION

Constraining the rotational DOF of the tubes results in the trained policy with worse error metrics. Moreover, with the trigonometric representation, the rotational constraints are redundant. We aim to further analysis differences in joint sampling and testing on a hardware system.

REFERENCES

- [1] H. B. Gilbert, D. C. Rucker, and R. J. Webster III, "Concentric tube robots: The state of the art and future directions," *Robotics Research*, pp. 253–269, 2016.
- [2] H. Alfalahi, F. Renda, and C. Stefanini, "Concentric tube robots for minimally invasive surgery: Current applications and future opportunities," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 410–424, 2020.
- [3] R. Grassmann, V. Modes, and J. Burgner-Kahrs, "Learning the forward and inverse kinematics of a 6-dof concentric tube continuum robot in se(3)," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 5125–5132.
- [4] D. C. Rucker, B. A. Jones, and R. J. Webster III, "A geometrically exact model for externally loaded concentric-tube continuum robots," *IEEE Transactions on Robotics*, vol. 26, no. 5, pp. 769–780, 2010.
- [5] K. Iyengar and D. Stoyanov, "Deep reinforcement learning for concentric tube robot control with a goal-based curriculum," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1459–1465.

Evaluation of a driving mechanism for percutaneous needle insertion

Jeref Merlin^{*1}, Lukas Lindenroth^{*1}, Agostino Stilli¹, Saurabh Singh^{1,2},
Steve Bandula^{1,2}, and Danail Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London

²Interventional Oncology Service, University College London Hospital

^{*}Indicates equal contribution

INTRODUCTION

To treat and diagnose different kinds of malignancies, minimally invasive image-guided interventions possess major benefits over traditional surgical approaches. Real time medical imaging such as interventional MRI or CT, ultrasound and fluoroscopy, in combination with medical image computing, aid in the execution of these procedures. MRI is typically preferred in this application as it offers high soft tissue contrast, multi-planar imaging capabilities and without exposing patient or clinician to ionizing radiation [1]. These interventional approaches are executed percutaneously, which entails the insertion of one or more needle-shaped probes or devices towards a target lesion through the skin. Thus, applications such as tumor biopsy or ablation can be performed across various anatomical regions such as liver, kidneys and breast. However, extensive planning and preparation are required as direct visual and tactile information are limited, with complications such as needle bending and tissue deformations [2]. The procedures are commonly executed by trained clinicians through iterative processes of imaging and manual adjustment of the needle. This is proved to be challenging as rigorous needle positioning is performed in a confined bore of an MRI and CT scanner iteratively, which is labor intensive and is often unergonomic for the operating clinicians. Robotic technologies have been investigated to achieve precise manipulation of probes and needle, while minimizing strain on clinicians. Fluid-actuated systems provide major benefits since pumps and motors can be placed away from the patient, thus minimizing the system's footprint. In the case of MRI, this eases the efforts for MRI compatibility since non-compatible components can be placed outside the magnetic field of the MRI machine. In this paper, we introduce a proof-of-concept based on the novel design of a fluid-driven needle insertion mechanism with inflatable hyperelastic membranes. We investigate its performance and introduce the use of rigid sliders incorporated to the membrane to improve consistency in the inflation, the consequent passive mechanical behaviour of the system and the maximum needle driving force.

MATERIALS AND METHODS

The presented design builds on the system proposed in our previous work [3]. It is composed of two actuation stages which are circular, inflatable membrane actuators as

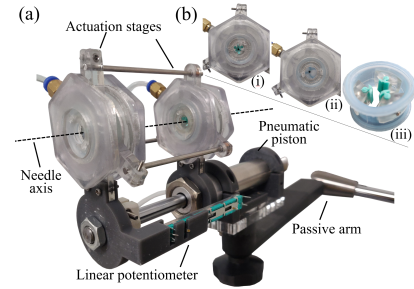


Fig. 1: Pneumatically-actuated needle driving system overview - Perspective view of the prototype (a). Front view of an inflated stage with sliders (b(i)) and without sliders (b(ii)). Sliders attachment to the membrane (b(iii)).

shown in Fig. 1a). The stages presented in Figs. 1b(i-ii) contain a central channel each, through which the needle can be fed and grasped by means of inflatable hyperelastic membranes, as shown in Fig. 1b(iii). Inflation of these membranes provides contact with the needle, resulting in the grasping force being transmitted onto the needle to facilitate insertion. However, utilising membrane alone to control a needle is inefficient as the buckling effect of the membrane when inflated radially inwards lead to inhomogeneous and inconsistent contact surface. Therefore, rigid slider elements are incorporated into the design. These sliders are axially clamped between ball bearings within the central housings to ensure free translation of the sliders in the actuation stage.

The central housings are made of photopolymer resin, printed with a 3D-printer (Formlabs©, Somerville, MA, US). Three circumferentially-positioned bushings are placed around each central housing to enable axial sliding over 3mm guiding rods. The housings are attached to a pneumatic piston to provide the axial motion of the system. These housings also support a potentiometer for closed loop position control of the piston. The piston is secured via a chassis to a passive articulated arm (Northern Digital Inc., Waterloo, Canada) to support and facilitate the global positioning of the system. The stages and piston are each actuated pneumatically via pressure regulators (VPPX-6F-L-1-F-0L10H-S1, FESTO GmbH, Esslingen, Germany), where the pressure is controlled through a microcontroller (F446RE, ST Microelectronics, Plan-les-Ouates, Switzerland) connected to a DAC (DA4C010BI, APTINEX Ltd., Maharagama, Sri Lanka).

To evaluate the efficacy of the employed sliding ele-

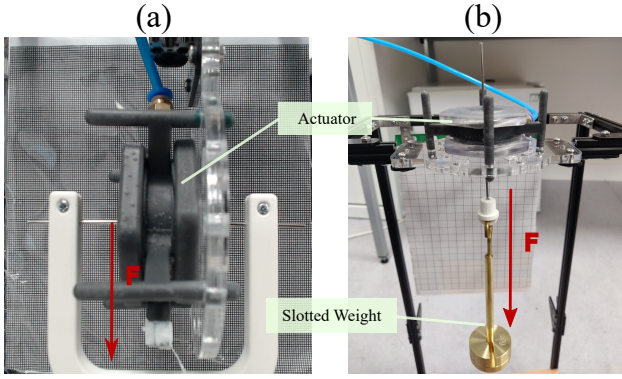


Fig. 2: Camera view of the experimental setups for compliance (a) and grasping force (b).

ment, experimental setups were constructed to compare the compliance and insertion force between the systems with and without sliding elements. Dragon Skin™ 20 silicone rubber (Smooth-On Inc., Easton, PA, US) is used for the inflatable membrane, which is fabricated with a custom injection mould, resulting in a membrane with a thickness of 1mm as shown in Fig. 1b(iii). Each slider is equipped with a small tip fabricated from Smooth-Sil 960™ (Smooth-On Inc., Easton, PA, US), which is screwed through the membranes. The systems are then actuated with the aim to grasp an 18G needle. In Fig. 2a, the compliance characteristic is investigated by inducing transversal force through slotted weights onto a stainless-steel rod with a diameter of 1.5 mm to a single stage and measuring the respective displacement of the rod at 2.5 kPa. The experiment was performed with a stainless steel rod to eliminate possible deflection. The rod displacement is measured through a camera (C922, Logitech, Switzerland) by manually tracking the end-points of the rod. This is then repeated across all angles at 30° increments to determine possible directional variability in compliance. In Fig. 2b, the grasping force between the systems was investigated through a similar setup by increasing slotted weight in increments along the direction of the rod at different pressures. The grasping force was then determined by the corresponding weight when the rod is displaced.

RESULTS

The displacement as a function of the applied load is presented in Fig. 3a. For the No-sliders profile, highly discontinuous trends are noticeable which was observed to be caused by the rod, slipping through the gaps generated by the buckling effect of the inflated balloon. This led to high discrepancies for forces of 2.5 N and significantly larger variability in the force response for the profile without slider. Linear interpolation is performed for each angle on each profile, to determine the lumped compliance of the system, as shown in Fig. 3b. The variability of compliance for the No-sliders profile occurred with a range of 0.6 mm/N while the sliders profile has a three times smaller range of 0.2 mm/N. Fig. 3c illustrates the grasping forces of each profile w.r.t the given inflation pressure up to 3.0 kPa, as pressure beyond this cause the membrane to fail. The Sliders profile could achieve

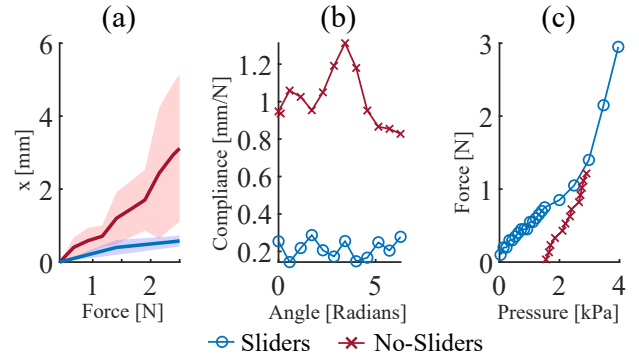


Fig. 3: The mean and standard deviation in shaded area error plot for rod displacements $[X]$ Vs. applied force $[F]$ (a), resulting interpolated compliance across varying angle of force (b) and grasping force of the stage for Sliders and Non-sliders profiles under load applied by slotted mass for different pressures (c).

approximately three times the amount of force for the same applied pressure in comparison to the No-sliders profile.

CONCLUSIONS AND DISCUSSION

In this work, a proof-of-concept of a fluid-actuated system with a soft membrane for percutaneous needle insertion is presented with the goal of high precision control and straightforward translation to being MRI-compatible. Rigid sliders are introduced to tackle limitations associated with utilizing membranes alone. The compliance test suggests that integration of sliders to the membrane could provide a more consistent compliance across all directions in the transversal plane. This passive deformation of the system is crucial as to avoid harming healthy tissues and the needle by conforming with the trajectory of the needle, while yielding the appropriate stiffness for maneuvering the needle. Furthermore, a test to evaluate the grasping force suggests that the sliders profile requires around three times the amount of applied pressure to achieve the same force. The system without sliders could not reach the required force of 2.3 N required for percutaneous needle application during liver biopsy found in [2]. The counterpart however could reach a maximum force of 2.5 N at a pressure around 1.0 kPa and up to 4 N at 3.0 kPa. Further compliance and grasping force experiments will be investigated for different types of silicone materials to achieve the appropriate compliance and insertion force.

REFERENCES

- [1] A. H. Mahnken, A. Buecker, E. Spuentrup, G. A. Krombach, D. Henzler, R. W. Günther, and J. Tacke, "Mr-guided radiofrequency ablation of hepatic malignancies at 1.5 t: Initial results," *Journal of Magnetic Resonance Imaging*, vol. 19, pp. 342–348, 3 2004.
- [2] C. Simone and A. M. Okamura, "Modeling of needle insertion forces for robot-assisted percutaneous therapy," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2, pp. 2085–2091, 2002.
- [3] L. Linderoth, A. Stilli, S. Singh, S. Bandula, and D. Stoyanov, "A fluid-driven mechanism for percutaneous needle insertion and fine manipulation under image guidance," *Joint Workshop on New Technologies for Computer/Robot Assisted Surgery (CRAS)*, 2020.

A Linguistic Comparison Between Textual Datasets to Assess the Complexity of Surgical Robotic Procedural Descriptions

Marco Bombieri¹, Diego Dall’Alba¹, and Paolo Fiorini¹

¹Department of Computer Science, University of Verona, Italy
Contact: marco.bombieri_01@univr.it

INTRODUCTION

Natural language processing (NLP) has rapidly expanded over the last 20 years thanks to increasing computational resources and high investment from both many research groups and private companies, with the aim of automating various tasks previously performed manually [1]. A notable rise has occurred recently also in the medical domain, with the introduction of deep learning technologies and pre-trained language models, such as BERT [2]. Despite this rapid growth, only few studies have attempted to apply deep learning NLP methods to the surgical domain, which is a sub-category of the medical one. The majority of the studies in surgery deals with the extraction of outcomes from textual records and with other text classification applications [3].

Robotic devices, also known as Surgical Robotic Systems (SRSs), are increasingly pervasive in operating theaters. Together with the introduction of surgical robots, numerous books, manuals, academic papers and online-resources about robot-assisted surgery have been published. They are used by medical trainees to acquire or refine their knowledge.

Extracting structured workflows from surgical textual resources would be beneficial both for the development of autonomous SRSs and for the intra-operative assistance to the surgeon. However, the *manual* extraction of these workflows requires substantial human effort and expertise. NLP mining technologies for robot-assisted surgical texts will open new frontiers of research. In particular, they would enable:

- the extraction of high-level information from surgical text-books (e.g. phases, steps or actions). This would improve computer-assisted planning of surgical procedures.
- the development of cognitive SRSs, that recognize surgical entities (e.g. anatomical parts, tools, actions) in the texts and link them with existing a-priori knowledge represented in ontologies. They would allow to infer knowledge not explicitly stated in the texts improving situation awareness algorithms and robotic assistance;
- the use of a-priori knowledge by the SRS. This will be an essential step toward cognitive surgical robotics [4].

From surgical textbooks, we can extract two types of knowledge:

- *Declarative knowledge*, usually formalized into ontologies. In surgery, declarative knowledge describes for example the characteristics of anatomical structures (e.g., the tissue color or the organ size), the technical specifications of the robotic tools and exceptional events that can happen both intra- or post-operatively.
- *Procedural Knowledge*, possessed by an intelligent agent that is able to perform a task. In surgery, the intelligent agent is a surgeon or an autonomous SRS, and procedural knowledge describes how surgical interventions should be performed.

A limited number of studies (in non-medical domains) have addressed the problem of *Procedural Knowledge* extraction, from annotated instructional texts of recipes or maintenance/repair manuals [5]. The number of published works is even lower in the biomedical field, and only one preliminary work has been published in the robotic surgery domain [6]. This is due to the lack of annotated *publicly* available data in the procedural surgical literature that could be used to develop supervised deep-learning methods. Furthermore, annotating datasets for NLP is a very expensive task that requires expert personnel. A possible solution could be to use *procedural* datasets belonging to other domains. This work therefore wants to investigate the linguistic differences between robot-assisted surgery books and other domains, such as cooking recipes or repair manuals. This is an essential information for enabling the direct application of knowledge extraction algorithms designed and developed for other domains to surgical robotic.

I. MATERIALS AND METHODS

In this work, we propose the first linguistic comparison between the procedural language used in robotic-surgery textbooks and in other domains. We have compared three representative datasets, SPKS dataset¹ and the two procedural datasets used by [5]. The first consists of 20 descriptions of robot-assisted surgical procedures belonging to the urological, gynecological, gastrointestinal and thoracic domains [6]. [5] instead deals with 30 cooking recipes from the BBC recipe website² (referred as "BBC" later on) and 30 "How-Tos" from eHow.com website referred as ("E-HOW" later on).

¹<https://gitlab.com/altairLab/spks-dataset>

²<https://www.bbc.co.uk/food/recipes>

TABLE I: Values of the metrics considered in this work for the 3 analyzed datasets.

	SPKS	BBC	E-HOW
α	13.8	5.6	6.4
β	3.64	2.0	2.35
γ	1.74	1.75	1.79
δ	> 30%	<5%	<5%

The comparison is based on the following metrics:

- (α): Flesh-Kincaid’s readability score³:

$$FKGL = 0.39(ASL) + 11.8(ASW - 15.59)$$

where *ASL* (Average Sentence Length) is the number of words divided by the number of sentences, *ASW* (Average Number of Syllables per Word) is the number of syllables divided by the number of words. The higher the score, the harder the text.

- (β) average number of verbs per sentence identified with POS-Tagging and lemmatization techniques [7].
- (γ) number of average semantic roles per verb using state-of-the-art transformer-based semantic role detection algorithms [2].
- (δ) Percentage of sentences using domain verbs not found in [8]. The coverage is evaluated in terms of presence of the lemma and of an appropriate frame in the resource.

The first metric is designed to measure the difficulty of an English sentence for an human reader. The second metric instead investigates the syntactic and structural aspect, analyzing how many tokens with the role of verb (i.e. possible actions) are present in the sentence. It gives clues on the procedural complexity of a sentence. The third metric investigates how complex it could be for an information extraction algorithm to understand and disambiguate the sentence and link the lexical tokens semantically. The fourth metric measures how much state-of-the-art semantic propositional banks, built manually using statistical analysis of general-English texts, cover the lemmas and frames often used in procedural languages.

RESULTS AND DISCUSSIONS

Table I shows metrics described in the previous section for the 3 datasets considered.

The SPKS dataset has a much higher FKGL score than the other two. A value of 13.8 means *difficult to read: college graduate expertise*. The other two datasets have scores below 7, which correspond to *fairly easy to read: primary school expertise*. The implication is that a domain expert (doctor or linguist) is needed to annotate the surgical dataset for supervised algorithms. SPKS also has a higher number of verbs per sentence (1.83 times more than BBC and 1.54 times more than E-HOW). This indicates that the procedural sentences written in the surgical manuals are significantly more complex than those of the other domains. All 3 datasets have an almost

³This formula has obvious limitations because considers only structural features: for example, the pangram "Cwm fjord-bank glyphs vext quiz" is considered *very easy to read*. It is, however, representative, since it captures useful information.

equal number of semantic roles per verb. This is expected, since the languages are procedural and the number of elements involved (e.g action, agent, target or tool) is almost the same, regardless of the domain. Finally, the surgical domain makes use of a larger number of domain verbs (for instance *kocherize*, *extraperitonealize*, *grasp*) not covered by semantic textual resources. This suggests that current semantic banks developed for general-English are unable to fully represent the semantic complexity of procedural surgical language and should therefore be extended.

CONCLUSIONS

The linguistic comparison presented in this work confirms that the descriptions of robotic-assisted surgical procedures are much more demanding to interpret: (i) they use more verbs per sentence; (ii) they require expert personnel to annotate dataset; (iii) they make use of terms that are not covered by existing semantic banks.

This preliminary investigation suggests that, in order to obtain suitable performance in mining textual descriptions for surgical robotics, modifications to existing semantic banks and the development of specialized processing methods are required.

ACKNOWLEDGEMENTS

This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 742671 “ARS”).

REFERENCES

- [1] J. Wang, H. Deng, B. Liu, A. Hu, J. Liang, L. Fan, X. Zheng, T. Wang, and J. Lei, “Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: Bibliometric study on pubmed,” *J Med Internet Res*, 2020.
- [2] P. Shi and J. Lin, “Simple BERT models for relation extraction and semantic role labeling,” *CoRR*, vol. abs/1904.05255, 2019.
- [3] J. Mellia, M. Basta, Y. Toyoda, S. Othman, O. Elfanagely, M. Morris, L. Torre-Healy, L. Ungar, and J. Fischer, “Natural language processing in surgery: A systematic review and meta-analysis,” *Ann Surg*, 2021.
- [4] M.-C. Fiazza and P. Fiorini, “Design for interpretability: Meeting the certification challenge for surgical robots,” in *2021 IEEE Int. Conf. on ISR*, 2021.
- [5] Z. Zhang, P. Webster, V. S. Uren, A. Varga, and F. Ciravegna, “Automatically extracting procedural knowledge from instructional texts using natural language processing,” in *Proceedings of LREC 2012, Istanbul, Turkey*, 2012.
- [6] M. Bombieri, M. Rospocher, D. Dall’Alba, and P. Fiorini, “Automatic detection of procedural knowledge in robotic-assisted surgical texts,” *Int. J. CARS*, vol. 16, 04 2021.
- [7] D. Jurafsky and J. H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, 2nd Edition, ser. Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education International, 2009. [Online]. Available: <https://www.worldcat.org/oclc/315913020>
- [8] M. Palmer, P. R. Kingsbury, and D. Gildea, “The proposition bank: An annotated corpus of semantic roles,” *Comput. Linguistics*, 2005.
- [9] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 NAACL-HLT Conference, Minneapolis, MN, USA*, 2019.

Predicting Image Relationships with Image Box Embeddings During Endoscopy

Anita Rau, Lourdes Agapito, and Danail Stoyanov

University College London
a.rau@cs.ucl.ac.uk

INTRODUCTION

Mapping 3D structures during endoscopy has the potential to provide guidance, control, and training for surgical interventions. For instance, a map of the colon can help improve cancer screening during colonoscopy. As endoscopic scenes lack features, depth and pose networks can be an alternative to feature-based mapping methods. But especially predicting relative camera poses between two images remains challenging because regressing novel viewpoints from training images is a complex geometric problem. In this work, we show that image box embeddings provide valuable information about the camera motion during endoscopy. Our trained model can distinguish between forward and backward movement of the camera on synthetic data — a difficult task in real colonoscopy. We demonstrate that the same model can precisely relocate after occlusion and generalize to laparoscopy sequences for some tasks without fine-tuning.

MATERIALS AND METHODS

Symmetric measures in the world, like the similarity between two scenes, can be learned with symmetric measures in embedding space. If image **a** is similar to image **b**, then image **b** must be similar to image **a** as well. In embedding space, two similar images would have a small Euclidean distance. But what if we want to learn an asymmetric measure?

Box embeddings were initially developed for natural language processing [1] but were recently applied to images as well [2]. Image box embeddings allow us to describe asymmetric measures in embedding space and can be used to learn asymmetric relationships between images. One such relationship is the normalized surface overlap (NSO) between two images. The $NSO \in [0, 1]$ describes what ratio of the surface in image **x** is also visible in image **y** and vice-versa:

$$NSO(\mathbf{x} \mapsto \mathbf{y}) = \text{overlap}(\mathbf{x} \mapsto \mathbf{y}) / N_{\mathbf{x}}, \text{ and} \quad (1)$$

$$NSO(\mathbf{y} \mapsto \mathbf{x}) = \text{overlap}(\mathbf{y} \mapsto \mathbf{x}) / N_{\mathbf{y}}, \quad (2)$$

where $\text{overlap}(\mathbf{x} \mapsto \mathbf{y})$ is the number of pixels in **x** also visible in **y** and N is the total number of pixels.

To distinguish between the two directions, $NSO(\mathbf{x} \mapsto \mathbf{y})$ is referred to as *enclosure*, and $NSO(\mathbf{y} \mapsto \mathbf{x})$ is called *concentration*. Depending on the enclosure and concentration four different interpretable image relationships can be observed in Figure 1. These relationships arise from the asymmetry of the NSO and cannot be learned with symmetric vector embeddings.

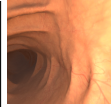
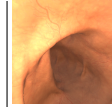
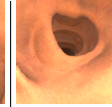
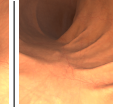
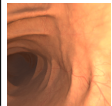
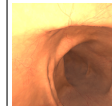
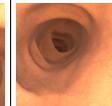
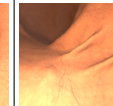
Image x				
Image y				
Enclosure	73 (79)	63 (58)	36 (40)	20 (35)
Concentration	72 (75)	25 (29)	9 (17)	62 (57)
Relationship	clone-like	zoom out	oblique	zoom in

Fig. 1: Ground truth (predicted) enclosure and concentration in %. The four examples show different relationships between **x** and **y**. If the enclosure is high but the concentration low, **y** must be a zoom out of **x**. If both NSOs are low, **x** and **y** must be observed from oblique views.

L_1	$L_1 < 0.1$	$L_1 < 0.05$	$L_1 < 0.01$
0.0698	79%	54%	12 %

TABLE I: Surface overlap errors on colonoscopy data.

To learn NSO we train a neural network to embed an image **x** as 32-dimensional box $\mathbf{b}_{\mathbf{x}}$. The NSO can then be approximated with the asymmetric normalized box overlap NBO, computed as the intersection of $\mathbf{b}_{\mathbf{x}}$ and $\mathbf{b}_{\mathbf{y}}$ divided by the volume of $\mathbf{b}_{\mathbf{x}}$ or $\mathbf{b}_{\mathbf{y}}$, respectively. The box embedding is learned using the loss

$$\mathcal{L}_{\text{box}} = \|\text{NSO}(\mathbf{x} \mapsto \mathbf{y}) - \text{NBO}(\mathbf{b}_{\mathbf{x}} \mapsto \mathbf{b}_{\mathbf{y}})\|_2^2. \quad (3)$$

We train the network proposed in [2] on synthetic data generated according to the pipeline in [3]. Our data consists of about 15k random image pairs with ground truth overlap from four traversals through the virtual colon. In each trajectory, each camera pose is randomly offset by up to 2mm and rotated by up to 20°. Our test data consists of 100 random image pairs from a different trajectory. We also evaluate the model on sequences from the Hamlyn laparoscopy dataset [4]. This datasets depicts real images recorded with the *da Vinci* surgical system.

RESULTS

We conduct and report four experiments. First, we show that box embeddings can learn the NSO between two images. We train the model proposed in [2] for 20 epochs on our training set and show example predictions for the NSO in gray in Figure 1. For each image pair we predict enclosure and concentration and report the average error in Table I. On average the NBO and NSO differ by roughly

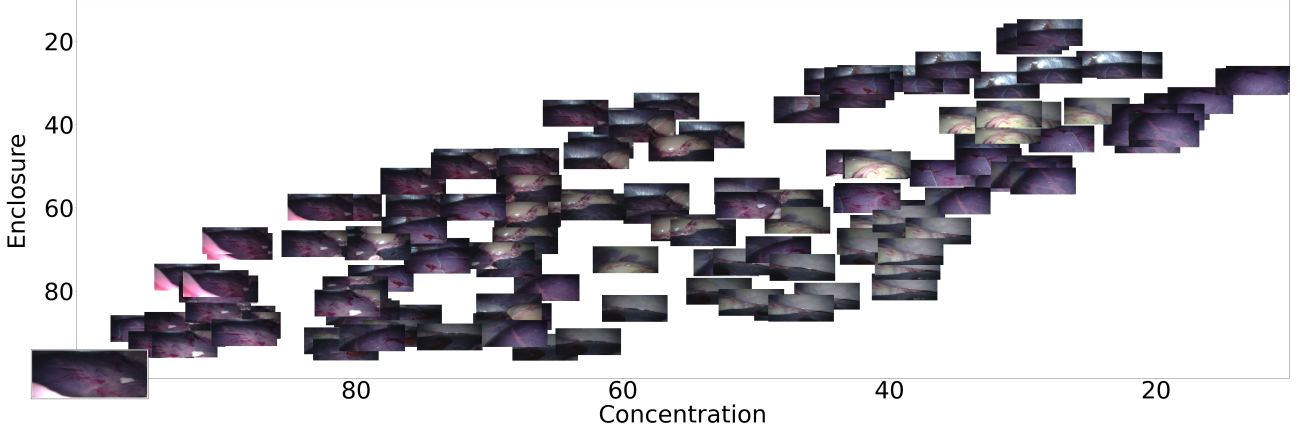


Fig. 2: For a query image (bottom left) we plot database images at the position of their predicted enclosure and concentration. Images in the lower left corner have high enclosure and concentration and are similar to the query. As the network was trained on Notre Dame it has no knowledge of the geometric layout of the laparoscopic scene and misinterprets the relative position of landmarks. Yet, the model is able to cluster images that depict the same region of the abdomen. Note how images between clusters sometimes depict landmarks from each one of the adjacent clusters.



Fig. 3: Predicted NNs (bottom) for three query images (top). Visually the retrieved image can barely be distinguished from the query—the model is able to retrieve relevant images for a camera relocation pipeline.

seven percentage points. We also report the accuracy as percentage of images pairs with an L_1 error

$$L_1(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\text{NSO}(\mathbf{x} \mapsto \mathbf{y}) - \text{NBO}(\mathbf{b}_\mathbf{x} \mapsto \mathbf{b}_\mathbf{y})\|_1 + \frac{1}{2} \|\text{NSO}(\mathbf{y} \mapsto \mathbf{x}) - \text{NBO}(\mathbf{b}_\mathbf{y} \mapsto \mathbf{b}_\mathbf{x})\|_1. \quad (4)$$

of less than a threshold in Table I. We conclude that the model predicts surface overlap accurately, despite the colon's repetitive structure.

Secondly, we show that box embeddings predict forward and backward movements accurately. Zooming into an image implies that the camera moved forward. Thus, high concentration and low enclosure indicate a forward movement—and vice versa. Of 100 test image pairs, the model predicts the correct direction in **90%** of the cases.

Thirdly, we show that box embeddings can relocate a camera after the view was temporarily occluded, for instance by water. NSO becomes symmetric with

$$\text{NSO}^{\text{sym}}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (\text{NSO}(\mathbf{x} \mapsto \mathbf{y}) + \text{NSO}(\mathbf{y} \mapsto \mathbf{x})). \quad (5)$$

Accordingly, we can approximate NSO^{sym} as $\frac{1}{2} \text{NBO}(\mathbf{b}_\mathbf{y} \mapsto \mathbf{b}_\mathbf{x}) + \frac{1}{2} \text{NBO}(\mathbf{b}_\mathbf{x} \mapsto \mathbf{b}_\mathbf{y})$. For each image in the test trajectory we retrieve the nearest neighbor (NN) as the image with the highest symmetric box overlap. We report the Euclidean distance between query images and their NNs in Table II.

L_2	$L_2 < 5\text{cm}$	$L_2 < 1\text{cm}$	$L_2 < 0.5\text{cm}$	$L_2 < 0.1$
0.43 cm	94%	78%	53%	5%

TABLE II: Median L_2 distance between query images and their predicted NNs. We also report the accuracy as the percentage of L_2 errors that are smaller than a threshold. For reference, the total trajectory length is 105.1 cm.

Lastly, we use the pre-trained and publicly available original model trained on Notre Dame and apply it to the Hamlyn dataset. During laparoscopy, the abdomen is insufflated providing a more diverse camera trajectory than the predominantly forward-backward movement during colonoscopy. For a query image, we plot database images at the position of their predicted NBOs in Figure 2. We further plot three images and their NNs in Figure 3.

CONCLUSIONS AND DISCUSSION

We demonstrate that image box embeddings can be a useful tool to describe camera movements during endoscopy. The asymmetric measure can interpret relationships between image pairs and relocate an endoscope during surgery even when trained on entirely different data. A drawback is a need for camera pose and depth ground truth during training.

Acknowledgments: This work was supported by EU-H2020 grant 863146: ENDOMAPPER.

REFERENCES

- [1] X. Li, L. Vilnis, D. Zhang, M. Boratko, and A. McCallum, “Smoothing the geometry of probabilistic box embeddings,” in *ICLR*, 2018.
- [2] A. Rau, G. Garcia-Hernando, D. Stoyanov, G. J. Brostow, and D. Turmukhambetov, “Predicting visual overlap of images through interpretable non-metric box embeddings,” in *ECCV*, 2020.
- [3] A. Rau, P. E. Edwards, O. F. Ahmad, P. Riordan, M. Janatka *et al.*, “Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy,” *IJCARS*, 2019.
- [4] D. Stoyanov, M. V. Scanzanella, P. Pratt, and G.-Z. Yang, “Real-time stereo reconstruction in robotically assisted minimally invasive surgery,” in *MICCAI*, 2010.

Accuracy Assessment of a Forward Kinematic Quasi-Static Model of a Concentric Tube Robot for Spina Bifida Repair

X. Zhang¹, G. Borghesan^{1,3}, M. Ourak¹, T. Vandebroek¹, T. Vercauteren^{1,4}, S. Ourselin⁴, J. Deprest^{1,2}, E. Vander Poorten¹

¹ Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Belgium

² Department of Obstetrics and Gynaecology, UZ Leuven, Belgium ³ Core Lab ROB, Flanders Make, Belgium

⁴ School of Biomedical Engineering & Imaging Sciences, King's College London, United Kingdom

INTRODUCTION

Spina bifida aperta (SPA) is a congenital central nervous system malformation caused by incomplete closure of the neural tube, creating a defect in the bony spine. It is one of the most common birth defects affecting between 0.1 to 5 per thousand births in Europe and US [1]. The defect occurs during the first month of the gestational stage and causes progressive nerve damage to the fetus. To reduce risks of complications and mortality, earlier prenatal interventions in the uterus are highly recommended [2]. Conventional minimal invasive surgery where multiple incisions are required to access the uterus would weaken the membrane's strength and increase the risk of rupture. Membrane rupture at early gestational stage leads to extreme premature delivery with possible fatal consequences [2]. Single port access surgery (SPA) is currently explored as an alternative for prenatal spina bifida repair. To support SPA-SBA repair, a Macro-Micro Multi-Arm robot is under development at KU Leuven (Fig. 1).

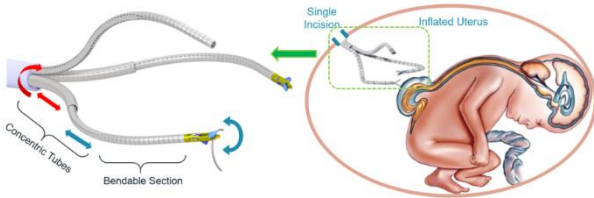


Fig.1 Macro-Micro Multi-Arm robot for spina bifida repair with single port access intervention.

The robot has multiple instrument arms capable of conducting various surgical tasks. Each arm has 3 degree of freedoms (DoFs). Two “Macro” DoFs are realized by a pair of concentric tube forming a so-called “triangulation stage”. By sliding the pair of tubes relative to each other, the overall shape is altered from a straight line (required for entry via the incision) to a first outwards and then inwards bending triangulation shape. It is to be noted that the relative orientation of the pair of tubes does not change such that the triangulation takes place in a single plane. The pair of tubes can be jointly rotated about their shared proximal axis. This allows to rotate the entire triangulation plane and offers a 3rd Macro DoF per instrument arm (tube pair). A fourth “Micro” DoF is provided by a distal bendable section driven by a pneumatic artificial muscle.

Because of the constraints for miniaturization and sterilization, integrating sensors inside this system is not

trivial. Hence, miniature optical sensors (PAT9125EL, PixArt Imaging Inc., China) were installed in the drive system to measure the two proximal displacement DoFs of the concentric tube. This abstract introduces a model to estimate the distal tip's pose as function of the macro DoFs and more in particular the insertion length s_i of the tubes and provides an assessment of the model accuracy.

MATERIALS AND METHODS

The model receives the rotation angle α and insertion distances s_1, s_2 as input and computes the position and shape of the concentric tube structure in 3D workspace. Apart from a pre-bent section at the distal tip, both tubes have straight sections which are connected to linear actuators. At this location, it is possible to integrate miniature optical sensors. The optical sensors measure motion in two perpendicular directions in a plane parallel to the optical sensor's surface (Fig. 2a). After calibration the sensors were found able to measure incremental rotation and translation of the tubes (Fig. 2b) with

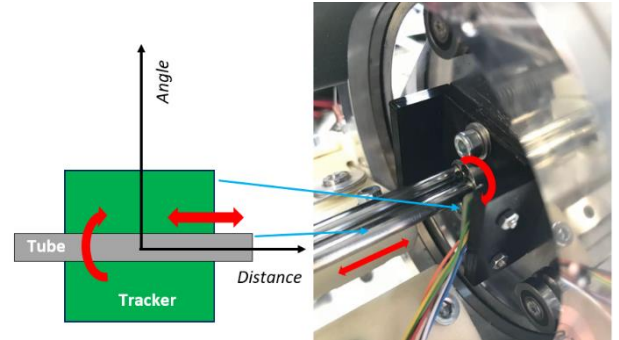


Fig. 2 (a) Data acquisition illustration (b) Optical tracker installation

respectively a resolution of 0.15 degrees and 22 micrometer per count. Due to the relative motion between the inner and outer tubes, separate trackers are necessary per tube.

The pre-bent sections of the tubes have different curvatures κ_i and curve lengths l_i at the distal section as:

$$\kappa_1 = \begin{cases} \frac{1}{55} \text{ mm}^{-1} & \in l_1 = [0, 64.5\text{mm}] \\ 0 \text{ mm}^{-1} & \in l_1 = [64.5, 70\text{mm}] \end{cases} \quad (1)$$

$$\kappa_2 = \begin{cases} \frac{1}{57} \text{ mm}^{-1} & \in l_2 = [0, 53\text{mm}] \\ 0 \text{ mm}^{-1} & \in l_2 = [53, 120\text{mm}] \end{cases} \quad (2)$$

When the inner tube slides inside the outer tube both bent and straight sections will overlap and interact. The combined curvature $K_{com,j}$ of the overlapping concentric

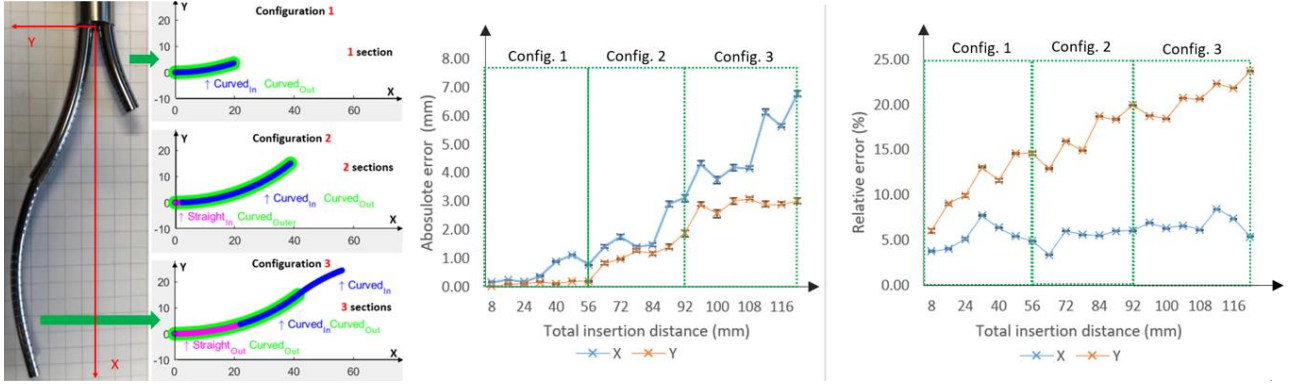


Fig. 3(a) 2D reconstruction from the model for all configurations.

3(b) Absolute and relative errors in 2D space

section j follows the curvatures of both tubes as in

$$K_{com,j} = \frac{\sum_{i=1}^n E_i I_i \kappa_i}{\sum_{i=1}^n E_i I_i} \quad (3)$$

where E_i is the Young's modulus of the Nitinol alloy and I_i is the moment of inertia of the tubes. A key assumption applied here is that the original pre-bent curvatures κ_i of both tubes do not alter over time. During the deployment of the instrument arm, the sliding motion of the two tubes changes the combined curvatures $K_{com,j}$ along the arm. The instrument tube has three different configurations in the final shape (Fig. 3a). With the curvature $K_{com,j}$ and curve length $l_{s,j}$ of one section derived from the insertion distances s_1, s_2 , the transformation matrices T_j^{j-1}

$$T_j^{j-1} = \begin{bmatrix} \cos(K_{com,j} l_{s,j}) & -\sin(K_{com,j} l_{s,j}) & 0 & \frac{\sin(K_{com,j} l_{s,j})}{K_{com,j}} \\ \sin(K_{com,j} l_{s,j}) & \cos(K_{com,j} l_{s,j}) & 0 & \frac{1 - \cos(K_{com,j} l_{s,j})}{K_{com,j}} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

and the 2D shape could be determined following [3]. Because the pre-bent orientations of two tubes share one common plane, the model can be multiplied with another transformation matrix $R(\alpha)$ to produce additional rotation needed for 3D experiments. In the experiments the model was compared to a ground truth measurement of the tip. Hereto the tip position was measured by integrating two electromagnetic sensors (Aurora Micro 6DoF, NDI, Canada) at the tip and the base of the instrument arm.

It was found that the overall accuracy is mainly determined by the quality of the 2D shape estimation. Since the input measurement could be decoupled from the optical trackers, the model was first validated with the tip position in 2D. Twenty insertion distances with equal increments were chosen to cover all three concentric tube configurations. For each set, 5 iterations were conducted to get averages and standard deviations of the estimates.

RESULTS

The average position errors at insertion X and bending Y directions were shown in absolute and relative values with respect to the total insertion distances (Fig. 3b). The total insertion distances were divided into groups of three configurations described in Fig. 3a.

The errors increased with the deployment of the instrument arm. At two transition stages between three configurations, the error changed drastically and then remained within the same range until the next transition. The maximum absolute error in tip position occurred when the tubes were fully inserted by actuators. The absolute error in the insertion direction X reached 6.32 mm and the one in the bending direction Y reached 3.09 mm at the final stage. The level of accuracy was not sufficient for the intended medical application where the required error should be less than 1 mm. Modifications are required for the model to reduce the error in 2D reconstruction before 3D validations.

Because the ranges of motion in two directions were different, error analysis in percentage were necessary for determining the main source of error. The results indicated that the maximum error in percentage in direction Y was 18.3% and the one in X was 8.5%.

CONCLUSION AND DISCUSSION

This abstract introduced a forward kinematic model together with optical sensor to estimate the distal motion of an SPA instrument for SBA treatment. Estimation errors were up to 6.32 mm indicating that the model can be used for gross positioning. The error accumulation was mainly because the default curvature of the two tubes changed during their interactions. Hence, adjusting the constant curvature assumption in the model and quantifying the changing curvatures could help improve the model's accuracy.

ACKNOWLEDGMENTS

This work was funded by PASSAGE (3E201024), an FWO-project, and from an innovative Engineering for Health Award by Wellcome Trust (WT101957) and EPSRC (NS/A000027/1).

REFERENCES

- [1] Sacco, A (2019). Fetal surgery for spina bifida. *Obstetrics Gynaecology & Reproductive Medicine*. 10.1016/j.ogrm.2019.10.005.
- [2] Blumenfeld, Yair J (2018). Updates in fetal spina bifida repair. *Current Opinion in Obstetrics and Gynecology* 30.2 123-129.
- [3] Vandebroek, T (2019). Macro-Micro Multi-Arm Robot for Single-Port Access Surgery. *IROS*, 2019, pp. 425-432.
- [4] Paul GC (2018). Single Port Access (SPATM) Minimal Access Surgery Through a Single Incision. *Surgical technology international*. 18. 19-25.

A Modular Organ Perfusion Research Platform for Computer Assisted Imaging and Intervention

K. Doyle¹, M. Magbagbeola¹, Z.L. Rai^{1,2,3}, L. Lindenroth¹, G. Dwyer¹,
D. Waterhouse¹, A. Stilli¹, B.R. Davidson^{1,2,3}, D. Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London,

²Centre for Surgical Innovation, Organ Repair and Transplantation (CSIORT), University College London

³ Royal Free Hospital NHS Trust, London
katie.doyle@ucl.ac.uk

INTRODUCTION

Improvements in surgical technique and pharmacological anti-rejection medications have made organ transplantation not only a viable treatment option for many but a preferred one in recent times [1]. While demand for organ transplants has increased in recent decades, the availability of viable organs has not kept pace. One solution that is currently being evaluated is machine perfusion (MP). MP has promising research-based applications. Previous studies have demonstrated the increased clinical relevancy of perfused *ex vivo* organ models over their nonperfused counterparts [2]. Clinical MP systems are specialised for organ transplantation, with tightly controlled operation parameters and a high degree of sterility. This reduces the overall versatility of the systems, rendering them unsuitable for use as research platforms. We present an MP system specifically designed as a platform for a wide range of research applications. This system provides an affordable alternative to commercially available clinical MP systems, with a total parts cost of approximately £11,000. Alongside its cost-effectiveness, this MP system is designed to be versatile and dynamic, allowing real-time monitoring of parameters of most interest and applicability to the specific needs of the user. Here, we present preliminary findings wherein the system is used to record spectroscopic data in a perfused porcine liver which is correlated to real-time blood oxygen saturations of the perfusate.

MATERIALS AND METHODS

Livers were obtained from domestic pigs immediately after termination at an abattoir and prepared for transport in situ. Whole red blood cells from autologous porcine donation were collected to be used as a perfusate. The portal vein (PV) and hepatic artery (HA), designated for inflow, and the hepatic vein (HV), designated for outflow, were cannulated using 6.35mm tubing. The bile duct was ligated to prevent bile leakage and blood flushed from the organ using cooled heparinised saline solution. The liver was weighed prior to connection to the system. The oxygenator heated the perfusate to 29°C and, during the oxygenated phase of data collection, pure oxygen was supplied at 0.2bar.

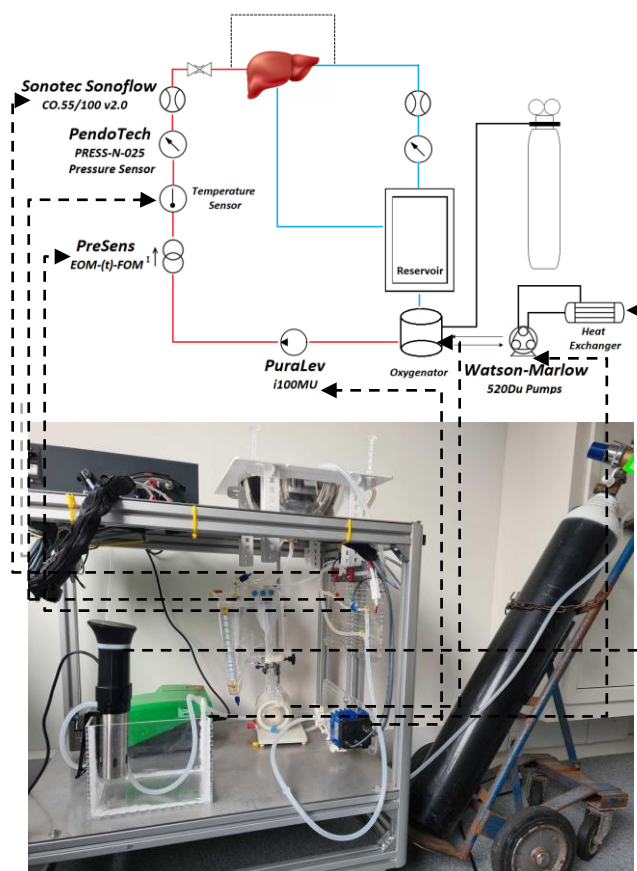


Figure 1: An overview of a research-orientated perfusion system.

System monitoring and control was coordinated by a raspberry pi microcontroller and used a MODBUS and RS232 serial communication protocol to communicate with the sensors. The raspberry pi recorded and displayed the raw data from all sensing equipment continuously in real-time and recorded it for further analysis. A predetermined inflow pressure was maintained by a control algorithm which altered pump speed based on inflow pressure readings via a closed-loop PID controller. Additionally, the system has the capability of maintaining an input flow rate through the use of a cascade controller. The weight of the organ was monitored using four load cells located at the base of the organ chamber. Pressure (PendoTech, Press-n-075), flow

(Sonotec Sonoflow, co.55/100), temperature (PreSens, Pt100), and blood oxygen (PreSens, EOM-(t)-FOM) sensors were connected in series in the inflow circuit branch. The outflow branch was monitored by additional pressure and flow sensors, as shown in Figure 1. Blood from the outflow branch and runoff collected in the organ chamber was returned to the reservoir, where it was filtered. The oxygenator re-oxygenated blood and warmed it via a heat exchanger and was controlled via a peristaltic pump (Watson-Marlow, 520Du).

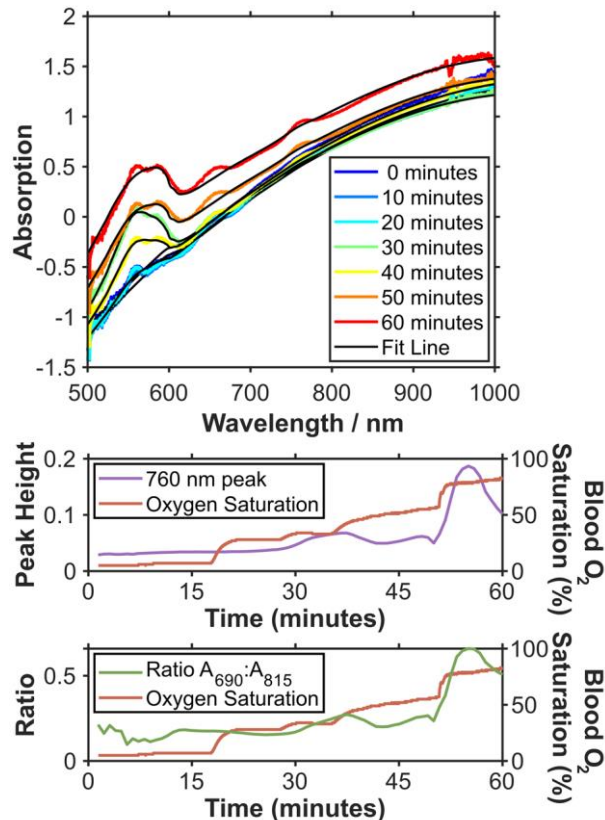


Figure 2: Top: Average tissue absorption spectra in 10 minute intervals. Middle: Change in height of 760 nm peak in the absorption spectra, compared to blood oxygen concentration. Bottom: Ratio of change in intensity of 690 nm and 815 nm spectral regions, compared to blood oxygen concentration.

Diffuse optical spectroscopy was carried out by securing a fibre optic reflection probe, supplied by a tungsten-halogen light source, against the surface of the liver. Reflected light was collected by a charge coupled device spectrometer. Calibration measurements with the light source on and off were taken with the probe against a white calibration card, and against the liver. The system was run and the liver perfused for one hour, during which, measurements were taken at regular intervals prior to and during perfusion with blood. Following perfusion with deoxygenated blood, the blood was gradually oxygenated. Once processed, the spectroscopy measurements were synchronised with and compared to measurements recorded by the MP system perfusate oxygen sensor.

RESULTS

Results from the spectroscopy data were post-processed in MATLAB; results are shown in Figure 2. A model of blood attenuation was fitted to the acquired absorption spectra to determine a value for tissue oxygen saturation; however, it did not account for the peak at 660 nm, which may result from the heparin present in the blood. The 760 nm peak height was analysed as a marker of deoxygenated blood, as this peak only exists in the spectrum of deoxyhaemoglobin (Hb). The 690 nm spectral region was selected as the region of greatest difference between Hb and oxyhaemoglobin (HbO₂) spectral intensity. The 815 nm spectral region was selected as it is the isobestic point of the spectrums of Hb and HbO₂. Figure 2 shows a clear change in tissue absorption over the duration of the perfusion. While there is a delay between measured blood oxygen saturation and the corresponding change in absorption spectra peaks, there is an increase in the 760 nm peak, as the organ is perfused with blood, and a decline in the peak, as the ratio of deoxygenated to oxygenated blood decreases, as expected.

CONCLUSION AND DISCUSSION

Machine perfusion is proving to be invaluable in clinical practice [3][4], and we propose that a modified computer assisted MP system can also benefit the research community. By providing an alternative research platform, such a system can assist researchers in monitoring key parameters relevant to their discipline and area of interest. Integration of several sensing components makes such a system modular and adaptable. Our results demonstrate that whole *ex-vivo* organ perfusion with a computer assisted system is both feasible and a promising avenue for the development of an innovative research platform. Multiple physiological parameters can be integrated within the perfusion circuit and within this study we have successfully validated oxygen saturation. Our perfusion circuit allows for a greater range of monitoring with the addition of various parameters, as per the requirements of individual experiments. More work is underway to demonstrate the versatility of such a cost-effective research platform.

REFERENCES

- [1] D. Abramowicz *et al.*, "Recent advances in kidney transplantation: A viewpoint from the Descartes advisory board," *Nephrol. Dial. Transplant.*, vol. 33, no. 10, pp. 1699–1707, 2018, doi: 10.1093/ndt/gfx365.
- [2] P. N. Siriwardana *et al.*, "Effect of Hepatic Perfusion on Microwave Ablation Zones in an Ex Vivo Porcine Liver Model," *J. Vasc. Interv. Radiol.*, vol. 28, no. 5, pp. 732–739, 2017, doi: 10.1016/j.jvir.2016.03.006.
- [3] D. Nasralla *et al.*, "A randomized trial of normothermic preservation in liver transplantation," *Nature*, vol. 557, no. 7703, pp. 50–56, 2018, doi: 10.1038/s41586-018-0047-9.
- [4] A. M. Rappaport, "Hepatic blood flow: morphologic aspects and physiologic regulation.," *Int. Rev. Physiol.*, vol. 21, no. 21, pp. 1–63, 1980, [Online]. Available: <https://europepmc.org/article/med/6993392>.

Concurrent Hierarchical Autonomous Surgical State Estimation during Robot-assisted Surgery

Yidan Qin^{1,2} and Joel W. Burdick²

¹Intuitive Surgical Inc., 1020 Kifer Road, Sunnyvale, CA, 94086, USA

²Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA, 91125, USA

INTRODUCTION

Robot-Assisted Surgery (RAS) technology has been widely adopted for its improved surgical precision and outcomes [1]. RAS systems generate a wide variety of synchronized data, which can drive artificial intelligence (AI) applications [2]–[4] such as the presentation of advisory information [5] and surgery workflow analysis [6], etc. These applications share the constitutive prerequisite of the real-time awareness about the current stage of a surgery [7]. A surgical Hierarchical Finite State Machine (HFSM) models the temporal evolution of an RAS procedure at multiple levels of temporal granularity, including the current surgical task (superstates) and the fine-grained states that model a surgeon’s actions and operation environments. At each instant, the current surgical superstate and fine-grained state are often highly correlated. E.g., during tissue dissection, tissue cutting is a commonly observed fine-grained state. However, it is not observed during suturing. The fine-grained state of needle manipulation is frequently observed during suturing, but not during tissue dissection. Such hierarchical correlations are commonly present in a surgical HFSM, and can improve the accuracy of surgical (super)state estimation. Previous methods have separately explored surgical superstate or fine-grained state estimations [8]–[13] with Convolutional Neural Networks (CNNs), Long-Short Term Memory (LSTM), etc.; however, these methods relied only on *direct data sources* such as the surgical robot’s kinematics [8], [14], endoscopic video streams [2], [9], and the surgical system’s events [4], [7]. The surgical (super)state time series are *inferred data sources* that have yet to be incorporated into the estimation process at other levels of temporal granularity. Since hierarchical correlations between surgical states are common and informative, the knowledge of the current surgical superstate can aid the fine-grained surgical state estimation, and vice versa.

We propose Concurrent Hierarchical Autonomous Surgical State Estimation Network (CHASSEN): a learning-based hierarchical surgical state estimation model that estimates surgical superstate and fine-grained states during RAS. CHASSEN achieves accurate and efficient surgical (super)state estimation through the learning and use of hierarchical correlations between states at multiple levels of temporal granularity. Our contributions include:

- Using both direct (robot kinematics, endoscopic vision, system events) and inferred data (current sur-

gical (super)state) as input sources for concurrent hierarchical surgical (super)state estimation;

- Learning the hierarchical correlations between surgical states at two levels of temporal granularity through an alternating training schematics;
- Improving surgical (super)state estimation accuracy by 4.3% and processing time by 31% comparing to existing methods in a real-world RAS dataset.

Comparing to state-of-the-art hierarchical surgical state estimation methods [7], CHASSEN’s learning and utilization of hierarchical correlations between surgical states at multiple levels of temporal granularity allowed it to achieve a higher state estimation accuracy with a more lightweight network architecture and higher estimation speed, which shows a strong promise for more effective and efficient hierarchical surgical state estimation ability.

MATERIALS AND METHODS

We used HERNIA-20 - a real-world RAS dataset containing 20 robotic inguinal hernia repair surgeries [4] - to train and evaluate CHASSEN. HERNIA-20 is annotated with 8 surgical superstates and 21 fine-grained surgical states. Hierarchical correlations exist commonly among the (super)states, including the likelihood distributions of state occurrence, the surgery’s temporal progresses, etc.

CHASSEN includes feature extraction and surgical (super)state estimation modules and accepts both direct RAS data and inferred data sources as inputs. The surgical state-related features are extracted from direct RAS data sources via a neural network-based feature extraction module (Fig. 1). Hierarchical correlations between surgical superstates and fine-grained states are captured through CHASSEN’s training schematics (Fig. 2). Following [4], CHASSEN

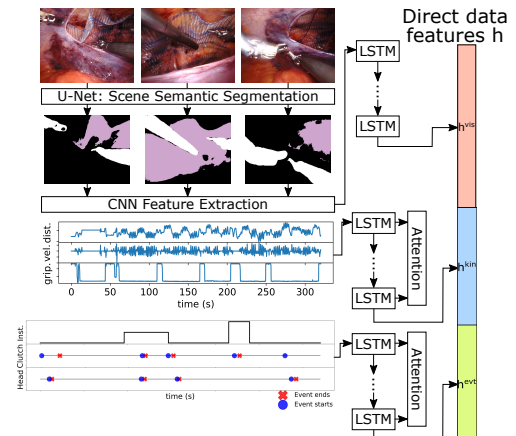


Fig. 1: CHASSEN’s feature extraction module.

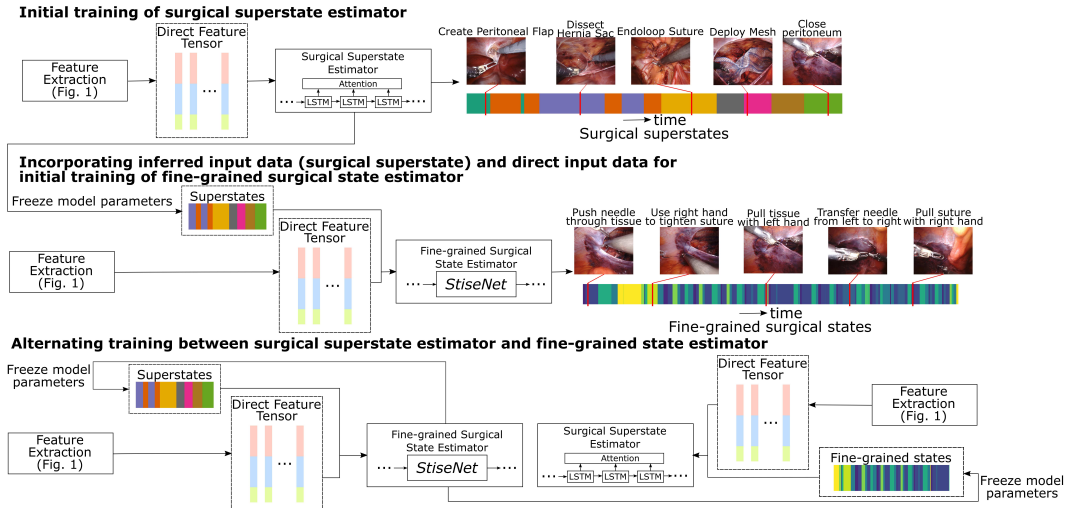


Fig. 2: CHASSEN's model architecture and its alternating training schematics.

uses synchronized time series data streams to leverage a rich representation of the surgical scene. To eliminate variability in the anatomical background, a semantic mask of the endoscopic vision image containing three scene classes (instruments, tissue, others) is first generated by a U-Net [15] based surgical scene segmentation model that was extensively trained and frozen. LSTM encoders embed the surgery's temporal progress into direct data features $\mathbf{h} = [\mathbf{h}^{\text{vis}}, \mathbf{h}^{\text{kin}}, \mathbf{h}^{\text{evt}}]$. An attention mechanism [16] manages the diversity of the robot kinematics and system events data sources. The feature extraction module will be further discussed in the final submission.

Knowledge of the current surgical superstate substantially improves the estimation of the current fine-grained surgical state, as the likelihood of occurrence and probability distribution of the fine-grained surgical states is correlated with the superstate. Similarly, knowledge of the current fine-grained state improves the superstate estimation process. We therefore adapt an alternating training schedule that learns such hierarchical correlations (Fig. 2). CHASSEN initially trains the superstate estimator only with direct input data. The embedded latent representation of direct input data forms a feature tensor over an observation window, and an LSTM decoder with an addition attention mechanism [16] was used for surgical superstate estimation [7]. The trained superstate estimator then generates the inferred data source - the surgical superstate time series. An LSTM encoder extracts the temporal correlations between superstates and their corresponding fine-grained surgical states, which is incorporated for the initial training of the fine-grained state estimator. CHASSEN implements *StiseNet* [4] - a robust model for fine-grained surgical state estimation. We then adopt an alternating training schedule between the fine-grained surgical state estimator and the surgical superstate estimator: in each training iteration, a previously trained and frozen estimator is used to generate either the surgical superstate or fine-grained state estimation result, which is concatenated with direct features to fine-tune the parameters of the other state estimator. The iteration repeats until convergence of the surgical (super)state estimation performance. Further details and hyperparameters of CHASSEN's training schematics will be discussed in the final submission.

CONCLUSIONS AND DISCUSSION

CHASSEN's hierarchical surgical (super)state estimation performance is compared against our previous work *HESS-DNN* [7]. *HESS-DNN* produces superstate and fine-grained state estimations through separate models in a decoupled manner, in which little knowledge of the surgical states at two levels of temporal granularity was shared during training or inference. We show the performance improvement of CHASSEN through two metrics: the percentage of time steps in the test set with correctly estimated surgical (super)states and the model's processing time on a workstation with an Intel Core i7-6700 CPU, 24GB RAM, and an NVIDIA RTX 2060 graphics card. To accommodate real-time applications, CHASSEN only uses data from the preceding and current time steps.

CHASSEN achieves a surgical superstate estimation accuracy of 88.4%, a 4.3% improvement over *HESS-DNN*, which estimates superstates without any knowledge about the related fine-grained states. CHASSEN also improves the fine-grained state estimation performance to 77.3% comparing to *HESS-DNN*'s 75.7%. CHASSEN's improvements in both superstate and fine-grained state estimation accuracy highlight the importance of correlations between hierarchical surgical states. Additionally, CHASSEN's lighter network architecture (as compared to *HESS-DNN*) significantly improves its state estimation processing time. Since *HESS-DNN* performs hierarchical state estimation in a decoupled manner, it uses multiple deep neural network architectures trained independently [7]. Comparing to *HESS-DNN*'s 7.1 frames per second (fps) operation, CHASSEN estimates both surgical superstate and fine-grained states at 9.3fps - a 31% gain.

In our initial study, CHASSEN performs more accurate hierarchical surgical state estimation with greater efficiency due to its incorporation of hierarchical correlations between surgical (super)states. Ongoing work evaluates CHASSEN more extensively and systematically to fully explore efficient hierarchical surgical state estimation. Our final submission will better describe CHASSEN's network architecture and training methods. A quantitative analysis of the contributions of correlations across hierarchical states is among our future work.

REFERENCES

- [1] G. I. Barbash, "New technology and health care costs—the case of robot-assisted surgery," *New Eng. J. Medicine*, vol. 363, no. 8, p. 701, 2010.
- [2] Y. Qin, S. A. Pedram, S. Feyzabadi, M. Allan, A. J. McLeod, J. W. Burdick, and M. Azizian, "Temporal segmentation of surgical sub-tasks through deep learning with multiple data sources," in *IEEE Int. Conf. Robotics and Automation*, 2020, pp. 371–377.
- [3] Y. Qin, S. Feyzabadi, M. Allan, J. W. Burdick, and M. Azizian, "davincinet: Joint prediction of motion and surgical state in robot-assisted surgery," *arXiv preprint arXiv:2009.11937*, 2020.
- [4] Y. Qin, M. Allan, Y. Yue, J. W. Burdick, and M. Azizian, "Learning invariant representation of tasks for robust surgical state estimation," *IEEE Robotics Automation Lett.s*, vol. 6, no. 2, pp. 3208–3215, 2021.
- [5] P. Chalasani, A. Deguet, P. Kazanzides, and R. H. Taylor, "A computational framework for complementary situational awareness in surgical assistant robots," in *IEEE Int. Conf. Robot. Comp.*, 2018, pp. 9–16.
- [6] N. Padoy, "Machine and deep learning for workflow recognition during surgery," *Minimally Invasive Therapy & Allied Technologies*, vol. 28, no. 2, pp. 82–90, 2019.
- [7] Y. Qin, M. Allan, J. Burdick, and M. Azizian, "Autonomous hierarchical surgical state estimation during robot-assisted surgery through deep neural networks," *IEEE Robotics and Automation Letters*, 2021.
- [8] C. Lea, R. Vidal, and G. D. Hager, "Learning convolutional action primitives for fine-grained action recognition," in *IEEE Int. Conf. Robotics and Automation*, 2016, pp. 1642–1649.
- [9] R. DiPietro, C. Lea, A. Malpani, N. Ahmadi, S. S. Vedula, G. I. Lee, M. R. Lee, and G. D. Hager, "Recognizing surgical activities with recurrent neural networks," in *Int. Conf. Med. image Comp. and Comp.-Assist. Intervention*, 2016, pp. 551–558.
- [10] G. Menegozzo, D. Dall'Alba, C. Zandonà, and P. Fiorini, "Surgical gesture recognition with time delay neural network based on kinematic data," in *Int. Symp. Medical Robotics*, 2019, pp. 1–7.
- [11] E. Mavroudi, D. Bhaskara, S. Sefati, H. Ali, and R. Vidal, "End-to-end fine-grained action segmentation and recognition using conditional random field models and discriminative sparse coding," in *IEEE Winter Conf. App.s Computer Vision*, 2018, pp. 1558–1567.
- [12] T. Yu, D. Mutter, J. Marescaux, and N. Padoy, "Learning from a tiny dataset of manual annotations: a teacher/student approach for surgical phase recognition," *arXiv preprint arXiv:1812.00033*, 2018.
- [13] A. Zia, C. Zhang, X. Xiong, and A. M. Jarc, "Temporal clustering of surgical activities in robot-assisted surgery," *IJCARS*, vol. 12, no. 7, pp. 1171–1178, 2017.
- [14] L. Tao, L. Zappella, G. D. Hager, and R. Vidal, "Surgical gesture segmentation and recognition," in *Int. Conf. Med. Image Comp. and Comp.-Assist. Intervention*. Springer, 2013, pp. 339–346.
- [15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Int. Conf. Med. image Comp. and Comp.-Assist. intervention*, 2015, pp. 234–241.
- [16] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv preprint arXiv:1704.02971*, 2017.

Proof-of-Concept Medical Robotic Platform for Endovascular Catheterization

Mohammad Hasan Dad Ansari^{1,2,3,*}, Beatriz Farola Barata^{3,4,*}, Fabian Trauzettel^{3,5,*}, Zhen Li^{5,6,*}, Di Wu^{3,5,*}, Diego Dall'Alba⁴, Gianni Borghesan³, Mouloud Ourak³, Veronica Iacovacci^{1,2}, Selene Tognarelli^{1,2}, Jenny Dankelman⁵, Elena De Momi⁶, Paul Breedveld⁵, Paolo Fiorini⁴, Jos Vander Sloten³, Arianna Menciassi^{1,2}, and Emmanuel Vander Poorten³

¹The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa 56025, Italy

²Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Pisa 56025, Italy

³Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Leuven 3001, Belgium

⁴Altair Robotics Laboratory, Department of Computer Science, University of Verona, Verona 37134, Italy

⁵Department of Biomechanical Engineering, Delft University of Technology, Delft 2628 CD, Netherlands

⁶NearLab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan 20133, Italy
hasan.mohammad@santannapisa.it, hasan.mohammad@kuleuven.be

INTRODUCTION

In endovascular interventions, vessels serve as access routes to deep and remote anatomic regions [1]. Navigation through narrow, fragile, and deformable vessels in endovascular procedures requires considerable skill [2]. Robotic distributed control approaches could improve catheter navigation to bypass anomalies and prevent tissue damage [3]. For particularly narrow vessels, remote actuation through magnetic fields could help in performing complex coordinated motion in three-dimensional (3D) space as magnetic fields are safe and highly controllable [4]. Visualization during endovascular catheterization procedures mainly relies on fluoroscopy, an X-ray based imaging modality that only offers two-dimensional (2D) views of the interventional scene [5]. Conventional approaches are thus characterized by poor situational awareness as constructing a 3D representation from 2D views is mentally demanding. Generating a 3D view via a C-arm is time-consuming and can only be done sporadically to limit patient radiation. Intra-operative 3D vessel representations from non-ionising imaging sources could greatly improve ease of navigation of the medical instruments to target anatomic sites. Another challenge is to provide safe guidance when navigating in fragile vessels from a restricted access point [6]. This could be addressed by path planning to search for a feasible path connecting a start to a goal configuration, while considering the robotic systems constraints and characteristics.

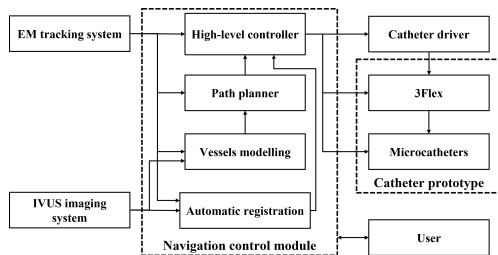


Fig. 1: Overall schematic of proposed medical robotic platform.

In this work, a prototype of a medical robotic platform aimed at endovascular catheterization is developed (Fig. 1). The system integrates several components: a multi-lumen catheter shaft, termed 3Flex, magnetically actuated microcatheter and a navigation control module conceived to help teleoperation and/or autonomous navigation of the 3Flex. This module comprises 3D vessel modelling, automatic registration, path planning, and catheter control. The 3Flex catheter is

designed to carry within it two 6 degrees-of-freedom (DOFs) electromagnetic (EM) tracking sensors, one intravascular ultrasound (IVUS) catheter, and two magnetically-actuated microcatheters envisioned to be controlled by a permanent magnet carrying KUKA robotic arm. However for this prototype, only one microcatheter was used.

MATERIALS AND METHODS

The 3Flex (Fig. 2) is a 3D printed catheter with an outer diameter of 12mm, a length of 500mm, and a 75mm long 2 DOF steerable tip actuated by four integrated pneumatic artificial muscles (PAMs). Its EM tracking sensors are located in two 10mm long sections proximally and distally adjacent to the steerable tip; both are fixed in place in their own lumens and provide position data to the control module. While the design can accommodate a coaxial Fibre Bragg Grating (FBG) for shape reconstruction of the distal segment, it was not implemented here.

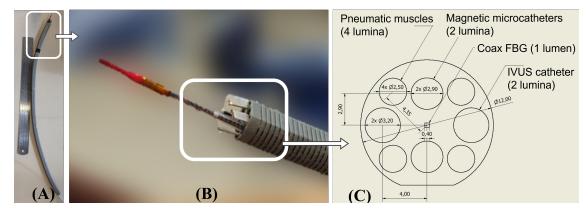


Fig. 2: A) 500mm long 3Flex system (beside a 300mm long ruler); B) view on tip of 75mm active distal section C) cross-sectional view showing embedded functionalities in the 3Flex's lumina.

Autonomous trajectory planning is designed into the integrated system. Given the preferable anchor region, targets (i.e. two coronary ostia), and start region (patient groin), an optimal path and anchor pose were generated and forwarded to the user. During the trajectory planning, following constraints and system characteristics were incorporated: tip length, bending capability, 3Flex's outer diameter, magnetic microcatheter length and offset between magnetic microcatheter and 3Flex catheter's main axis. Real-time IVUS and EM tracking-based vessel modelling technology is integrated in this system. With it, a local representation of the vasculature of the vessel as a 10mm long cylinder-shape at the level of the catheter tip is estimated in real time. This 3D vessel modelling approach provides a 3D representation of the local vasculature for visualization and assists catheter navigation. Navigation assistance is achieved by outputting the 3D vessel model as well as its positioning relative to the 3Flex tip to the control strategy. Furthermore, an online automatic registration approach is implemented based on the intra-operative IVUS and EM tracking data. This registration method requires

*These authors contributed equally to this work.

This work was supported by the ATLAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813782.

minimal user interaction by exploiting either a stochastic branch and bound method or an iterative closest point algorithm. It produces a transformation matrix between the EM tracker's coordinate frame and a pre-operative mesh frame [7]. This automatic registration runs continuously. Additional information regarding the vessels contour (points) are used to refine the registration between the pre-operative and the intra-operative (reconstructed) geometries. From the above IVUS-based strategies, enhanced catheter navigation is attained as the pre-operative geometry can be overlaid with the estimated local vessel models and the catheter pose. Consequently, the catheter pose and shape awareness can be significantly improved. To communicate this information the estimated local cylinder models and the reconstructed vessel registered with the pre-operative mesh was rendered both in a graphical user interface (GUI) based on the Unity engine (Fig. 3, left) and visualized in augmented reality using Microsoft's® HoloLens™ 2 (Fig. 3, right).

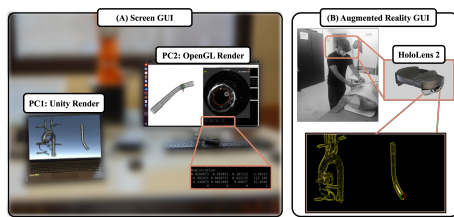


Fig. 3: Overview of GUI of intra-operative 3D vessel modelling, presenting (A) screen GUI including Unity rendering (PC1) and OpenGL rendering (PC2); (B) augmented reality GUI via HoloLens 2. The transformation matrix computed by the automatic registration strategy is also displayed in the middle.

The 3Flex has three DOFs of which two are catheter bending, while one is catheter insertion/retraction. Four PAMs are configured in two antagonistic pairs, allowing bi-directional bending of the 3Flex tip with a 80° maximum bending angle. To control the bending angle, input pressures to the four PAMs are modulated by four proportional air pressure valves (Festo group, Germany). These valves receive control signals from an analog output module NI-9144 (NI, Texas, USA). Currently tele-operated, 3Flex bending will be semi-autonomous in the future. A sleeve-based catheter driver, which has two pneumatically actuated grippers that grasp the catheter alternately and insert the catheter in a relay fashion, (Fig. 1, top right) is used to advance/retract the 3Flex.

The magnetic microcatheter carried by the 3Flex is composed of a soft polymer (UV Electro 225-1, Momentive, Germany) with embedded magnetic particles (MQFP-15-7, Magnequench, Germany). Once protruding from the 3Flex, it can be remotely actuated by a permanent magnet that is either robot-mounted or hand-held. The hollow microcatheter can introduce other sensors or guidewires for easier cannulation or for returning to earlier reached locations.

RESULTS

Trajectory planning was employed for the integrated system and the visualization of the inputs and outputs of the path planner is presented in Fig. 4. Fig. 3 depicts the output of the 3D vessel modelling approach and the online automatic registration strategy. The former determines, at each time step, an estimated local cylinder in the vicinity of the 3Flex tip. The latter helps to compute the transformation matrix of the intra-operative IVUS data with respect to the pre-operative geometry and overlay the estimated cylinder for catheter navigation. Both methods rely on the EM tracking data

relative to the 3Flex tip pose and the vessel lumen contours, segmented from IVUS images.

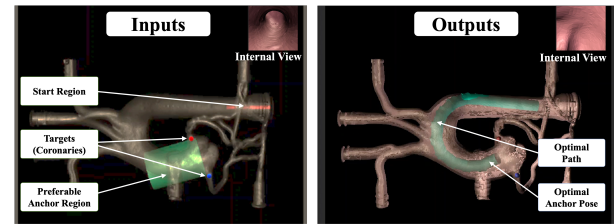


Fig. 4: Visualization of the inputs and outputs of the path planner.

To qualitatively evaluate the system, the 3Flex catheter was steered in an aortic phantom by using a joystick. The planned path was displayed to the user controlling the joystick. The catheter was successfully steered from the descending aorta to the coronary ostia (Fig. 5), by utilizing deflection from contact points. The optimal deflection points were determined by the navigation control module by considering the maximum bending angle/radius of the 3Flex. The magnetic microcatheter was guided to the right coronary artery (RCA) based on external visual feedback (Fig. 5) for now. However, in the future, integration of FBG- and EM- based shape reconstruction will replace the need for visual feedback.

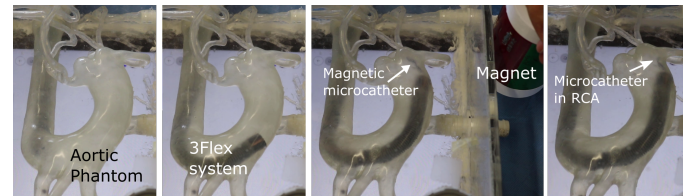


Fig. 5: 3Flex system being steered in the aorta and delivering the microcatheter to the aortic root from where the external magnet guides the magnetic microcatheter to the right coronary artery.

DISCUSSION AND CONCLUSION

The 3Flex was successfully tele-operated to the target site by joystick, further showing good stability during the manipulation of the magnetic microcatheter. Planned future work consists of integrating an autonomous control strategy of the 3Flex catheter to follow the pre-planned trajectories, while considering the intra-operative vessel model in the decision-making. Though only one microcatheter was used for this demonstration, the 3Flex has been designed to carry two within it which can be pushed one after the other and guided to both coronary arteries. Finally, while the proposed prototype is focused on endovascular catheterization, the developed technologies can be transferable to other intraluminal clinical scenarios, e.g. colonoscopy or ureteroscopy.

REFERENCES

- [1] D. Guez *et al.*, "Recent trends in endovascular and surgical treatment of peripheral arterial disease in the medicare population," *Journal of Vascular Surgery*, vol. 71, no. 6, p. 2178, 2020.
- [2] R. Aggarwal *et al.*, "Virtual reality simulation training can improve inexperienced surgeons' endovascular skills," *European Journal of Vascular and Endovascular Surgery*, vol. 31, no. 6, 2006.
- [3] Y. Yang *et al.*, "Adaptive distributed control of a flexible manipulator using an iterative learning scheme," *IEEE Access*, vol. 7, 2019.
- [4] F. Carpi *et al.*, "Stereotaxis Niobe® magnetic navigation system for endocardial catheter ablation and gastrointestinal capsule endoscopy," *Expert Review of Medical Devices*, vol. 6, no. 5, 2009.
- [5] H. Rafii-Tari *et al.*, "Current and emerging robot-assisted endovascular catheterization technologies: A review," *Annals of Biomedical Engineering*, vol. 42, no. 4, 2014.
- [6] T. da Veiga *et al.*, "Challenges of continuum robots in clinical context: A review," *Progress in Biomedical Engineering*, vol. 2, no. 3, 2020.
- [7] C. Grijthuijsen *et al.*, "An automatic registration method for radiation-free catheter navigation guidance," *Journal of Medical Robotics Research*, vol. 1, no. 3, 2016.

Autonomous image guided control of endoscopic orientation for OCT scanning

Fernando Gonzalez Herrera^{1,2,*}, Ameya Pore^{3,4,*}, Luca Sestini^{1,5,*}, Sujit Kumar Sahu^{1,6,*}, Guiqiu Liao^{1,3,*}, Philippe Zanne¹, Diego Dall'Alba³, Albert Hernansanz⁴, Benoit Rosa¹, Florent Nageotte¹, Michalina J Gora⁷

¹ ICube, University of Strasbourg, Strasbourg 67000, France

² Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Leuven 3001, Belgium

³ Altair Robotics Laboratory, Department of Computer Science, University of Verona, Verona 37134, Italy

⁴ Biomedical engineering, Universitat Politècnica de Catalunya, Barcelona 08028, Spain

⁵ NearLab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan 20133, Italy

⁶ The BioRobotics Institute and Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Pisa 56025, Italy

⁷ Wyss Center for Bio and Neuroengineering, Campus Biotech, Geneva 1202, Switzerland

gonzalezherrera@unistra.fr

INTRODUCTION

Minimally invasive gastrointestinal procedures such as Endoscopic Submucosal Dissection (ESD) are challenging to perform with standard flexible endoscopes [1]. Hence, robotic solutions have been proposed to improve dexterity and manipulation in intraluminal surgeries. Conventional systems such as Da Vinci (Intuitive Surgical) have been used for transanal access, but the usage of rigid instruments limits the application to operating in the vicinity of the natural access and makes manipulation in constrained areas highly challenging. Recently, STRAS robotic platform [1] has shown feasibility for intraluminal procedures in terms of safety and dissection speed with respect to manual systems.

Optical Coherence Tomography (OCT) embedded in continuum robots offers minimally invasive scanning of internal tissues and organs with micrometre resolution and millimetre penetration depth. However, due to the limited depth perception, to achieve precise scanning, contact between the tissue should be maintained during scanning. Performing such tasks requires controlling several DoFs while relying on both the endoscopic camera and OCT images. This procedure has been proven to be difficult to realise by users, even in telemanipulation. In this context, automatic repositioning of the endoscope could allow deploying the OCT probe accurately and more easily.

The objective of this study is to develop autonomous endoscope positioning to facilitate OCT sensing in a phantom environment (Fig.1a).

MATERIALS AND METHODS

Autonomous positioning of the endoscope is carried out by developing an image-based controller. We describe the methodology and the OCT sensor data in this section.

System Overview: The STRAS robotic system consists of a main endoscope, which allows to house three working channels for instruments [1]. The main endoscope is equipped with a camera at the distal tip, a lighting system and, a channel for fluids such as air insufflation and water to cleanse the camera. The distal part of the endoscope can be deflected in two orthogonal directions, which are actuated by antagonist tendons. In total, the

*These authors contributed equally to this work.

This work was supported by the ATLAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813782.

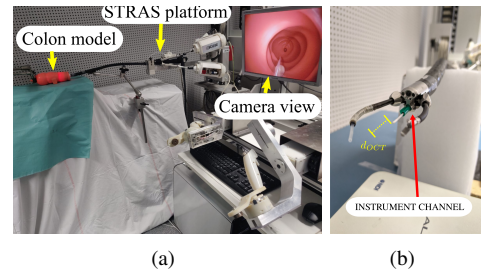


Fig. 1. (a) STRAS robotic setup. (b) Motorised OCT probe.

system has 10 degrees of freedom, 3 per arm (bending, rotation and translation) and 4 on the body (vertical and horizontal bending, rotation and translation). In our setup, a motorised OCT probe is inserted in the central instrument channel and extends 25mm out its distal tip (Fig.1b).

OCT: The probe comprises an external motor, an optical fibre, a distal ball lens, a casing and the associated electronic and optical elements. The ball lens deflects the light passing through the optical fibre, creating an orthogonal imaging beam to the probe. The reflected light is processed to obtain A-line scans with a perception range of 3.5mm outside the probe sheath. The high-speed rotation of the external motor enables the creation of radial images at 50Hz. In this study, we use planar radial B-scans images. The OCT image stream is firstly stabilised with a CNN-based method [2] and then segmented to calculate the distance and direction between the scanning centre and the surrounding tissue.

Polyp Detection: Polyps in the phantom are automatically detected using supervised deep learning techniques. The training set is created by capturing images through telemanipulating the robot. These images (720×576 pixels) are manually annotated by selecting the polyp area from the background; the 100 images collected are augmented to ensure robust training. A U-net [3] is implemented and trained in a supervised manner on the annotated image dataset. Using the output of the trained U-net, the centre of mass of the detected polyp (P_p) is estimated.

Endoscope Orientation: An image-guided position control is developed to reduce the distance error between the image centre P_c and the detected coordinates of polyp P_p . The controller outputs the desired joint position proportional to the error, $E = P_c - P_p$, on both coordinates $E = [E_x, E_y]$. With the horizontal bending q_{horz} reducing

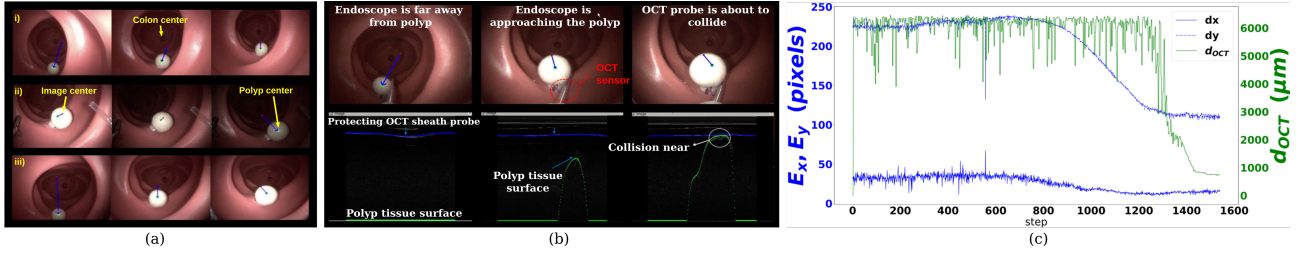


Fig. 2. a) Image guided control. Sequence of frames obtained from the monocular camera while aligning towards the polyp (i) in normal lighting conditions, (ii) in varying lighting conditions, (iii) Experiment two: translating and aligning towards the detected polyp. b) Top: Deployment of the OCT probe. Bottom: Data generated while tissue scanning. c) Plot of evolution of E_x , E_y and d_{OCT} for the second experiment performed.

the E_x component and, q_{vert} acting on E_y . When the error is below $\theta_1 = 60$ pixels and $\theta_2 = 100$ pixels, for E_x , E_y , the orientation control ends.

Endoscope translation: To ensure polyp-probe contact, a small motorised translation of $\sim 5cm$ is carried out at a constant speed. For simplification, the navigation is carried out at a constant speed. The OCT signal is used as feedback to terminate the translation when the tip is close to the polyp or colon wall, i.e. $d_{OCT} < \theta_3$, where d_{OCT} represents the distance measured by the OCT probe and $\theta_3 = 750\mu m$ is the distance threshold.

For the experiments performed, the STRAS robot was teleoperated by the user inside the colon model; the position allows polyp visibility. Then, after a set time, the autonomous technique takes control of the whole-DOF to perform the alignment.

RESULTS

We use the ascending colon part of the LM-107 Colonoscopy Simulator (KOKEN, Japan). The model has a white polyp that can be inserted in a predefined location within the selected section. Our aim is first to show robust polyp detection in varying lighting conditions; second to deploy an OCT probe for tissue scanning. For the first objective, we implement the image position control as described in the previous section, and evaluate detection in changing luminosity conditions. Note that the OCT probe is not employed in these experiments. For the second objective, we deploy the OCT probe and translate near the polyp to scan the tissue. We consider two experimental scenarios as follows (i) align first and then translate towards the polyp; (ii) align towards the detected polyp while translating.

The image position control and translation experiments are illustrated in Fig.2a. The distance from the centre of the image (red dot) to the centre of mass of the detected polyp (green dot) is shown in blue. Fig.2b (bottom) shows the detection of the polyp surface in the OCT image when the endoscope is approaching it, with the corresponding endoscopic image in the top line. The green line in this figure shows the polyp surface, while the blue shows the sheath of the OCT catheter probe. It can be observed that when the endoscope is not aligned with the polyp, both the line remains straight as the polyp remains outside the field of view of the OCT probe. However, while approaching towards polyp, a peak point occurs in the green line showing the presence of an object near

the probe. The peak point continues growing more with endoscope advancement and finally touches the blue line when the probe is about to collide with the polyp. This way, the feedback from the OCT image avoids collision. In Fig. 2c, the results of the second experiment are shown. A 10 second (830 step) safety buffer is set for visual inspection at a 50Hz step rate. After this, pixel error E starts reducing; the d_{OCT} start decreasing after the 1300 step, in which the probe has come close enough to the polyp to provide a measurement with less error.

CONCLUSIONS AND DISCUSSION

The LM-107 colon model used has a 4cm diameter, making navigation with the arms outside the channels prone to collision. A model with a realistic colon diameter, from $\sim 4-8$ cm, would resolve the collision problems and allow for independent arm-body motions [4]. Increasing the complexity of the testbed would require a more robust control system, e.g., the one present in [5] where feedback from the endoscopic camera and OCT is used to actuate the translation and bending intelligently. The prior changes on the setup would require retraining the neural network to ensure polyp detection in real-time, given the background image modification. Polyp dimensions from the image can also serve as feedback on the selected thresholds θ_1 , θ_2 to better suit the geometry of the polyp and OCT probe w.r.t. the camera frame; given dimensions of the polyp does not have any effect in our approach.

Another simplification considered in this work is the constant Jacobian matrix that maps the tendon displacement to camera displacement. This assumption can be valid for small orientation changes; future work will be directed towards estimating the Jacobian matrix online.

REFERENCES

- [1] Nageotte Florent *et al.*, "Stras: A modular and flexible telemanipulated robotic device for intraluminal surgery," in *Handbook of Robotic and Image-Guided Surgery*. Elsevier, 2020, pp. 123–146.
- [2] Liao Guiqiu *et al.*, "Data stream stabilization for optical coherence tomography volumetric scanning," *IEEE Transactions on Medical Robotics and Bionics*, 2021.
- [3] Ronneberger Olaf *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [4] Mora Caravaca Oscar *et al.*, "Oct image-guidance of needle injection for robotized flexible interventional endoscopy," in *Endoscopic Microscopy XVI*, vol. 11620. International Society for Optics and Photonics, 2021, p. 116200D.
- [5] Zhang Zhongkai *et al.*, "Image-guided control of an endoscopic robot for oct path scanning," *IEEE Robotics and Automation Letters*, 2021.

3D Reconstruction of Local Anatomy based on 2D Ultrasound for Fetal Laser Surgery

Yuyu Cai¹, Mirza Awais Ahmad^{1,2}, Mouloud Ourak¹, Ruixuan Li¹, Kenan Niu¹,
Tom Vercauteren³, Sebastien Ourselin³, Jan Deprest², and Emmanuel Vander Poorten¹

¹Robot-Assisted Surgery Group, Department of mechanical engineering, KU Leuven, Belgium

²University Hospital Leuven, Department of Development and Regeneration, Leuven, Belgium

³King's College London, School of Biomedical Engineering and Imaging Science, London, United Kingdom

INTRODUCTION

Twin-to-twin transfusion syndrome (TTTS) increases the risk of disability and death in monochorionic twin pregnancies. Fetoscopic elective laser ablation (ELA) has been shown to be the preferred treatment for TTTS [1]. During the intervention, a surgeon relies on in-utero images from a fetoscope and a 2D-ultrasound (US) scan performed by an assistant. The US view can be used to guide the scope, and hence the laser fiber, to obtain an optimal (90°) ablation angle and distance (approximately 10mm) relative to the placental surface during ablation. In this context, 2D US imaging is somewhat limiting as a 3D model of the visited placental area is needed for efficient in-utero navigation.

A good estimation of the local 3D geometry may decrease the surgeon's difficulty to navigate the fetoscope to the target while satisfying the aforementioned requirements. An intuitive 3D model of the uterus could be reconstructed in real-time provided 2D US images are properly calibrated, segmented and localized in 3D space. So-called local geometric features are essential components for accurate 3D rigid data registration [2].

However, for TTTS surgery, the accuracy of local geometry features should be further investigated. This abstract reports on a fast US surface reconstruction framework for TTTS ELA surgery. Compared to other US surface reconstructions in the clinical application (mean distance error < 5 mm) [3], our work achieved good accuracy (0.7 mm mean distance error). Our work includes three parts: calibration, segmentation, and reconstruction, as Fig. 1. A curved surface and a hemispherical cage were used for experimental validation. The offline reconstruction takes 0.5 ms per image, yielding a total processing time of 75.3 s for the cage model and 106 s for the curved surface model. Representation errors of the reconstructions for the cage model and the curved surface were 0.75 mm and 1.65 mm on average, respectively.

MATERIALS AND METHODS

The experimental setup consists of: (1) a US machine (Sonosite, FUJIFILM, USA) with a 5.2-MHz curvilinear probe; (2) an electromagnetic (EM) tracking system (NDI Aurora System, Northern Digital Inc., Canada); (3) a laptop (Intel i7, CPU @2.8 GHz, 8G RAM) to store and to process all data on-the-fly. Fig.2 shows the setup. The corresponding poses of the individual US images were

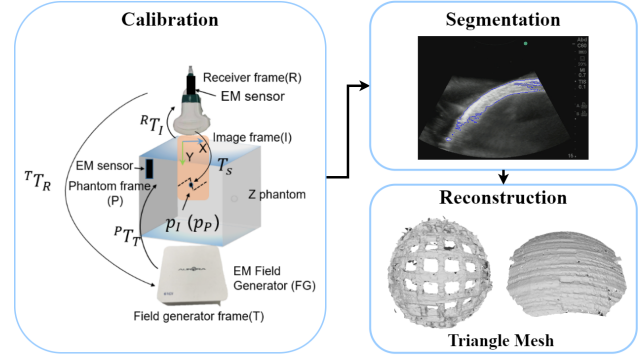


Fig. 1: The reconstruction approach contains three parts: calibration, segmentation, and reconstruction.

measured by an EM sensor rigidly attached to the US probe. The calibration phantom and reconstruction models were placed in a water tank filled with distilled water.

US calibration provides the scale factors and transformation of the scanned object between US images and the real world. The calibration matrix can improve the quality of reconstruction. A $150 \times 100 \times 200$ mm Z-phantom containing three Z-fiducial modules was used for US calibration. Each module was made of three layers of "Z" shaped crossing nylon wires. The Least-Square method and similar triangles theorem were used to derive the spatial relationship between the US images and the EM sensor on the probe holder following the method described by Li *et al.* [4]. Our calibration achieved an accuracy of 0.72 mm and a precision of 0.31 mm.

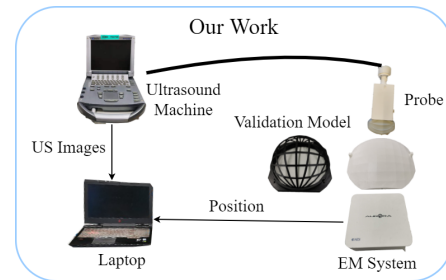


Fig. 2: Overview of experimental setup consisting of an ultrasound machine, laptop and EM tracking system. After acquiring all the US images and positions of the US probe, the reconstructed surface mesh was generated offline.

The boundaries were extracted by three steps: a Gaussian filter, a threshold binarization, and a segmentation step were conducted. In the reconstruction, the segmented pixels were converted from US images into 3D point-cloud by:

$$\begin{bmatrix} x^t \\ y^t \\ z^t \\ 1 \end{bmatrix} = {}^T T_R \cdot {}^R T_I \cdot T_s \cdot \begin{bmatrix} u \\ v \\ 0 \\ 1 \end{bmatrix}, \quad (1)$$

where u, v denote the column and row pixel indexes in the US image; $[x^t \ y^t \ z^t \ 1]^T$ are the positions in the EM tracking system (coordinate frame $\{T\}$). ${}^T T_R$ is the transformation matrix from the sensor on the probe holder $\{R\}$ to the frame $\{T\}$, measured by the EM tracking system. ${}^R T_I$ and T_s were found from the calibration. ${}^R T_I$ represents the spatial relationship from the frame $\{I\}$ (the US image plane) to the frame $\{R\}$. Matrix T_s is a scaling matrix that converts the pixel indices of the US image to SI units (in mm). After point-cloud reconstruction, the surface mesh was obtained through the ball pivoting algorithm [5]. Two 3D print models (Fig 2) have been designed to validate the ability to reproduce local geometry features in the 3D reconstruction. To validate the US reconstruction for fetal surgery, a curved surface and a customized cage with known geometry were designed to mimic the curved shape of the uterine boundary.

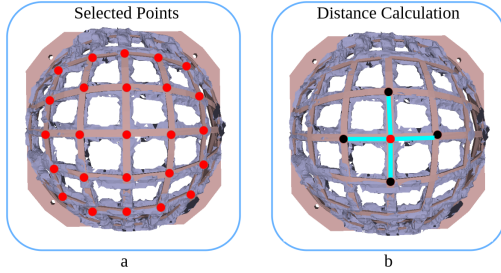


Fig. 3: Results: in red ground truth model, in blue reconstructed mesh; a) shows in red the selected points at the intersection between ribs; b) shows distance calculation between neighbouring points (cyan lines).

The iterative closest points (ICP) algorithm [6] was used to register the generated point cloud with the CAD model. The 3D representation error is the distance between corresponding point pairs of reconstruction and CAD model after registration. To better evaluate the local geometrical feature of the cage model, the key feature points were identified manually (marked red in Fig. 3.a). The 25 selected points are the center of the curved ribs intersection area. Per point four adjacent points were identified (marked black in Fig. 3.b). The distances between each red point and the 4 neighboring points were calculated and compared to the ground-truth from CAD. In total, distances for the selected points have been evaluated. Table II reports the mean and SD of the errors.

RESULTS

The reconstruction experiments were repeated 3 times per model. Table I summarizes the representation errors

TABLE I: Results of the representation error (mm)

Experiments	Hemisphere cage	Curved surface
1	0.37	1.77
2	0.85	1.40
3	1.02	1.79
average	0.75	1.65

TABLE II: Results of the local distance error (mm)

Experiments	Mean	Standard deviation (SD)
1	0.69	0.65
2	0.79	0.59
3	0.63	0.45
average	0.70	0.56

of the reconstructions of both the curved and cage surface (0.75 mm and 1.65 mm on average). Note that the size of the reconstructed curved surface was $183 \times 185 \times 41.5 \text{ mm}^3$, close to the ground truth size $184 \times 184 \times 41.5 \text{ mm}^3$. The entire reconstruction took less than 2 min for each model. The image processing time for every image was 0.5 ms using Anaconda's numba library.

CONCLUSIONS

This paper focuses on the capability to reconstruct local geometry features from US for TTTS ELA surgery. As ELA surgery requires a certain angle and distance between the fetoscope and the placenta, communicating such local features could be helpful to guide surgeons. The presented work offered representation errors of 1.65mm and 0.75mm with mean and SD of local distance errors of 0.70mm and 0.56mm, respectively. In combination with robotic control the reconstruction and navigation assistance could operate in a more autonomous fashion in the future. The future work will therefore need to consider issues related to non-rigid deformation of soft tissue.

ACKNOWLEDGEMENT

Yuyu Cai was supported by the China Scholarship Council. This project received funding from FLEXFET, KU Leuven funded C2 project and a grant from Wellcome Trust (WT101957) and the Engineering and Physical Sciences Research Council (EPSRC) (NS/A000027/1).

REFERENCES

- [1] L. Van Der Veeken *et al.*, "Laser for twin-to-twin transfusion syndrome: a guide for endoscopic surgeons," *Facts, views & vision in ObGyn*, vol. 11, no. 3, p. 197, 2019.
- [2] Jiaqi Yang *et al.*, "Evaluating local geometric feature representations for 3d rigid data matching," *IEEE Transactions on Image Processing*, vol. 29, pp. 2522–2535, 2019.
- [3] Duc V. Nguyen *et al.*, "Validation of 3d surface reconstruction of vertebrae and spinal column using 3d ultrasound data – a pilot study," *Medical Engineering & Physics*, vol. 37, no. 2, pp. 239–244, 2015.
- [4] Li Ruixuan *et al.*, "Comparative quantitative analysis of robotic ultrasound image calibration methods," *20th International Conference on Advanced Robotics*, 2021.
- [5] Fausto Bernardini *et al.*, "The ball-pivoting algorithm for surface reconstruction," *IEEE transactions on visualization and computer graphics*, vol. 5, no. 4, pp. 349–359, 1999.
- [6] Szymon Rusinkiewicz *et al.*, "Efficient variants of the icp algorithm," in *Proceedings third international conference on 3-D digital imaging and modeling*. IEEE, 2001, pp. 145–152.

Unsupervised Denoise and Deblur of OCT Images via Disentangled Representation

Shunlei Li, Muhammad Adeel Azam, Ajay Gunalan, Alessandro Casella, and Leonardo S. Mattos

Department of Advanced Robotics, Istituto Italiano di Tecnologia
shunlei.li@iit.it

INTRODUCTION

Optical coherence tomography (OCT) is a rapidly evolving imaging technology that combines a broadband, low coherent light source with interferometry and signal processing to produce high-resolution images of living tissues. Most OCT devices used in clinical studies have a resolution of approximately $10 \mu\text{m}$ and a depth of penetration of up to 2 mm in most soft tissues. However, the speckle noise introduced by the low coherence interferometry imaging process and blur from device motion significantly degrades imaging quality. This has a strong impact on subsequent analysis and makes clinical application challenging. Therefore, efficient OCT image enhancement methods are urgently required.

By improving the light source, hardware-based approaches reduce the noise of the detector and scanner to some extent, but the speckle or white noise in the imaging system cannot be eliminated. Software-based approaches such as non-local means or block-matching and 3D filtering (BM3D) can provide good results, but need laborious efforts of parameter tuning for different noise levels [1]. Also convolutional neural networks (CNNs) are a potential solution for such image enhancement task, but supervised learning methods are laborious in terms of training data annotation and the use of standard CNNs may lead to loss of details due to averaging processes [2].

A promising solution for image denoising is Generative Adversarial Network (GAN), in which two neural networks compete with each other in a game. There have been many recent efforts on learning disentangled representations to do denoise and deblur, which breaks down, or disentangles, each feature into narrowly defined variables and encodes them as separate dimensions [3], [4]. However, these methods have only been applied separately. In order to remove blur and speckle noise in OCT image with only one generator, we proposed an unsupervised learning method based on GAN and disentangled representation.

MATERIALS AND METHODS

The proposed method consists of three parts: 1) content encoders E^c , and feature encoders E^b, E^n for blur and noise; 2) blurred, noisy, and blurred noisy image generators G^b, G^n, G^{bn} , and clean image generator G^c ; 3) blurred, noisy, blur-noise, and clean image discriminators D^b, D^n, D^{bn}, D^c . Given an input blur-noise data X and unpaired clean data Y , the content encoder E^c can

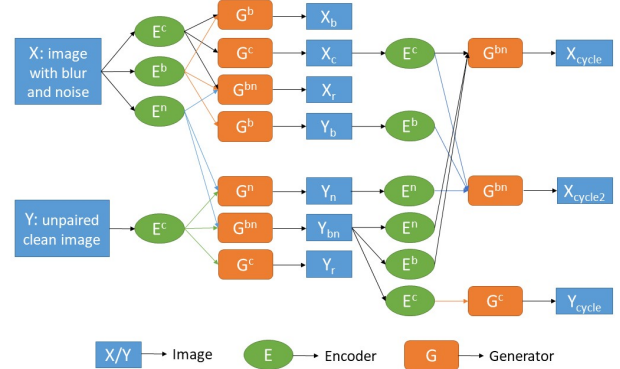


Fig. 1: Workflow of proposed image enhancement method. X and Y are inputs, subscript of b, c, r, n, bn and $cycle$ is blurred, clean, reconstructed, noisy, blurred and noisy, and cycled.

extract content information from corresponding samples, and E^b, E^n can estimate the feature information from X . G^b, G^n, G^{bn} then take features and content information to generate corresponding images and G^c generates clean image. The discriminators distinguish between the real and generated images. The framework is illustrated in Fig. 1. To do image enhancement on unpaired data, we use disentanglement to generate clean images and decode their features. For the blur domain, with content information of E^c and blur feature of E^b , the generated blur images guide the encoders E^c towards extracting content information from the blurred images. Similarly, generating and then distinguishing noisy images from clean ones guides E^c towards extracting content from noisy images. After training the model and addressing disentanglement, clean images can be obtained from using E^c and G^c .

Our proposed loss function includes five components:

(1) Domain adversarial loss includes content information loss L_{adv}^c , blur feature loss L_{adv}^b , noise feature loss L_{adv}^n , and blur-noise features loss L_{adv}^{bn} . In order to get better blur and noise features, we pretrained deblur and denoise module based on GAN with disentangled representations [3] and shared last layer of generators to G_b and G_n . The domain adversarial loss is defined as:

$$L_{adv} = \underset{E, G}{\operatorname{argminmax}} (L_{adv}^c + L_{adv}^b + L_{adv}^n + L_{adv}^{bn}) \quad (1)$$

(2) Cycle consistency loss:

$$L_{cycle} = \mathbb{E}[\|X - X_{cycle}\|_1] + \mathbb{E}[\|Y - Y_{cycle}\|_1] \quad (2)$$

(3) Reconstruction loss:

$$L_{recon} = \mathbb{E}[\|X - X_r\|_1] + \mathbb{E}[\|Y - Y_r\|_1] \quad (3)$$

(4) Noise patch loss: According to L_{noise}^X and L_{noise}^Y from noise patch discriminator, we can get noise patch loss as:

$$L_{noise} = \underset{E, G}{\operatorname{argminmax}} (L_{noise}^X + L_{noise}^Y) \quad (4)$$

(5) *KL* divergence loss is used to regularize the blur feature distribution to get close to normal distribution. *KL* divergence is minimized to obtain the *KL* loss:

$$L_{KL} = \frac{1}{2} \sum_{i=1}^N (\mu_i^2 + \sigma_i^2 - \log(\sigma_i^2) - 1) \quad (5)$$

Consideringe the equations above, the main loss function can be written as:

$$L = \lambda_{adv} L_{adv} + \lambda_{cycle} L_{cycle} + \lambda_{recon} L_{recon} + \lambda_{noise} L_{noise} + \lambda_{KL} L_{KL} \quad (6)$$

The subscripted λ are the coefficients of each corresponding loss. The architecture of the encoders and generators are based on DRGAN [4].

RESULTS

The proposed method was evaluated on a custom dataset obtained from pork vocals with a commercial OCT device (Thorlabs TEL320C1). After augmentation, the dataset included 15600 B-Scan images with resolution of 1024×1736 pixels (*height* \times *width*).

TABLE I: Ablation study

Method	Metrics	
	PSNR	SSIM
Original image	5.61	0.10
Denoise module	20.61	0.66
Deblur module	16.54	0.59
Proposed method	28.67	0.81

To validate the effects of each module in our method, an ablation study was performed (see Fig. 2). Table I presents the Peak signal-to-noise ratio (PSNR) and the Structured Similarity Indexing Method (SSIM) of: original OCT image with blur and noise, blur image obtained from denoise module, noise image obtained from deblur module, and image enhanced with the proposed method. Table II summarizes the results of PSNR and SSIM for BM3D, DeSpecNet (CNN method), SiameseGAN, DRGAN and proposed method.

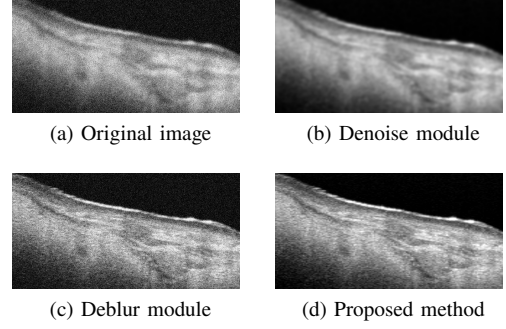


Fig. 2: Ablation study to evaluate each module and proposed method.

TABLE II: Comparison of image enhancement methods

Method	Metrics	
	PSNR	SSIM
BM3D	18.94	0.652
DeSpecNet[5]	22.57	0.713
SiameseGAN[6]	27.41	0.757
DRGAN[4]	26.11	0.772
Proposed	28.67	0.812

CONCLUSIONS AND DISCUSSION

We proposed a novel combined denoise and deblur method for OCT images via disentangled representation, which demonstrated promising results in terms of PSNR and SSIM. To the best of our knowledge, this is the first work proposing OCT image denoise and deblur based on a single processing step. For this, we used an unsupervised learning approach and performed noise, blur and content disentanglement with corresponding encoders and generators. In the future, we will further optimize this method for clinical application and evaluate the image enhancement performance on extended datasets from open sources and other OCT devices.

REFERENCES

- [1] B. Chong and Y.-K. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified bm3d filter," *Optics Communications*, vol. 291, pp. 461–469, 2013.
- [2] Y. Guo, K. Wang, S. Yang, Y. Wang, P. Gao, G. Xie, C. Lv, and B. Lv, "Structure-aware noise reduction generative adversarial network for optical coherence tomography image," in *International Workshop on Ophthalmic Medical Image Analysis*. Springer, 2019, pp. 9–17.
- [3] B. Lu, J.-C. Chen, and R. Chellappa, "Unsupervised domain-specific deblurring via disentangled representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 225–10 234.
- [4] Y. Huang, W. Xia, Z. Lu, Y. Liu, H. Chen, J. Zhou, L. Fang, and Y. Zhang, "Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2020.
- [5] F. Shi, N. Cai, Y. Gu, D. Hu, Y. Ma, Y. Chen, and X. Chen, "Despecnet: a cnn-based method for speckle reduction in retinal optical coherence tomography images," *Physics in Medicine & Biology*, vol. 64, no. 17, p. 175010, 2019.
- [6] N. A. Kande, R. Dakhane, A. Dukkupati, and P. K. Yalavarthy, "Siamesegans: a generative model for denoising of spectral domain optical coherence tomography images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 180–192, 2020.

Study of the curling behaviour of a self-shaping cochlear implant during insertion using a teleoperated robotic system

D. Bautista-Salinas^{1,2}, C. Kirby^{1,2}, M. E. M. K. Abdelaziz¹, B. Temelkuran¹,
C. T. Huins³, and F. Rodriguez y Baena^{1,2}

¹The Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, Imperial College London, UK

²Mechatronics in Medicine Lab, Department of Mechanical Engineering, Imperial College London, UK

³Queen Elizabeth Hospital Birmingham, UK

INTRODUCTION

A cochlear implant (CI) is currently the most effective treatment for people suffering from sensorineural hearing loss. They work by converting sound into electrical pulses, which are sent, through an electrode array (EA), to the cochlea. The EAs found in commercial CIs can be distinguished depending on their position within the cochlea. A lateral wall electrode array (LEA) sits close to the outer wall of the scala tympani (ST), the helical chamber of the cochlea where the EA is inserted. A perimodiolar electrode array (PEA) sits close to the inner wall of the ST. Both commercial LEAs and PEAs can generate intracochlear trauma [1], leading to residual hearing loss, which itself important to enhance the CI hearing outcomes for the patient. As an alternative, we proposed a PEA self-shaping CI that can generate contact forces below the rupture threshold and that has the potential to overcome the limitations of current devices [2]. The insertion procedure has to be studied and optimised to maximise the benefits derived from this implant. It is known that the additional skill required to insert PEAs results in worse reported outcomes [3]. Therefore, in this work, we present a teleoperated robot that we will use to study the recovery behaviour of our self-shaping EA.

MATERIALS AND METHODS

Self-Shaping implant. A thermally responsive shape memory polymer (SMP) can recover its original shape from a temporary one, achieved after deformation and fixation when under an external temperature stimulus (e.g. body temperature). To fabricate the self-shaping EA, we start from SMP pellets (SMP Technologies Inc., Japan) to create a cylindrical preform with the final device cross-section. The cross-section has an inner lumen that can be used for sensing and therapeutics, such as the corticoids used in clinical practice. The preform was used to obtain filaments of the desired EA diameter following a thermal drawing process; filaments were then moulded into a 2D shape of a 3x scaled-up version of the ST. The implant was moulded at 80°C for 15 minutes, using a CNC machined Teflon mould (Proto Labs Ltd., UK), and after cooling, it was attached to a fixture that was used to connect the implant to the robot. Then, the implant was programmed

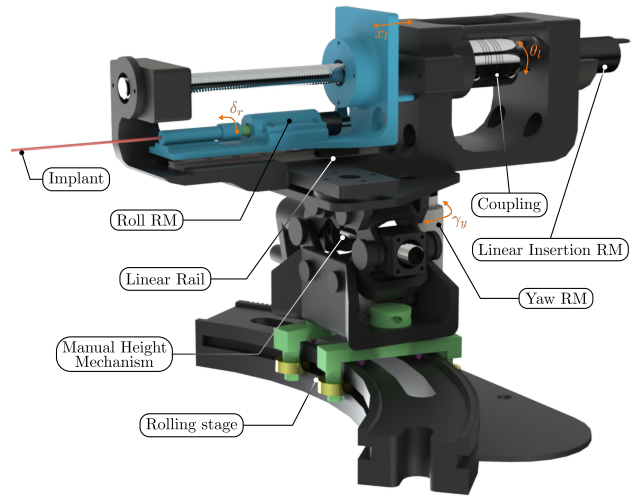


Fig. 1: A rendering of the proposed robot showing the key components. RM, rotary motor

into a straight shape. We warmed it up by placing it in water slightly above its transition temperature (35 °C) for 30 s. Immediately after, the implant was placed in a 4°C bath for another 30 s to fix its temporal position.

Robotic system design. A robot with 4 degrees of freedom was developed to manipulate our implant in a master-slave fashion. The robot design is compact to minimise its footprint in the operating theatre and lightweight to facilitate its manipulation. The linear insertion motion is achieved with a lead screw actuated with a brushless DC-Servomotor (2057S 024B K1155, Faulhaber, Germany). This motor can generate speeds that minimise trauma, and it uses analogue hall sensors, which provide a lightweight and compact alternative to encoders. The insertion device is attached to the lead screw and sits on a linear slider (Misumi, Germany) that acts as an anti-rotation feature and helps minimise friction. The motion is smooth with a total stroke length of 77 mm, but an inherent problem of lead screws is their backlash that is solved here by preloading the system manually before each insertion. The insertion tool has a second brushless DC-Servomotor with analogue hall sensors (1226A 012B K1855, Faulhaber, Germany) combined with a planetary gearhead with a

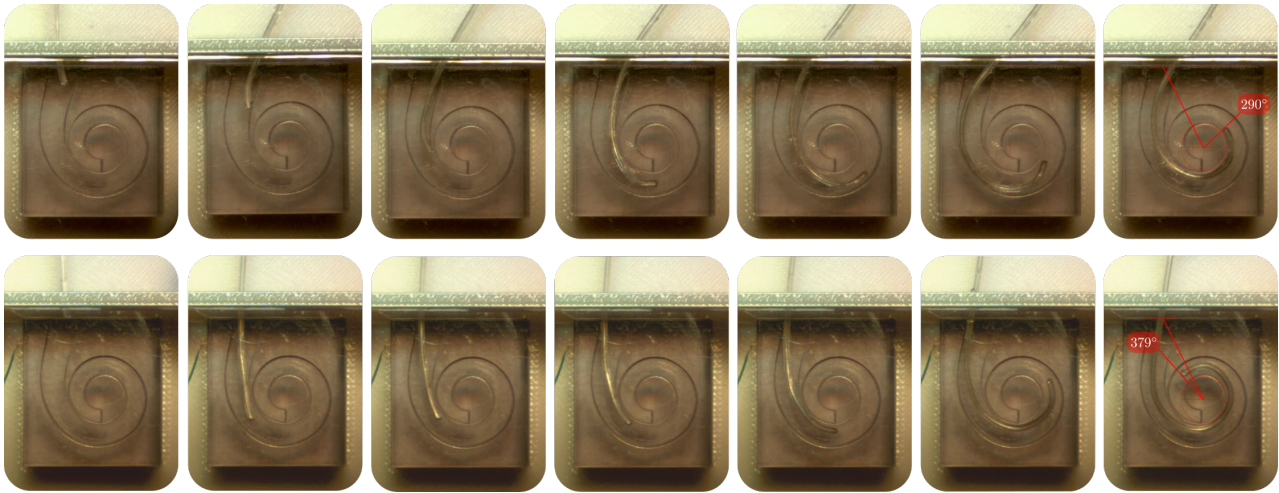


Fig. 2: Snapshots showing the self-shaping cochlear implant following the insertion approach 1 (top) and 2 (bottom)

reduction ratio of 16:1 (12/4, Faulhaber, Germany). This motor allows to accurately control the implant roll, which is necessary to avoid touching the cochlear walls and translocation from the ST to the SV. A fixture is pressed fit into the motor shaft to load the implants. The yaw motion is actuated via a custom brushless DC-Servomotor motor with analogue hall sensors (Maxon Group, Switzerland), which can accurately control the insertion approach angle. The height can be adjusted manually. A rendering of the robot is shown in Fig. 1

Robotic system control. To control the Faulhaber motors we used 2 EtherCAT controllers (MC 5004 P STO, Faulhaber, Germany). A serial EPOS2 Module 36/2 was used to control the Maxon motor. The linear motion was velocity-controlled to maintain safe speeds, while the roll and yaw were position-controlled to get the desired angles with minimum error. The controllers were programmed in C++ on a Linux PC. Teleoperation was performed in a master-slave configuration with a keyboard controller as the master device.

RESULTS

The experiment was carried out using a 3x scaled-up planar model of the cochlea adapted from an open-source design [4]. The model was enclosed into a chamber where the body temperature was simulated. A surgical microscope was used to visualise and assess the insertion process. An additional digital camera (Thorlab Inc., USA) was placed perpendicular to the microscope to visualise the implant roll. Before insertion, the implant was kept at 5°C, and after loading the device into the robot, the operator started the insertion immediately. We tested two different insertion approaches. In the first one, the operator would approach the cochlea entry point and then insert the device progressively. In the second one, the operator was instructed to follow an approach similar to the Advance Off-Stylet® (AOS), where the implant has to be inserted rapidly until the first turn to avoid curling and then progressively until the device is fully inserted.

The insertion time was similar in both cases: 59 s and 70 s, respectively. To validate the insertion, we measured the depth insertion angle, which is considered successful for depth angles between 330° and 390° [5]. The self-shaping implant was inserted up to a depth angle of 290° and 379°. Therefore, only approach 2 can be considered successful for insertion depth. It was also noted that the contacts with the cochlear walls (identified by visual inspection) were significantly minimised for the second approach, as shown in Fig. 2.

CONCLUSION AND DISCUSSION

The self-shaping CI used in this work had previously shown the ability to generate normal forces below the rupture threshold. However, its behaviour during insertion had not been studied. In this work, we have used a teleoperated robotic system to study the best approach to insert such an implant. The experiment shows that an AoS insertion technique has the potential to minimise contacts with walls as well as improve insertion depth. Future steps will focus on studying this technique in-depth, understanding how recovery forces evolve during insertion, and investigating different modes to operate the robotic system. We will also explore different sensing approaches to retrieve the position of the implant within the cochlea in a realistic way.

REFERENCES

- [1] F. Risi, "Considerations and Rationale for Cochlear Implant Electrode Design - Past, Present and Future," *The Journal of International Advanced Otolaryngology*, vol. 14, no. 3, pp. 382–391, 2018.
- [2] D. Bautista-Salinas, M. E. M. K. Abdelaziz, B. Temelkuran, E. M. Yeatman, C. T. Huins, and F. Rodriguez y Baena, "Towards a Functional Atraumatic Self-Shaping Cochlear Implant," *Macromolecular Materials and Engineering*, vol. 2100620, p. 2100620, 2021.
- [3] B. P. O'Connell, J. B. Hunter, and G. B. Wanna, "The importance of electrode location in cochlear implantation," *Laryngoscope Investigative Otolaryngology*, vol. 1, no. 6, pp. 169–174, 2016.
- [4] J. R. Clark, F. M. Warren, and J. J. Abbott, "A Scalable Model for Human Scala-Tympani Phantoms," *Journal of Medical Devices*, vol. 5, no. 1, p. 014501, 2011.
- [5] T. Stöver, P. Issing, G. Graurock, P. Erfurt, Y. ElBeltagy, G. Paasche, and T. Lenarz, "Evaluation of the Advance Off-Stylet insertion technique and the cochlear insertion tool in temporal bones," *Otology and Neurotology*, vol. 26, no. 6, pp. 1161–1170, 2005.

Estimation of a Vessel Side-Branches Model for Robotic Intravascular Navigation

Beatriz Farola Barata^{1,2,*}, Wim-Alexander Beckers^{1,*}, Gianni Borghesan^{1,3}, Diego Dall'Alba², Paolo Fiorini², Jos Vander Sloten¹, and Emmanuel Vander Poorten¹

¹Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Leuven 3001, Belgium

²Altair Robotics Laboratory, Department of Computer Science, University of Verona, Verona 37134, Italy

³Core Lab ROB, Flanders Make, Belgium

*These authors contributed equally to this work (Corresponding author: beatriz.barata@kuleuven.be)

INTRODUCTION

Endovascular catheterization is a complex minimally invasive procedure which requires external imaging for intra-operative visual guidance. Fluoroscopy is the commonly chosen imaging modality. Yet, it is an X-ray based modality, limited to two-dimensional views of the imaged scene. Consequently, endovascular interventions still remain challenging and improved visual feedback would thus be advantageous [1]. Intra-operative modelling of blood vessels using local information from sensors embedded in the catheter has the potential to greatly improve surgical robots' or clinicians' awareness during navigation, while reducing the use of X-ray radiation [2].

In a previous work [3], a method for constructing a real-time intra-operative 3D approximation of the vessel geometry around a catheter tip was proposed. Intravascular Ultrasound (IVUS) imaging and Electromagnetic (EM) pose sensing were used together with an Unscented Kalman Filter (UKF) to estimate the best fitting local cylinder. The method showed good approximation capability of the vessel geometry, but in the absence of side branches. Nonetheless, the modelling of side branches is imperative as in some cases, the catheter must be steered into them to reach the desired target site. Additionally, they could serve as important navigation landmarks.

In this work, a new vessel model is proposed to allow estimating and representing side branches. In particular, as shown in Fig. 1, the side branches are modeled as holes in the local cylinder model that represents the main vessel (MV). Also, an UKF is implemented for inferring the parameters of the cylinder holes from synthetic IVUS and EM data. The proposed method is tested in a virtual vessel model making use of simulated IVUS and EM data.

MATERIALS AND METHODS

Two UKFs are used to: (1) locally approximate the MV by a cylinder around the catheter tip, as described in [3] and (2) approximate vessel side branches as holes in the aforementioned cylinder model.

Side-branches modelling: A side-branch is represented by a hole characterized by three variables with respect to the cylinder frame $\{cyl\}$, as shown in Fig. 2a: *i*) an angle

This work was supported by the ATLAS project. The ATLAS project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813782.

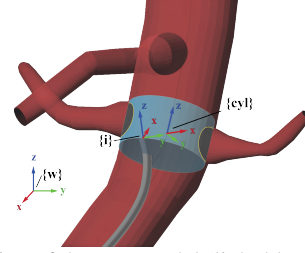


Fig. 1: Illustration of the MV model (light blue cylinder) and the side-branches models (holes in the cylinder, outlined in yellow). The proposed method is used to estimate the hole(s) state.

$\alpha \in [0, 2\pi)$ in the xy plane of the cylinder perpendicular to its z -axis, *ii*) a displacement dz of the hole along the z -axis of the cylinder, and *iii*) a radius r . For each side branch j among the B detected side branches around the catheter tip ($j = 1, \dots, B$), the hole parameters α_j , dz_j and r_j are added to the state vector to estimate, here defined as:

$$\mathbf{x}_k = [\alpha_1 \ dz_1 \ r_1] \cdots [\alpha_B \ dz_B \ r_B]^T. \quad (1)$$

Side-branch detection: In the case at hand, synthetic IVUS data is directly generated as vessel contour points by intersecting the xy plane of the IVUS frame $\{i\}$ (see Fig. 1) with the simulated vessel. For each extracted contour, side-branch detection is carried out. A side branch is considered detected when *i*) the distance between the center of the side-branch ostium and its projection on the xy plane of $\{i\}$ is smaller than the side branch radius; and *ii*) the distance of the IVUS probe to the projected side branch ostium's center is smaller than a user-defined value (e.g. 20 mm) (see Fig. 2b). Note that in simulation, both the side-branch ostium center and the radius are known from the mesh geometry. After this step and the estimation

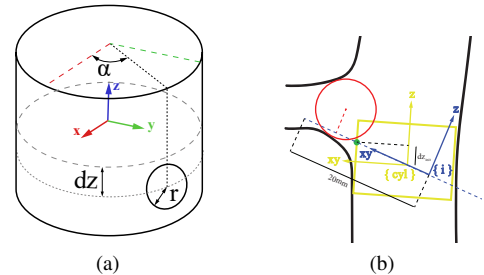


Fig. 2: (a) Overview of the three variables that characterize a hole state in the MV cylinder model. α and dz determine the position of the side-branch hole in the cylinder, while r determines its radius. (b) Depiction of the side branch detection and initial dz -value calculation strategies. The red circle represents the side branch ostium characterized by a center location and a radius; the yellow rectangle represents the MV cylinder model.

of the local MV cylinder model, the hole(s) state estimation is performed. From the side branch detection strategy, the initial α -values are obtained by projecting the branch center point onto the xy -plane of $\{i\}$ and then determining, on the same plane, the angle between the origin of $\{i\}$ and the projected point. Similarly, the initial dz -values are calculated first, by finding the intersection point with the MV estimated cylinder and a line along α (defined in the xy -plane of $\{i\}$) and second, by taking the distance from the cylinder position (origin) to the intersection point projected onto the cylinder z (longitudinal) axis. A representation of how these initial values are determined is shown in Fig. 2b. Finally, considering that a side branch is stationary relative to the moving local cylinder modelling the MV, a side branch is deleted from the state vector whenever dz exceeds a given threshold. This threshold is computed based on the height of the cylinder model from the MV and it is an indication that the side branch is no longer close to the catheter tip.

Filtering for side-branches model estimation: In the UKF implementation hereby described, the hole(s) state is first predicted based on the transformation of the MV approximated cylinder between the current and the previous frames, and then updated by considering the contour points of the vessel lumen and side branches in the xy -plane of the IVUS probe. Both α and dz at time $k - 1$ are expressed in the new frame $\{cyl\}_k$ by, respectively, considering the rotation and the translation of the cylinder around its z -axis between $k - 1$ and k , as described in

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{u}(\mathbf{e}, \mathbf{f}, \mathbf{R}, \mathbf{p}), \quad \text{where}$$

$$\mathbf{u} = \begin{bmatrix} -atan2((i_k \mathbf{R}^{i_{k-1}} \mathbf{f})_{i_k} \mathbf{e}, (i_k \mathbf{R}^{i_{k-1}} i_{k-1} \mathbf{e})_{i_k} \mathbf{e}) \\ -(w_k \mathbf{p} - w_{k-1} \mathbf{p}) \cdot w_{k-1} \mathbf{d} \\ 0 \end{bmatrix}. \quad (2)$$

$i_k \mathbf{e}$ and $i_k \mathbf{f}$ are unit vectors along the cylinder x and y axes, respectively, at time k expressed in the IVUS frame $\{i\}$; $i_k \mathbf{R}^{i_{k-1}}$ is the rotation matrix of the IVUS frame $\{i\}$ between time $k - 1$ and k ; $w_k \mathbf{p}$ is the cylinder position in the world frame $\{w\}$ at time k ; and $w_{k-1} \mathbf{d}$ is the direction of the cylinder in frame $\{w\}$ at time $k - 1$.

In order to describe the observation function $\mathbf{h}(\mathbf{x}_k)$ output, the cylinder with holes is intersected with the xy -plane of $\{i\}$. First, the distances of the origin of $\{i\}$ to M evenly spaced (every $\frac{2\pi}{M-1}$ radians) intersection points $i_k \tilde{\mathbf{c}}(k = 1, \dots, M = 10)$ of the xy -plane of $\{i\}$ with an infinite cylinder are determined. Second, the distances of the intersection points S within the side-branch section (subset of M) are adjusted by a user-defined factor of 1.3:

$$\mathbf{h}(\mathbf{x}_k) = [||i_k \tilde{\mathbf{c}}|| \dots [1.3 \cdot ||i_k \tilde{\mathbf{c}}|| \dots 1.3 \cdot ||i_k^S \tilde{\mathbf{c}}||] \dots ||i_k^M \tilde{\mathbf{c}}||]^T \quad (3)$$

RESULTS

An experiment was carried out with a virtual catheter advancing inside a simulated aortic model with 4 side branches (see an example with 3 side branches in Fig. 1). **Experimental setup (simulation):** The catheter was steered repeatedly by translating it 10 mm forward at 2 mm/s, followed by a series of bending motions at 5%:

i) 36° bending of the catheter in one plane, *ii)* 360° rotation of the bending plane and *iii)* -36° bending in the bending plane, returning to the original orientation. For more realistic conditions, zero mean Gaussian noise with a standard deviation of 0.3 mm and 0.5° was added to the translation and rotation components of the catheter EM data, respectively; and zero mean Gaussian noise with a standard deviation of 1 mm was added to the IVUS data. The IVUS probe, the EM sensor and the catheter tip were assumed to be aligned by construction. Also, Gaussian noise with zero mean and 5° standard deviation was added to the α -values obtained from side-branch detection.

TABLE I: UKF parameters for side-branches model estimation

\mathbf{x}_0	$[\alpha_{\text{detected}} \ dz_{\text{detected}} \ 4 \text{ mm}]$
$\mathbf{P}_{\mathbf{x}_0}$	$\text{diag}([\frac{\pi}{18} \ 5 \text{ mm} \ 0.1 \text{ mm}])^2$
$\mathbf{P}_{\mathbf{v}_k}$	$\text{diag}([\frac{\pi}{180} \ 0.5 \text{ mm} \ 0.1 \text{ mm}])^2$
$\mathbf{P}_{\mathbf{n}_k}$	$\text{diag}([4 \text{ mm} \ \dots \ 4 \text{ mm}])^2$
$[\kappa \ \alpha \ \beta]$	$[0 \ 1 \ 2]$

Hole estimation evaluation: The filter performance was evaluated by means of two error metrics, computed at each simulation step: *i)* the distance error of the approximated side-branch representation and the ground-truth side-branch model (hole), and *ii)* the difference between the estimated radius and the ground-truth radius. The errors progression for each side branch and the errors distribution for all branches, including the median and interquartile range values, are shown in Fig. 3.

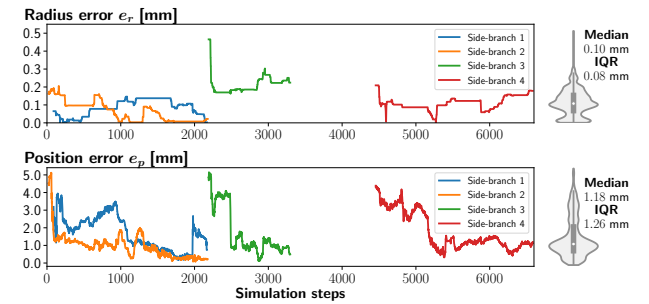


Fig. 3: Position and radius errors progression (left) and distribution (right) of the different side-branches models.

CONCLUSIONS AND DISCUSSION

The proposed method aims at determining reliable 3D vessels including side branches by expanding a prior cylinder model. From the low medians and IQRs shown in Fig. 3 (Radius: 0.10 mm and 0.08 mm; Position: 1.18 mm and 1.26 mm, respectively), the obtained results demonstrate the potential of using holes to model the ostium of side-branches. Future work is planned to compare the hole model with more detailed side-branch models and to conduct *in vitro* validation with real IVUS and EM data.

REFERENCES

- [1] H. Rafii-Tari *et al.*, “Current and emerging robot-assisted endovascular catheterization technologies: A review,” *Annals of Biomedical Engineering*, vol. 42, no. 4, pp. 697–715, 2014.
- [2] E. Vander Poorten *et al.*, “Cognitive autonomous catheters operating in dynamic environments,” *Journal of Medical Robotics Research*, vol. 1, no. 3, pp. 1–25, 2016.
- [3] B. Farola Barata *et al.*, “Ivus-based local vessel estimation for robotic intravascular navigation,” *IEEE robotics and automation letters*, vol. 6, no. 4, pp. 8102–8109, 2021.

One-Shot Boundary Detection Network for Multi-Modal Side-Viewing Imaging

Beatriz Farola Barata^{1,2,*}, Guiqiu Liao^{2,3,*}, Diego Dall’Alba², Gianni Borghesan^{1,4}, Benoit Rosa³, Michel de Mathelin³, Florent Nageotte³, Paolo Fiorini², Michalina J. Gora³, Jos Vander Sloten¹, and Emmanuel Vander Poorten¹

¹Robot-Assisted Surgery Group, Department of Mechanical Engineering, KU Leuven, Leuven 3001, Belgium

²Altair Robotics Laboratory, Department of Computer Science, University of Verona, Verona 37134, Italy

³ICube Laboratory, University of Strasbourg, Strasbourg, France

⁴Core Lab ROB, Flanders Make, Belgium

*These authors contributed equally to this work (Corresponding author: beatriz.barata@kuleuven.be)

INTRODUCTION

Nowadays, catheter-based imaging systems are increasingly used in a variety of clinical applications in order to obtain side-viewing luminal and transmural images. Mainstream side-viewing catheters often use Intravascular Ultrasound (IVUS) or Optical Coherence Tomography (OCT) to acquire cross-sectional views of the intraluminal environment. IVUS is commonly used for imaging intravascular pathologies such as aneurysms or atherosclerotic plaque [1]. Side-viewing OCT can also be used to visualize vascular structures, as well as larger lumens, such as the colon or the respiratory lung airways. Both imaging modalities have their characteristics: IVUS can see through blood, but experiences problems seeing through air. OCT does not have that problem, but it relies on injection of saline solution or contact with tissue for imaging in blood. The resolution of OCT is typically higher than of IVUS whereas IVUS has a larger image depth. Despite their differences, in side-viewing catheters, both methods tend to produce images with a similar appearance. Automatic segmentation of both OCT and IVUS is an appealing feature for supporting real-time diagnosis or offline image analysis. However, the resemblance between these two imaging modalities has not yet been exploited to produce robust tissue segmentation algorithms. Traditional methods for segmentation of OCT or IVUS use hand-crafted edge detection filters, mathematical morphology, Otsu’s automatic thresholding, intersection of radial lines with lumen boundaries, Markov-random fields and light back-scattering [2].

This work proposes a novel deep Convolutional Neural Network (CNN) architecture, shown in Fig. 1, based on explicit coordinate encoding networks designed for multi-modal image processing. Aside from improved contour segmentation efficiency, the proposed approach directly provides the relative distance between tissue surface and scanning center. The proposed architecture directly encodes in a coordinates vector the position of the detected object contours in the polar domain, without pix-wise segmentation nets [3] or additional detection networks. Moreover, the same architecture is applied to OCT and

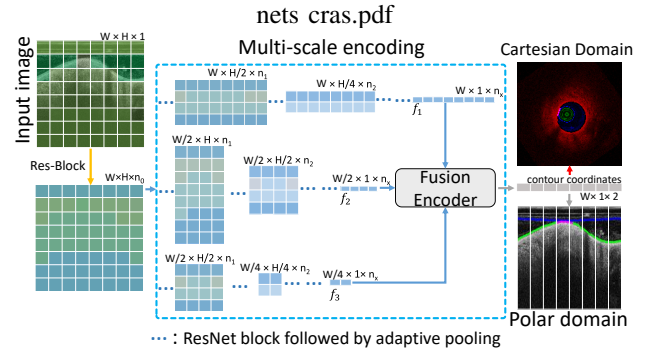


Fig. 1: Proposed network architecture featuring a parallel multi-scale encoding scheme.

IVUS images for the task of lumen segmentation and then compared with the state-of-the-art U-net architecture [3].

MATERIALS AND METHODS

Inspired by shape encoding, the proposed network architecture predicts the contour coordinates of surgical objects and/or intraluminal structures. In particular, the final boundary coordinates prediction occurs in one shot and does not rely on detection nets or segmentation nets as backbone. The proposed method, shown in Fig. 1, first adapts an initial shallow Resnet block to produce raw features with the same dimension as the 2D input image (orange arrow in Fig. 1). Then, it follows a parallel multi-scale encoding scheme (blue dashed box in Fig. 1). Adaptive pooling modules, after identity Resnet blocks, extract multi-scale hierarchical coordinates position descriptors which potentially contain information for predicting different scales of contour coordinates. For instance, the first order of position descriptor $f_1 \in \mathbb{R}^{W \times 1 \times n_x}$ matches the width W of the input image $I \in \mathbb{R}^{W \times H}$. f_1 extracts features with higher spatial correspondence. In contrast, lower scale descriptors f_2 and f_3 represent the position information with lower scale in the horizontal direction, but they extract more abstract features that are less sensitive to noise. Note that Fig. 1 just illustrates the schematic of the proposed networks. Applications can thus have more than 3 coordinates descriptors f_i ($i=1,2,3,\dots$). The higher parallelism allows for much faster inference compared to extracting hierarchical information in a cascaded way, as implemented in the U-Net architecture. The fusion encoder (gray block in Fig. 1) is deployed after the sub-branches to re-organize the multi-scale information,

This work was supported by the ATLAS project. The ATLAS project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813782.

TABLE I: Comparison with post-processed U-net on the different datasets with both region- and boundary- based metrics

	Jaccard index(\uparrow)			Dice coefficient (\uparrow)			Boundary error/pixel (\downarrow)	
	Lumen	Catheter	Tissue	Lumen	Catheter	Tissue	R1 v R2	R2 v R3
Phantom IVUS UNet	0.953 \pm 0.032	0.970 \pm 0.014	0.984 \pm 0.013	0.976 \pm 0.018	0.985 \pm 0.007	0.992 \pm 0.007	2.32 \pm 0.34	3.39 \pm 2.36
Phantom IVUS Ours	0.974\pm0.017	0.988\pm0.006	0.992\pm0.006	0.987\pm0.009	0.994\pm0.003	0.996 \pm 0.003	0.26\pm0.11	1.42\pm1.04
Phantom OCT UNet	0.973 \pm 0.035	0.937 \pm 0.043	0.600 \pm 0.297	0.985 \pm 0.019	0.966 \pm 0.025	0.699 \pm 0.242	2.54 \pm 0.38	5.51 \pm 11.2
Phantom OCT Ours	0.988 \pm 0.013	0.984\pm0.011	0.851\pm0.105	0.993 \pm 0.007	0.991\pm0.005	0.915\pm0.065	0.73\pm0.40	1.87\pm1.59
<i>In vivo</i> OCT UNet	0.704 \pm 0.117	0.916 \pm 0.019	0.894 \pm 0.051	0.820 \pm 0.092	0.956 \pm 0.010	0.943 \pm 0.029	2.06 \pm 0.47	11.8 \pm 9.84
<i>In vivo</i> OCT Ours	0.918\pm0.088	0.946\pm0.015	0.958\pm0.070	0.954\pm0.066	0.972\pm0.007	0.976\pm0.059	0.99\pm0.28	4.39\pm2.83

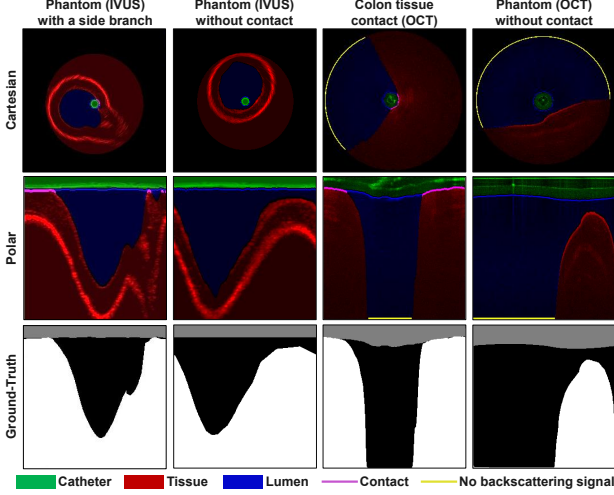


Fig. 2: Qualitative segmentation results.

which aligns all lower scale features to the original image scale, and produces A-line level contour coordinates as the boundary detection result.

RESULTS

An IVUS probe embedded at the tip of a robotic catheter with an active distal segment was steered in a poly(vinyl alcohol) (PVA) cryogel vessel phantom to collect the first dataset (3500 images). OCT images were acquired by steering an OCT probe in a colon phantom with layered tissue [4] (3000 images) as well as in an *in vivo* swine colon [5] (2000 images).

Qualitative results: Fig.2 shows representative results of the boundary detection experiments. Segmentation results are shown both in Cartesian and Polar domain. The detected boundaries are used to segment images. The catheter is shown in green, the tissue area is marked red, lumen (air or blood) is shown in blue and contact between catheter/sheath and tissue is highlighted in purple. Finally, A-lines without back scattering are marked yellow. The first column of Fig. 2 shows the IVUS catheter in contact with the tissue and close to a side branch, visible in the image. The second column shows a case where the catheter with IVUS is inside the phantom (small lumen) with good lumen contrast and no contact with the tissue. The third column is taken with OCT in a large lumen. The catheter is pressed in a tight contact between plastic sheath and tissue. The fourth column, in the same large lumen, shows the tissue mostly as a flat surface with the rest of the image showing background noise.

Validation strategies: For both IVUS and OCT, the acquired images were split into train and test dataset by 2:1. The region segmentation accuracy was computed

by means of the Jaccard index and the Dice coefficient, comparing the proposed method with a U-net trained with GAN loss [3]. As shown in Table I, the proposed architecture shows higher accuracy for all datasets. Furthermore, the U-net does not directly output the contour position. Through post-processing of the boundaries extracted from the U-net, the Euclidean distance with the ground-truth boundary was computed as an additional validation metric. Errors for 2 boundaries between 3 regions were calculated (regions of catheter, tissue and lumen). The last two columns of the Table I show the pix errors of each boundary ("R1 v R2" and "R2 v R3" denote boundaries of catheter/lumen and lumen/tissue, respectively). The proposed method has significantly lower errors compared to the prediction based on U-net. Mean distance errors are reduced from 5.4 \pm 5.5 to 1.9 \pm 1.2 for OCT images, and from 2.9 \pm 1.8 to 0.8 \pm 0.9 for IVUS images.

CONCLUSIONS AND DISCUSSION

In this work, a novel boundary detection architecture is proposed for polar domain processing of images acquired from side-viewing catheters. The proposed approach is applied to two different imaging modalities and evaluated by both area and boundary based metrics. The results show that, in comparison to state-of-the-art methods, higher accuracies are obtained for both OCT and IVUS. Moreover, no post-processing is required to predict the contour coordinates. The proposed method showed superior performance with accuracy improvements by 20% for side-viewing catheter images. Future work will investigate the use of this architecture for other tasks requiring boundary prediction such as e.g. for inspection of the esophagus by a capsule catheter. Further improvements could follow from predicting the existence probability for every object per A-line.

REFERENCES

- [1] Chaoyang Shi *et al.*, "Three-dimensional intravascular reconstruction techniques based on intravascular ultrasound: A technical review," *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 806–817, 2018.
- [2] Yan Ling Yong *et al.*, "Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography," *Journal of biomedical optics*, vol. 22, no. 12, p. 126005, 2017.
- [3] Isola Phillip *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [4] Natalia Zulina *et al.*, "Colon phantoms with cancer lesions for endoscopic characterization with optical coherence tomography," *Biomedical optics express*, vol. 12, no. 2, pp. 955–968, 2021.
- [5] Oscar Caravaca Mora *et al.*, "Steerable oct catheter for real-time assistance during teleoperated endoscopic treatment of colorectal cancer," *Biomedical optics express*, vol. 11, no. 3, pp. 1231–1243, 2020.

Semi-Supervised Placental Vessel Segmentation with Majority Voting for Pseudo-Label Updating*

Christina Bornberg, Sophia Bano, Francisco Vasconcelos, and Danail Stoyanov

Wellcome / EPSRC Centre for Interventional and Surgical Sciences and Department of Computer Science
University College London, UK

INTRODUCTION

Twin-to-twin transfusion syndrome (TTTS) is a condition where the blood flow in placenta anastomoses are unevenly distributed and occurs in 10-15% of mono-chorionic twin pregnancies. Without a surgical procedure, such as fetoscopic laser photocoagulation, this can lead to the death of both fetuses [1]. To assist the identification of placental blood vessels deep learning-based vessel segmentation can be applied.

Vessel segmentation of fetoscopy images. Sadda et al. [2] proposed a U-Net for segmenting placental blood vessels with class-weighted cross-entropy loss to deal with imbalance between vessels and non-vessel segments. Bano et al. [3] proposed a vessel segmentation approach with combined BCE and Jaccard loss as a preprocessing step to perform mosaicking.

Semi-supervised learning (SSL) is a technique to combine a (small) labelled dataset with a (larger) unlabelled dataset. The combined dataset helps the model to generalise better and it has proven to be useful in the medical image analysis field where manual image annotation is very costly and requires experts. Most SSL solutions are based on assumptions such as smoothness, consistency, clustering, or low-density [4].

Semi-supervised vessel segmentation has previously shown applications for retina, brain and coronary data [5]. Commonly used approaches include proxy-labelling (self-training, co-training), consistency regularisation (temporal ensembling, mean-teacher), hybrid methods (MixMatch, FixMatch), GAN (SGAN) and graph-based methods (label propagation) [4].

MATERIALS AND METHODS

In this work, a self-training network following the semi-supervised pseudo-labelling approach with majority voting using test-time augmentation is proposed. A U-Net is used to perform binary vessel segmentation.

A ResNet-18 backbone was used. The model was trained for 500 epochs, with a batch size of 32, and an image size of 512x512, with 1 channel (grayscale). Weighted BCE loss, adam optimizer and a learning rate of 0.0003 were chosen.

*This research was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z]; the Engineering and Physical Sciences Research Council (EPSRC) [EP/P027938/1, EP/R004080/1, EP/P012841/1]; and the Royal Academy of Engineering Chair in Emerging Technologies Scheme.

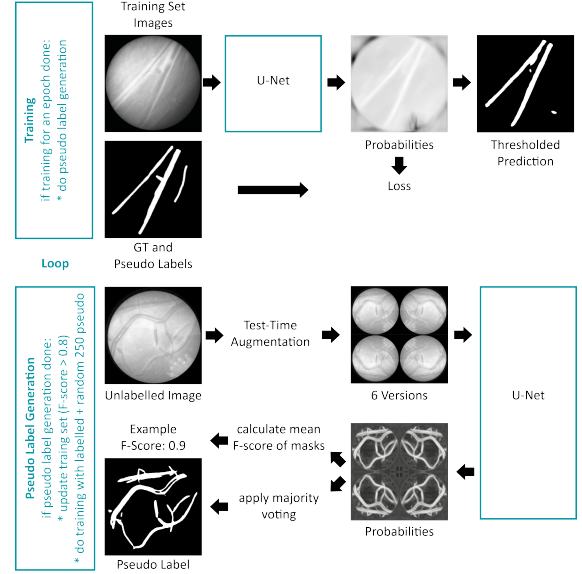


Fig. 1: Method overview: Loop of training and pseudo label generation after initial training on only manually annotated data until model reaches F-score > 0.5.

The used **fetoscopy dataset** [6] consists of 18 videos and is randomly split into train set (14), validation set (2) and test set (2) without mixing extracted images between the sets. Images from the 14 videos of the training set are extracted, 1607 labelled images are used for supervised learning and 5922 unlabelled images are used for the generation of pseudo labels and therefore training set enhancement.

Self-training first uses a small amount of labelled data to train a model. The model then assigns pseudo-labels to unlabelled data and are used as the ground truth for further training [4].

Test-time augmentation (TTA) can, similarly to test-time dropout, be used for the estimation of uncertainty by measuring the diversity of predictions with either variance or entropy [7]. An average binary mask can be created by majority pixel voting.

SEMI-SUPERVISED SELF-TRAINING NETWORK

For the proposed self-training network a U-Net is first trained in a supervised fashion using labelled data until it reaches an F-score above 0.5. From this point on whenever a model reaches a higher F-score, or after 15 epochs,

pseudo labels are generated/updated for the unlabeled data. The pseudo labels are generated using majority voting of the original and 5 augmented versions of the image (horizontal flip, vertical flip, horizontal and vertical flip, lowered contrast, smoothed). The prediction mask from the original image has a higher weight than the other augmentations. The majority mask is only used, if the mean between the F-score of the original and each of the augmented versions is higher than 0.8. Furthermore, more than 5 pixels need to be the foreground class. For each epoch, 250 randomly chosen pseudo-labelled images are used in addition to the training data. The overall workflow of our proposed semi-supervised self-training network is shown in Fig. 1.

CONCLUSIONS AND DISCUSSION

The network reached a validation F-score of 0.5 after 21 epochs. The final model was chosen with a validation F-score of 0.6599. By introducing the pseudo labels, generated by majority voting, the semi-supervised approach favours undersegmentation compared to an equivalent fully supervised baseline, only trained on labelled data. This can be seen by comparing the precision and recall values. The amount of false positives is higher in the baseline. See Table I.

Model	F-score	IOU	Precision	Recall
Base	0.6073	0.4699	0.56	0.78
Semi	0.6197	0.4844	0.62	0.72

TABLE I: Results of Semi-supervised approach and baseline, using a hold-out testset.

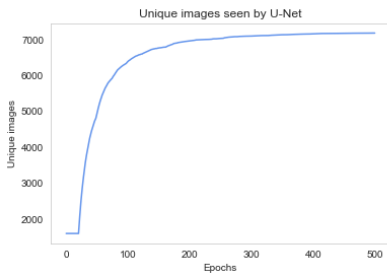


Fig. 2: Unique images seen by neural network, including 1607 labelled images. The labelled data and 250 randomly shuffled pseudo label images are used per epoch.

Conclusion: The proposed self-training approach for vessel segmentation gave a similar F-score as the supervised baseline, but reduced false positives.

Future work: Self-training networks have the disadvantage, to be unable to correct their own mistakes. Biased and wrong segmentations can be amplified, resulting in wrongly confident pseudo labels. More research in the choice of pseudo labels should be done. Furthermore, semi-supervised approaches with two models could be considered.

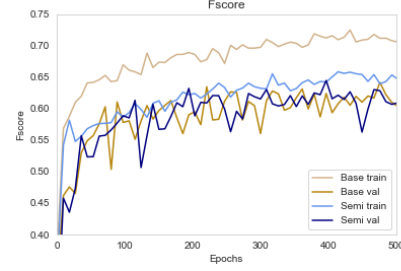


Fig. 3: F-score values for training and validation of base and semi. Showing a drop in train fscore when introducing pseudo labels from epoch 22.

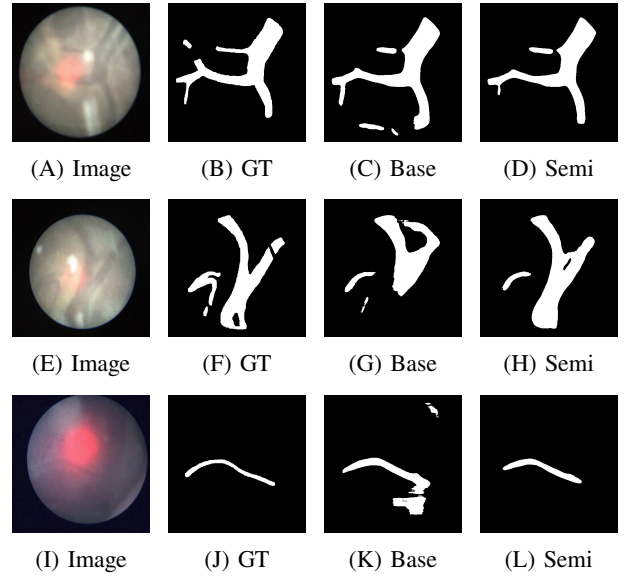


Fig. 4: Example images, showing oversegmentation from the baseline and undersegmentation of the semi-supervised approach, using the hold-out test set.

REFERENCES

- [1] A. Baschat, R. H. Chmait, J. Deprest, E. Gratacós, K. Hecher, E. Kontopoulos, R. Quintero, D. W. Skupski, D. V. Valsky, and Y. Ville, "Twin-to-twin transfusion syndrome (ttts)," *Journal of perinatal medicine*, vol. 39, no. 2, pp. 107–112, 2011.
- [2] P. Sadda, M. Imamoglu, M. Dombrowski, X. Papademetris, M. Bahtiyar, and J. Onofrey, "Deep-learned placental vessel segmentation for intraoperative video enhancement in fetoscopic surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, 11 2018.
- [3] S. Bano, F. Vasconcelos, L. M. Shepherd, E. Vander Poorten, T. Vercauteren, S. Ourselin, A. L. David, J. Deprest, and D. Stoyanov, "Deep placental vessel segmentation for fetoscopic mosaicking," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020.
- [4] Y. Ouali, C. Hudelot, and M. Tami, "An overview of deep semi-supervised learning," 2020.
- [5] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms — review of methods, datasets and evaluation metrics," *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 71–91, 2018.
- [6] S. Bano, A. Casella, F. Vasconcelos, S. Moccia, G. Attilakos et al., "FetReg: Placental Vessel Segmentation and Registration in Fetoscopy Challenge Dataset," *arXiv:2106.05923*, 2021.
- [7] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34–45, 2019.

DESIGN AND EVALUATION OF A HAPTIC CERVIX PHANTOM FOR HYSTEROSCOPY TRAINING

Vladimir Poliakov, Bas Van Delm, Igor Verbruggen, and Emanuel Vander Poorten

Department of Mechanical Engineering, KU Leuven, Belgium

INTRODUCTION

Hysteroscopy is an endoscopic gynaecological procedure that allows visual inspection of the uterine cavity and surgical treatment of uterine pathologies. During this procedure, a hysteroscope, which is a long slender instrument with a camera, is introduced in the uterus through the vagina and the cervical canal without making incisions. Recent advancements enabled outpatient hysteroscopic treatment, which also introduced a new challenge: as the cervical canal is the narrowest and the most sensitive part of the uterus, the clinician should be extra thorough and avoid any abrupt motion not to cause perforation or patient discomfort [1]. Future gynaecologists can train the skills required for successful hysteroscopy using box trainers, such as the commonly used HYST+T (European Academy of Gynaecological Surgery, Belgium) [2], which consists of a plastic uterus model placed inside a silicone vagina model. However, such simulators lack a realistic cervix model, impeding its use for practicing skills regarding pain management and careful cervix dilation.

This work presents the first prototype of an actuated cervix model for hysteroscopy training. To this end, an actuated device simulating the cervical dilation behaviour was designed and evaluated. Compact design and simplicity allows the system to be effortlessly integrated into existing box trainers. A series of stress tests was carried out to assess repeatability, stability and suitability of the proposed actuated phantom to represent cervical dilation behaviour. The results of the tests suggest an achievable repeatability below 0,5 mm. Therefore, the proposed device can be suitable for integration in existing box trainers. However, further research is necessary to finalize the haptic device. Additionally, user experiments should be carried out to validate the device and its use for training purposes.

MATERIALS AND METHODS

Other solutions for cervix dilation simulation have already been demonstrated previously [3]. However, this phenomenon has not been addressed in the field of hysteroscopic training, in which dynamic interaction and the level of force applied by the clinician plays the most important role. The working principle of the device is based on two arrays of rotary actuators attached to the silicone cervix model via a set of cables (Figure 1). This approach allows for independent dilation control both at the internal and external os of the phantom. The haptic device consists of four main parts: the tissue

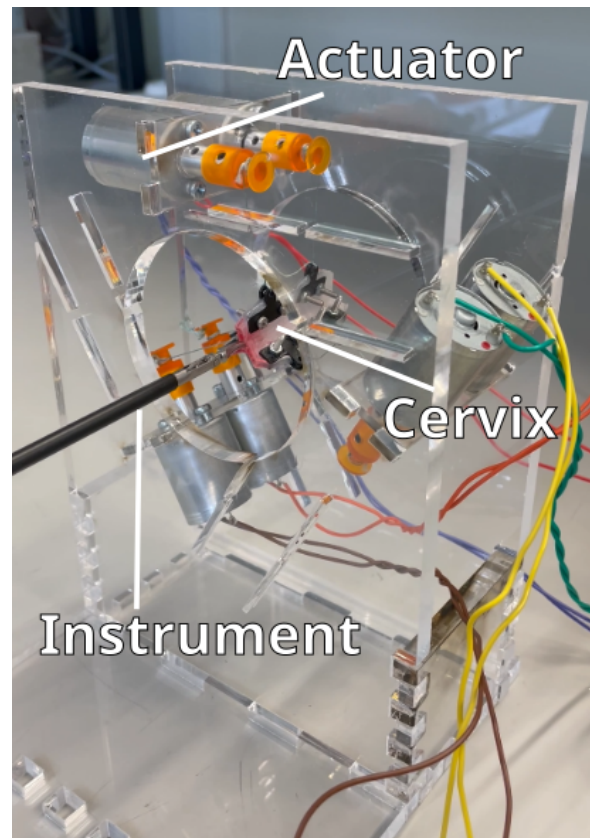


Fig. 1: The view upon the dilated haptic cervix phantom with medical forceps inserted inside.

phantom, the sensor, the actuator, and the controller. To simulate dilation behaviour, the sensor measures the forces applied to the cervical walls as a result of hysteroscope insertion. Based on these measurements, the controller can take the required action, controlling the dilation of the canal. Hence, the tissue allows for easier passage of a hysteroscope.

The selected material for the tissue model is Ecoflex (Smooth-On, US) silicone with a Shore hardness of 00-50. As determined in tensile tests by Liao et al. [4], E-modulus of this material is of an order of magnitude of 60 kPa. This value lies within the range of average cervical stiffness determined in a study by Castro et al. [5], and is, therefore, considered sufficient for the scope of this work.

Cyclic loading tests with visual measurements were executed with the prototype of the actuated phantom. This experimental setup consisted of the silicone phantom and

six actuators, four MCUs to control the linear actuators simultaneously, and a camera to measure dilation. The goal of the experiments was to verify the repeatability, cycling stability, and suitability of the proposed actuated phantom to represent cervical dilation behaviour. The same tissue phantom was subjected to three sets of 250 loading cycles, each separated in time by at least one hour to allow motors and electronics to cool down. Two pairs of Arduino boards, each board being supplied with 12 V DC and fitted with a motor shield, allowed for simultaneous control over the average motor terminal voltages by the means of pulse width modulation (PWM).

RESULTS AND DISCUSSION

A total of 13 835 video frames, split into three sets of measurements, was analyzed using Ilastik [6] and ImageJ [7]. It should be noted that, by evaluating the video footage, it was noticed that the motion resulting in dilation was not smooth. Instead, a stick-slip like behaviour was observed. A possible cause for this phenomenon is mechanical friction inside the motors, which was not easily counteracted by the slowly increasing motor torques, therefore resulting in interplay between static and dynamic friction. Moreover, because terminal voltages were directly controlled, rather than the rotor current, momentary torque was not necessarily proportional to the control inputs when the motors were not stalled.

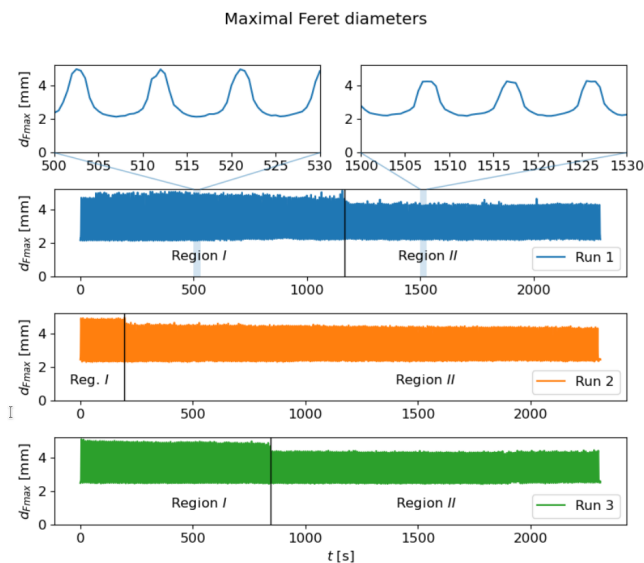


Fig. 2: The plot showing the maximum Feret diameter of the cervical canal across the loading cycles as a function of time.

The maximum Feret diameters of the cervical canal are plotted in Figure 2, as a function of time for all three runs of the experiment. Considering the cycle peaks for each run, the graph follows a similar course, which can be subdivided into two regions. The first region ranges between $t = 0$ and a timestamp at which a sudden drop is noticed in the top peaks. This timestamp is different time for each run but the drop has the same amplitude. From

the results and setup, there is no obvious cause for this drop. At minimal dilation, upward drift was observed in both the minimal and maximal Feret diameters, suggesting material creep. With standard deviations of 0,12 and 0,06 for maximal and minimal Feret diameter results, respective repeatability ranges are 0,46 and 0,24. Like the maximal dilation results, spread within runs is lower, with values below 0,2 mm.

CONCLUSIONS

To summarise, the proposed design fulfils most of the user requirements, only with curvature to be tested. To finish the haptic device, which can at least render the dilation behaviour of a straight cervix of average dimensions, without any pathologies, the following steps need to be undertaken:

- 1) The proposed sensor should be tested at the desired scale, before being calibrated and integrated into the haptic device.
- 2) Once the hardware components are assembled and tested, the controller can be implemented and tuned.
- 3) Final step involves integration of the platform into an existing box training system and user experiments with clinicians.

REFERENCES

- [1] S. Bettocchi and L. Selvaggi, "A vaginoscopic approach to reduce the pain of office hysteroscopy," *Journal of the American Association of Gynecologic Laparoscopists*, vol. 4, no. 2, pp. 255–258, 1997.
- [2] E. A. of Gynaecological Surgery, "Hystt: Hysteroscopic skills training and testing method," 2021, accessed: 2021-10-08. [Online]. Available: <https://europeanacademy.org/training-tools/hystt/>
- [3] C. Lunchini, S. Tognarelli, and A. Menciassi, "High-fidelity simulator of cervix changes during labour," in *Proceedings of the 10th Conference on New Technologies for Computer/Robot Assisted Surgery*, 2021.
- [4] Z. Liao, M. Hossain, and X. Yao, "Ecoflex polymer of different shore hardnesses: Experimental investigations and constitutive modelling," *Mechanics of materials*, vol. 144, p. 103366, 2020.
- [5] L. Castro, J. A. García-Mejido, E. Arroyo, J. Carrera, A. Fernández-Palacín, and J. A. Sainz, "Influence of epidemiological characteristics (age, parity and other factors) in the assessment of healthy uterine cervical stiffness evaluated through shear wave elastography as a prior step to its use in uterine cervical pathology," *Archives of gynecology and obstetrics*, vol. 302, no. 3, pp. 753–762, 2020.
- [6] T. ilastik developers, "ilastik: the interactive learning segmentation algorithm," 2021, accessed: 2021-10-10. [Online]. Available: <https://www.ilastik.org/>
- [7] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," pp. 671–675, 2012. [Online]. Available: <https://www.nature.com/articles/nmeth.2089>

DESIGN AND EVALUATION OF A HAPTIC CERVIX PHANTOM FOR HYSTEROSCOPY TRAINING

Vladimir Poliakov, Bas Van Delm, Igor Verbruggen, and Emanuel Vander Poorten

Department of Mechanical Engineering, KU Leuven, Belgium

INTRODUCTION

Hysteroscopy is an endoscopic gynaecological procedure that allows visual inspection of the uterine cavity and surgical treatment of uterine pathologies. During this procedure, a hysteroscope, which is a long slender instrument with a camera, is introduced in the uterus through the vagina and the cervical canal without making incisions. Recent advancements enabled outpatient hysteroscopic treatment, which also introduced a new challenge: as the cervical canal is the narrowest and the most sensitive part of the uterus, the clinician should be extra thorough and avoid any abrupt motion not to cause perforation or patient discomfort [1]. Future gynaecologists can train the skills required for successful hysteroscopy using box trainers, such as the commonly used HYST+T (European Academy of Gynaecological Surgery, Belgium) [2], which consists of a plastic uterus model placed inside a silicone vagina model. However, such simulators lack a realistic cervix model, impeding its use for practicing skills regarding pain management and careful cervix dilation.

This work presents the first prototype of an actuated cervix model for hysteroscopy training. To this end, an actuated device simulating the cervical dilation behaviour was designed and evaluated. Compact design and simplicity allows the system to be effortlessly integrated into existing box trainers. A series of stress tests was carried out to assess repeatability, stability and suitability of the proposed actuated phantom to represent cervical dilation behaviour. The results of the tests suggest an achievable repeatability below 0,5 mm. Therefore, the proposed device can be suitable for integration in existing box trainers. However, further research is necessary to finalize the haptic device. Additionally, user experiments should be carried out to validate the device and its use for training purposes.

MATERIALS AND METHODS

Other solutions for cervix dilation simulation have already been demonstrated previously [3]. However, this phenomenon has not been addressed in the field of hysteroscopic training, in which dynamic interaction and the level of force applied by the clinician plays the most important role. The working principle of the device is based on two arrays of rotary actuators attached to the silicone cervix model via a set of cables (Figure 1). This approach allows for independent dilation control both at the internal and external os of the phantom. The haptic device consists of four main parts: the tissue

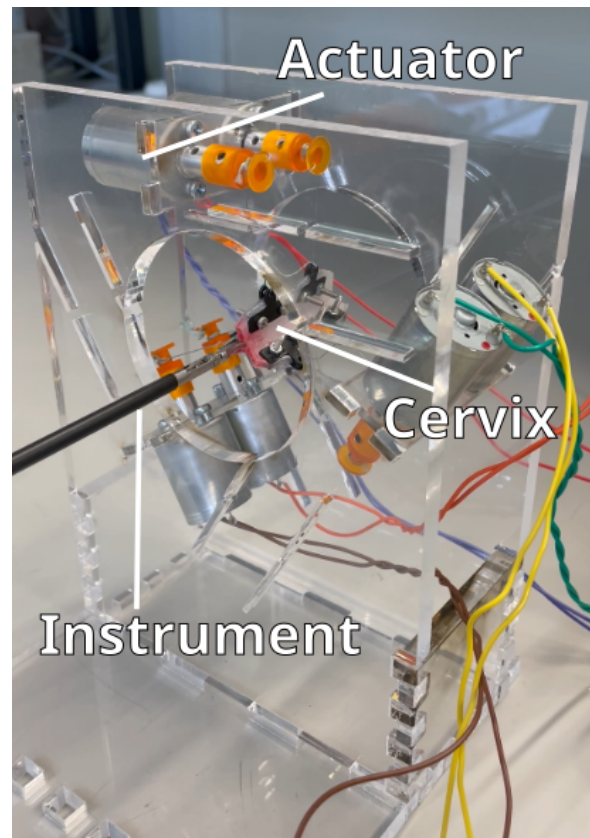


Fig. 1: The view upon the dilated haptic cervix phantom with medical forceps inserted inside.

phantom, the sensor, the actuator, and the controller. To simulate dilation behaviour, the sensor measures the forces applied to the cervical walls as a result of hysteroscope insertion. Based on these measurements, the controller can take the required action, controlling the dilation of the canal. Hence, the tissue allows for easier passage of a hysteroscope.

The selected material for the tissue model is Ecoflex (Smooth-On, US) silicone with a Shore hardness of 00-50. As determined in tensile tests by Liao et al. [4], E-modulus of this material is of an order of magnitude of 60 kPa. This value lies within the range of average cervical stiffness determined in a study by Castro et al. [5], and is, therefore, considered sufficient for the scope of this work.

Cyclic loading tests with visual measurements were executed with the prototype of the actuated phantom. This experimental setup consisted of the silicone phantom and

six actuators, four MCUs to control the linear actuators simultaneously, and a camera to measure dilation. The goal of the experiments was to verify the repeatability, cycling stability, and suitability of the proposed actuated phantom to represent cervical dilation behaviour. The same tissue phantom was subjected to three sets of 250 loading cycles, each separated in time by at least one hour to allow motors and electronics to cool down. Two pairs of Arduino boards, each board being supplied with 12 V DC and fitted with a motor shield, allowed for simultaneous control over the average motor terminal voltages by the means of pulse width modulation (PWM).

RESULTS AND DISCUSSION

A total of 13 835 video frames, split into three sets of measurements, was analyzed using Ilastik [6] and ImageJ [7]. It should be noted that, by evaluating the video footage, it was noticed that the motion resulting in dilation was not smooth. Instead, a stick-slip like behaviour was observed. A possible cause for this phenomenon is mechanical friction inside the motors, which was not easily counteracted by the slowly increasing motor torques, therefore resulting in interplay between static and dynamic friction. Moreover, because terminal voltages were directly controlled, rather than the rotor current, momentary torque was not necessarily proportional to the control inputs when the motors were not stalled.

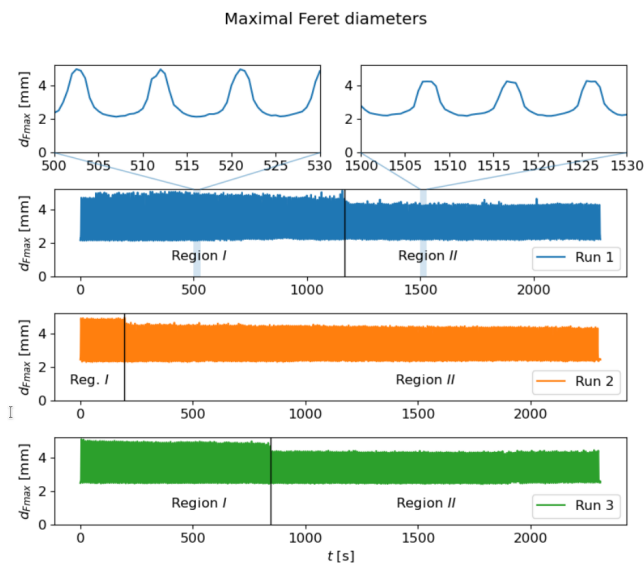


Fig. 2: The plot showing the maximum Feret diameter of the cervical canal across the loading cycles as a function of time.

The maximum Feret diameters of the cervical canal are plotted in Figure 2, as a function of time for all three runs of the experiment. Considering the cycle peaks for each run, the graph follows a similar course, which can be subdivided into two regions. The first region ranges between $t = 0$ and a timestamp at which a sudden drop is noticed in the top peaks. This timestamp is different time for each run but the drop has the same amplitude. From

the results and setup, there is no obvious cause for this drop. At minimal dilation, upward drift was observed in both the minimal and maximal Feret diameters, suggesting material creep. With standard deviations of 0,12 and 0,06 for maximal and minimal Feret diameter results, respective repeatability ranges are 0,46 and 0,24. Like the maximal dilation results, spread within runs is lower, with values below 0,2 mm.

CONCLUSIONS

To summarise, the proposed design fulfils most of the user requirements, only with curvature to be tested. To finish the haptic device, which can at least render the dilation behaviour of a straight cervix of average dimensions, without any pathologies, the following steps need to be undertaken:

- 1) The proposed sensor should be tested at the desired scale, before being calibrated and integrated into the haptic device.
- 2) Once the hardware components are assembled and tested, the controller can be implemented and tuned.
- 3) Final step involves integration of the platform into an existing box training system and user experiments with clinicians.

REFERENCES

- [1] S. Bettocchi and L. Selvaggi, "A vaginoscopic approach to reduce the pain of office hysteroscopy," *Journal of the American Association of Gynecologic Laparoscopists*, vol. 4, no. 2, pp. 255–258, 1997.
- [2] E. A. of Gynaecological Surgery, "Hystt: Hysteroscopic skills training and testing method," 2021, accessed: 2021-10-08. [Online]. Available: <https://europeanacademy.org/training-tools/hystt/>
- [3] C. Lunchini, S. Tognarelli, and A. Menciassi, "High-fidelity simulator of cervix changes during labour," in *Proceedings of the 10th Conference on New Technologies for Computer/Robot Assisted Surgery*, 2021.
- [4] Z. Liao, M. Hossain, and X. Yao, "Ecoflex polymer of different shore hardnesses: Experimental investigations and constitutive modelling," *Mechanics of materials*, vol. 144, p. 103366, 2020.
- [5] L. Castro, J. A. García-Mejido, E. Arroyo, J. Carrera, A. Fernández-Palacín, and J. A. Sainz, "Influence of epidemiological characteristics (age, parity and other factors) in the assessment of healthy uterine cervical stiffness evaluated through shear wave elastography as a prior step to its use in uterine cervical pathology," *Archives of gynecology and obstetrics*, vol. 302, no. 3, pp. 753–762, 2020.
- [6] T. ilastik developers, "ilastik: the interactive learning segmentation algorithm," 2021, accessed: 2021-10-10. [Online]. Available: <https://www.ilastik.org/>
- [7] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," pp. 671–675, 2012. [Online]. Available: <https://www.nature.com/articles/nmeth.2089>

Automatic bone breach detection for spine surgery based on bio-electrical conductivity sensing: Ex-vivo experimental validation

Jimmy Da Silva^{1,2}, Elie Saghbiny², Thibault Chandanson¹,
Stéphane Bette¹, Maurice Bourlion¹, and Guillaume Morel²

¹SpineGuard, Vincennes, France. ²ISIR, UMR 7222 Sorbonne University, CNRS, U1150 INSERM, Paris, France.

INTRODUCTION

Several procedures in spine surgery require the surgeon to insert screws in vertebrae to immobilize some parts of the spine with metallic rods, *e.g.*, for scoliosis correction. The drilling trajectory is chosen to pass via a narrow anatomical part called the pedicle. Misplacing the pedicle screws can induce many complications due to the proximity of critical neural or vascular elements [1].

The main difficulty associated to the manual free-hand procedure is that the precise location of the tool is not directly visible. To assist surgeons, X-ray imaging systems have been coupled to marker-based optical tracking devices to provide a real-time visual estimation of the tool position in Virtual or Augmented Reality. Recently, spine surgery robots have been introduced to these technologies to position autonomously a drill guide on top of the patient. However, these robots do not perform the drilling themselves.

To better understand what is happening at the tool tip, SpineGuard, a medical device company, designs tools embedding local bio-electrical conductivity sensing thanks to a bipolar sensor pulsing current flow at the tip of their instruments [2], [3]. The measured signal varies with the bone density and allows discriminating between cortical bone (dense), cancellous bone (spongy), and soft tissues (blood, muscles, etc).

With the idea to provide additional on-line safety check for robotized spine surgery, we proposed a new concept showcasing a robotic arm using a tool equipped with conductivity sensing in [4]. The present paper describes a more thorough experimental investigation of the concept. It discloses a set of 104 experimental drillings performed on *ex-vivo* lamb vertebrae, where 100% of the drillings were autonomously stopped at the interface between the bone and the spinal canal thanks to bio-electrical conductivity measurements.

MATERIALS AND METHODS

The mechatronic setup consisted of an LBR 7 Med redundant robotic manipulator from KUKA, a custom-made power drill, and a threaded drill bit embedding a conductivity sensor, prototyped by SpineGuard.

All the experiments were conducted on fresh lamb lumbar vertebrae acquired at the butcher shop. The *ex-vivo* pieces, once at room temperature, were fixed in a clamping vice, which was placed inside a transparent box. The container was filled with a saline solution to reproduce

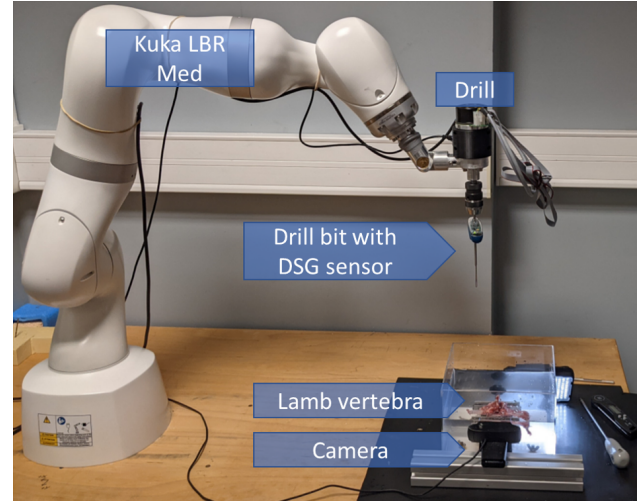


Fig. 1: Experimental setup.

the conductivity of the cerebrospinal fluid (CSF), usually present in the vertebral canal during real surgeries. A camera was positioned right outside the box to record the interior of the spinal canal and look for potential bone breaches. The overall setup is shown in Fig. 1.

An entry point was made manually in the spinous process for each trial. Then the tool was hand-guided to the entry point, and oriented towards the spinal canal. Next, the robot was controlled to keep the orientation fixed and apply a constant force of 10N, while drilling with the threaded instrument with a 1mm pitch. The power drill was controlled at a fixed rotation speed of 30 rpm, leading to almost a constant insertion speed (see Fig. 2c) of 0.5mm/s., thanks to the threads on the instrument.

A preliminary experiment, performed on 100 continuous drillings of lamb spinous processes, allowed for gathering electrical conductivity signals corresponding to bone breaches. Then, the collected data was used to develop and tune a bone breach detection algorithm to stop the robot for all drillings in a ± 2 mm zone from the border between the bone and the vertebral canal. Such breaches would correspond to grades A and B of the Gertzbein-Robbins classification of pedicle screw misplacement [5].

The resulting Algorithm 1 uses the conductivity σ and the depth z measurements to create a security flag *Alert* enabling to stop the system. Parameters σ_{\max} and α are used to create an adaptable threshold on the conductivity signal. At the same time, $\Delta\sigma$ and Δz are used to monitor

Algorithm 1 Bone breach detection, called for each new conductivity measurement; Initially $Alert = \text{false}$, $\sigma_r = 0$ and Σ, Z are empty lists. **Blue values** are the parameters tuned from calibrating experiments

Input: σ_k , conductivity signal in mV
Input: z_k , depth insertion in the bone in mm
Output: $Alert$, flag used to stop the drilling

$\Sigma \leftarrow [\Sigma \ \sigma_k], Z \leftarrow [Z \ z_k]$ \triangleright Constructing lists

if $z_k < 5$ **then**
 $Alert \leftarrow \text{false}$
 if $z_k < 3$ **then** $j \leftarrow k$
 else $\sigma_r \leftarrow \frac{(k-j-1)\sigma_r + \sigma_k}{k-j}$ **end if**
else
 $A_1 \leftarrow \sigma_k > \min(500, 2.4 \ \sigma_r)$
 $m \leftarrow \underset{i \in \{1 \dots k\}}{\operatorname{argmin}} |Z(i) - (z_k - 2)|$
 $n \leftarrow \underset{i \in \{m \dots k\}}{\operatorname{argmin}} \Sigma(i)$
 $A_2 \leftarrow \left(\left(\max_{i \in \{n \dots k\}} \Sigma(i) \right) - \Sigma(n) \right) > 230 \text{ mV}$
 $Alert \leftarrow Alert || A_1 || A_2$
end if

the conductivity increases over a few millimeters.

All parameters have been tuned to post-operatively trigger a stop for the data of the preliminary experiments, $\pm 2\text{mm}$ around the interface between cortical bone and the vertebral canal.

Then, a second experiment, comprising 104 new vertebrae, was performed with the previously defined automatic bone breach detection algorithm and parameters, with the intent to stop the threaded drill bit right around the interface between the cortical bone and the spinal canal.

The final position of the tool relatively to the bone/canal interface was then evaluated for each trial by using the synchronized data (robot displacement and videos) acquired during the experiment. To do so, we measured, in post-processing, the number of millimeters of robot displacement after seeing bone movement on camera. False positives (stops happening before reaching the interface) were assessed thanks to post-operative CT scans.

RESULTS

The presented robotic system successfully drilled all the 104 lamb lumbar vertebrae autonomously without breaching outside of the bone.

For each drilling, a surgeon verified via palpation with a ball-tip feeler that the instrument did not fully breach outside the bone. Moreover, the recorded video feed from the webcam allowed to visually verify that the instrument did not pass the vertebral wall (see Fig. 2b).

Also, a post-experiment CT scans analysis confirmed that all drillings were stopped within less than 2mm from the canal (0% false positive). A few vertebrae were passed through a micro-CT scan to better visualize the resulting hole drilled in the spinous processes (see Fig. 2a).

The post-processing of the synchronized robot logs and webcam videos permitted to estimate the amount of bone

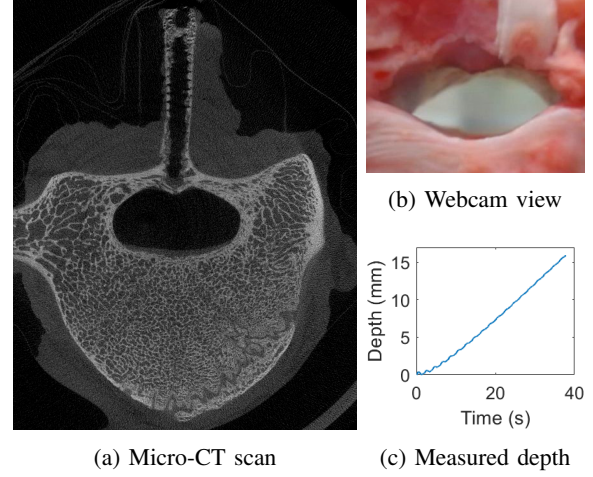


Fig. 2: Results on one autonomously drilled vertebra.

pushed inside the vertebral canal. The mean displacement inside the canal was 0.65mm, with a standard deviation of 0.4mm.

CONCLUSION AND DISCUSSION

A total of 204 vertebrae were drilled with the same robotic setup. The 100 first tests were unstopped drillings, used to collect data on bone breaches. The last 104 drillings were automatically stopped at the bone interface thanks to a detection algorithm.

The CT scans and the recorded videos showed that the detection happens when the cortical bone starts to crack, *i.e.*, before the hole is thoroughly drilled. The videos of the preliminary experiment also allowed to visualize the bone deformation (bump) happening in the vertebral canal before bone perforation.

The trajectory used in this experiment was perpendicular to the spinal canal. Even in this worst-case scenario, all the drillings were graded A or B with the Gertzbein-Robbins classification, which is clinically acceptable. Nonetheless, future work will need to validate the algorithm on actual pedicle trajectories.

ACKNOWLEDGEMENTS

This research has received funding from the EU's H2020 research and innovation program under grant agreement No. 101016985 (FAROS project) and from ANRT through the CIFRE program.

REFERENCES

- [1] J. E. Lonstein *et al.*, "Complications associated with pedicle screws," *JBJS*, vol. 81, no. 11, pp. 1519–1528, 1999.
- [2] C. Bolger *et al.*, "Electrical conductivity measurement: a new technique to detect iatrogenic initial pedicle perforation," *Eur. Spine J.*, vol. 16, no. 11, pp. 1919–1924, nov 2007.
- [3] O. Suess *et al.*, "Control of Pedicle Screw Placement with an Electrical Conductivity Measurement Device: Initial Evaluation in the Thoracic and Lumbar Spine," *Adv. Med.*, vol. 2016, pp. 1–7, sep 2016.
- [4] J. D. Silva *et al.*, "Robot-assisted spine surgery guided by conductivity sensing: first preclinical experiments demonstrate X-ray free breach detection," in *Hamlyn Symp. Med. Robot.* The Hamlyn Centre, Imperial College London, jun 2019, pp. 75–76.
- [5] S. D. Gertzbein *et al.*, "Accuracy of Pedicular Screw Placement In Vivo," *Spine (Phila. Pa. 1976)*, vol. 15, no. 1, pp. 11–14, jan 1990.

Force control of the KUKA LBR Med without external force sensor

Jimmy Da Silva^{1,2}, Saman Vafadar², Thibault Chandanson¹, and Guillaume Morel²

¹SpineGuard, Vincennes, France. ²ISIR, UMR 7222 Sorbonne University, CNRS, U1150 INSERM, Paris, France.

INTRODUCTION

In recent years, robots embedding joint torque sensors and impedance control have been industrialized and commercialized, aiming to deploy safe human-robot collaboration applications. Among them, the Kuka LBR Med benefits from all the certifications to ease its integration in medical applications. The embedded torque sensors open the opportunity to implement force control [1], a feature that is missing in most existing surgical robots. Yet, it is crucial for safety in a context where uncertainties and disturbances, such as tissue deformations or physiological movements, challenge robot stability at contact.

Impedance control preserves the robot's stability in the presence of uncertainties when the robot contacts its environment [2]. However, within this framework, explicitly controlling interaction forces applied by a robot to its environment is impossible when the environment position and stiffness are unknown. In this framework, force-regulation or tracking can be achieved using inner-outer loop control schemes [3], [4].

Installing a force/torque sensor on an end-effector to implement explicit force control brings two drawbacks: i) non-collocated modes that challenge stability [5]; ii) from the perspective of a translation to clinical application, adding costly and fragile equipment to a medical system is of high cost and risk.

In this paper, we experimentally investigate the possibility of finely controlling the forces applied by the robot to the environment based on KUKA's built-in joint impedance controller. We first use an external force/torque sensor to provide the feedback signal, leading to poor performance due to non-collocated dynamics. We then introduce a pseudo-force feedback approach using joint position sensors that significantly improves force response rapidity and robustness.

MATERIALS AND METHODS

We consider a robot equipped with an inner joint impedance controller. The dynamic relationship between the external joint torque vector τ_e and the joint position vector \mathbf{q} of an impedance-controlled robotic arm is:

$$M_d \ddot{\mathbf{q}} = K_d(\mathbf{q}_r - \mathbf{q}) + B_d(\dot{\mathbf{q}}_r - \dot{\mathbf{q}}) - \tau_e, \quad (1)$$

where M_d , B_d , K_d , and \mathbf{q}_r represent the desired mass, damping, stiffness matrices, and the reference joint position vector. Typically, diagonal constant matrices are used for K_d , B_d , and M_d , leading to independent linear dynamic equations.

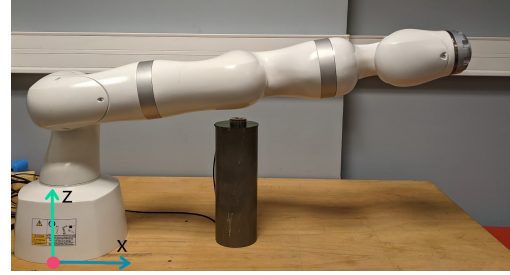


Fig. 1: Setup used for the 1-DoF experiment.

In a first step, we study a one-joint problem and use plain lowercase letters to symbolize scalars:

$$m_d \ddot{q} = k_d(q_r - q) + b_d(\dot{q}_r - \dot{q}) - \tau_e. \quad (2)$$

Regulating the torque to its desired value τ_d requires adjusting q_r in real-time by means of an outer torque loop. Two different outer loops are considered. The first one uses classical explicit torque regulation ([3], Fig. 2a):

$$\dot{q}_r = \lambda(\tau_d - \tau_e), \quad (3)$$

where λ is a proportional gain and τ_e is the measured at the end-effector level using the installed force sensor. The second one exploits a pseudo-torque signal $\tilde{\tau}_e = k_d(q_r - q)$ as the feedback signal (Fig. 2b):

$$\dot{q}_r = \lambda[\tau_d - k_d(q_r - q)]. \quad (4)$$

Note that according to Eq. (2), the pseudo-torque signal corresponds to the real external torque at the equilibrium, i.e.: if $\ddot{q} = \dot{q} = \dot{q}_r = 0$, then $\tilde{\tau}_e = \tau_e = k_d(q_r - q)$.

In a second step, we expand the 1-DoF case to a multi-DoF problem for force regulation. The control law is then:

$$\dot{\mathbf{q}}_r = J^\dagger \Lambda [F_d - (J^\dagger)^T K_d(\mathbf{q}_r - \mathbf{q})], \quad (5)$$

where F_d is the desired force/torque (wrench) vector; J^\dagger denotes the Moore-Penrose inverse of the Jacobian matrix J ; And, Λ is a diagonal matrix whose elements denote the proportional gains $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_6)$. Gain λ_i is set to zero for directions where the force is not controlled, leaving it free for conventional trajectory control, similar to hybrid position/force control.

Experiments, 1-DoF and 7-DoF, were conducted using the KUKA LBR 7 Med robotic arm running its built-in joint impedance controller. For the 1-DoF experiment, the second joint of the KUKA arm (Fig. 1) was controlled. The other joints were servoed to their initial position. An ATI Mini40 force and torque (F/T) sensor was placed at the interface between the robot and the environment to

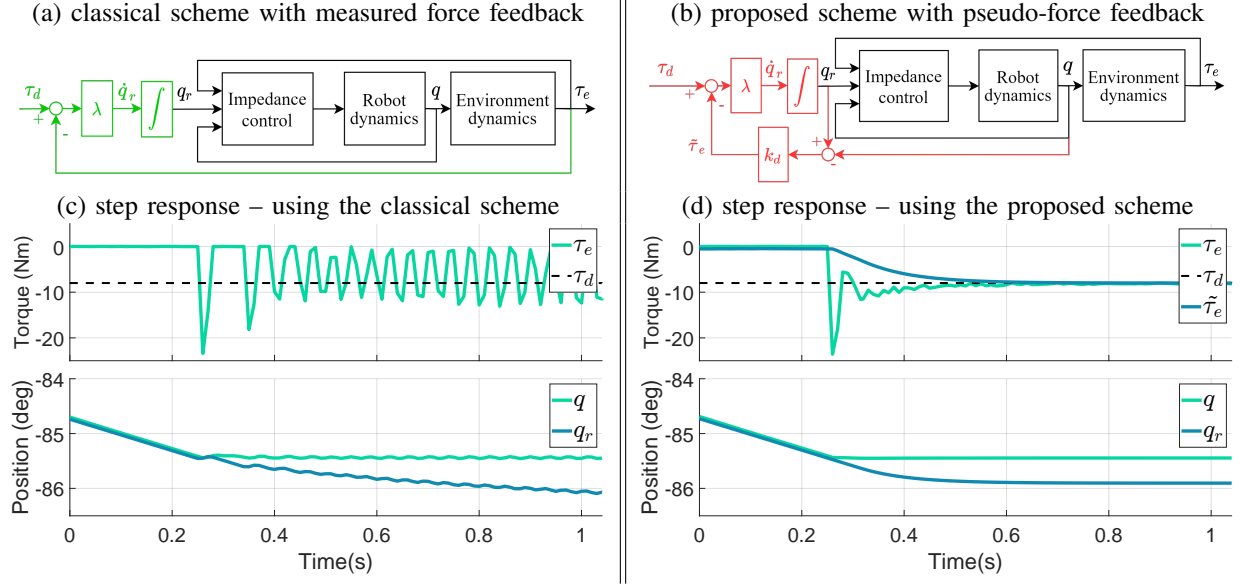


Fig. 2: The 1-DoF experiment – control schemes and results. Comparison between the control schemes.

provide τ_e by projection about the robot's second joint axis. In practice, for safety, a velocity saturation \dot{q}_{\max} was added to Eq. (3). This had the experimental effect of inducing limit cycles when stability conditions were not met [6]. The desired torque was set to $\tau_d = -8.0$ Nm. Tuning the controller consisted of experimentally selecting $\lambda = 0.01$ rad/s/Nm and $\dot{q}_{\max} = 0.05$ rad/s to ensure stability in various experimental conditions with the pseudo-force feedback. The same tuning was kept for a fair comparison with the external sensor force feedback.

For the 7-DoF experiment, all joints of the KUKA arm were controlled using the pseudo-force feedback control scheme. The F/T sensor was placed at the interface between the robot and the environment only for evaluation. Similar to the 1-DoF experiment, an operational space velocity saturation $\dot{x}_{\max} = 0.02$ m/s was added. The desired force was -20.0 N along the vertical axis to contact a hard horizontal plane (see Fig. 1). The gains λ_i were all set to zero but for $\lambda_3 = 0.03$ m/s/N, corresponding to the vertical force component.

RESULTS AND DISCUSSION

Figure 2 shows the results obtained during the 1-DoF experiment, when the robot started in the free-space and contacted the environment, using the external (left) or pseudo (right) joint torque signal. Although the gains are identical, the external sensory approach leads to instability (Fig. 2c), whereas the pseudo-torque feedback method leads to a stable and precise response and, compared to literature with this robot arm, relatively fast (Fig. 2d). The servoed signal $\tilde{\tau}_e$ smoothly converged towards the desired value within approximately 200 ms. Meanwhile, the real torque τ_e experienced a peak at the impact that was not fed back through the outer loop, thus not destabilizing the system. Observed robustness arises from the fact that the feedback signal $\tilde{\tau}_e$ excludes the high-frequency components. Yet, it does not involve any low-pass filtering,

that would add poles within the closed-loop bandwidth and induce stability issues.

Figure 3 shows the results of the 7-DoF experiment. Similar to the 1-DoF experiment, we observed a stable behavior. The servoed signal (pseudo-force, blue) was smooth and converged toward the desired force value rapidly. The observed slope, before reaching the desired value, corresponds to the speed saturation that was set.

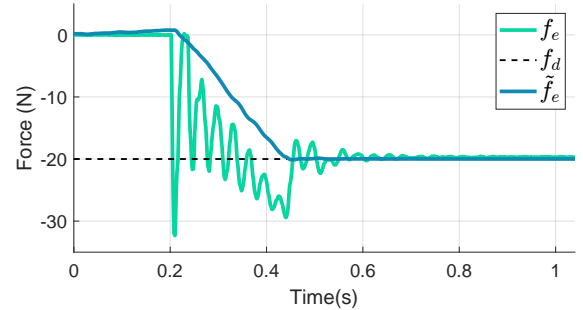


Fig. 3: The 7-DoF experiment – results.

ACKNOWLEDGEMENTS. This research has received funding from the EU's H2020 research and innovation program under grant agreement No. 101016985 (FAROS project) and from ANRT through the CIFRE program.

REFERENCES

- [1] T. Haidegger *et al.*, "Force sensing and force control for surgical robots," *IFAC*, 2009.
- [2] N. Hogan, "On the stability of manipulators performing contact tasks," *IEEE Journal on Robotics and Automation*, 1988.
- [3] T. A. Lasky *et al.*, "On force-tracking impedance control of robot manipulators," *Proc. - IEEE Int. Conf. Robot. Autom.*, 1991.
- [4] F. Almeida *et al.*, "Force-impedance control: a new control strategy of robotic manipulators," *Recent advances in Mechatronics*, 1999.
- [5] S. Eppinger and W. Seering, "Three dynamic problems in robot force control," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 6, pp. 751–758, 1992.
- [6] D. Liu *et al.*, "Asymptotic stability of discrete-time systems with saturation nonlinearities with applications to digital filters," *IEEE Transactions on Circuits and Systems I*, 1992.

Multi-level-assistance Robotic Platform for Navigation in the Urinary System: Design and Preliminary Tests

M. Finocchiario^{1,3,*}, X. T. Ha^{2,3,*}, J. F. Lazo^{4,7,*}, C. Lai^{4,5,*}, S. Ramesh^{6,7,*}, A. Hernansanz¹,
G. Borghesan², D. Dall'Alba⁶, S. Tognarelli³, B. Rosa⁷, A. Casals¹, N. Padoy^{7,8}, P. Fiorini⁶,
J. Dankelman⁶, E. Vander Poorten², A. Menciassi³, E. De Momi⁴

¹Universitat Politècnica de Catalunya, Spain, ²Katholieke Universiteit Leuven, Belgium, ³Scuola Superiore Sant'Anna, Italy,
⁴Politecnico di Milano, Italy, ⁵Delft University of Technology, Netherlands, ⁶University of Verona, Italy
⁷ICube, University of Strasbourg, CNRS, France, ⁸IHU Strasbourg, France

INTRODUCTION

Ureteroscopy is the gold standard procedure for treatment and diagnosis of upper urinary tract diseases such as carcinoma, urinary strictures and urolithiasis. As a minimally invasive procedure, it implies accessing the organs through natural orifices, avoiding incisions. This is translated in lower risk of infections and shorter recovery time for the patient. However, performing this procedure is a non trivial task and mastering it requires an extensive training. Current challenges related to navigation with traditional instruments inside narrow luminal organs, such as the ureter, could turn into a complex task due to the limited intuitiveness in controlling the endoscope movements, the poor visual feedback, and the absence of any type of guidance or assistance in current endoscopic systems [1].

In this context, Robotic Flexible Ureteroscopy offers an opportunity to overcome the mentioned challenges, ease the burden of the clinicians [2] and offer better treatments for patients. The advantages these robotic systems present include the use of ergonomic Human Machine Interfaces (HMI) which allow more precise and smoother movement of the tools and the possibility to operate the robot remotely at a safer distance from radiation. Nevertheless, since the first reported clinical use of a robotic ureteroscope, the Sensei Magellan system (Hansen Medical, no longer commercially available) [3], few other robotic platforms have been tested for urological applications. This lack of robotic platforms might be related to the fact that the platforms did not provide sufficient assistance to the operator (*e.g.* haptic feedback, augmented reality, autonomous navigation *etc.*), limiting the benefit of their use compared to manual approaches.

In this regard, this abstract introduces the design and a preliminary user test of an innovative robotic platform able to address the current challenges related to navigation in the urinary tract. The proposed platform presents different levels of assistance up to a level 4 of autonomy as defined in [4]. The platform is endowed with (1) a multi-steerable active catheter, (2) multi-level autonomy HMI, (3) a real-time tracking system of the position and shape of the device inside the lumen and (4) real-time suggestions and aids for the user (*e.g.*, current procedure phase, position of the center of the lumen *etc.*). The mechanical properties of the active robotic catheter together with its autonomous and semi-autonomous abilities could help clinicians to prevent perforations and get support during the procedure. Furthermore, thanks to the integrated tracking system, real-time position of the ureteroscope together with its shape mapped inside the patient's anatomy may become available

*These authors contributed equally to this work.

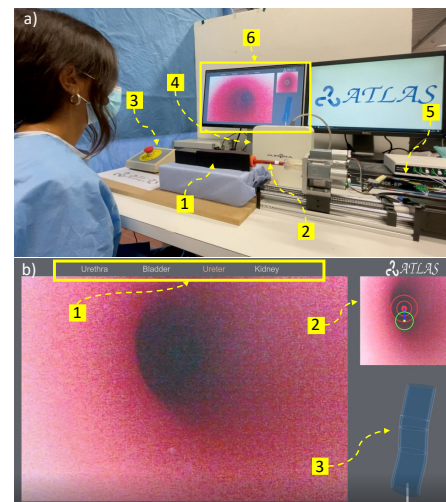


Fig. 1. a) General perspective of the Multi-level-assistance robotic platform including: 1) ureter phantom, presenting the visual conditions of a real ureter; 2) built robotic endoscope; 3) emergency stop button; 4) EM tracking system; 5) FBG interrogator and 6) Human Machine Interface. b) The Graphical User Interface: 1) current phase of the procedure; 2) processed output (center of the ureter) of our visual-servoing module and 3) 3D shape of the endoscope relative to the phantom expressed in the EM coordinate frame.

during the whole surgery, reducing the need of using X-rays for localization.

MATERIALS AND METHODS

The proposed system is depicted in Fig. 1 and it includes three main components:

- A *Visual-Servoing Module*, based on [5], comprises of (1) a cable-driven soft robotic endoscope which has a backbone and a helical structure with two bending directions. The steerable segment of the soft robotic endoscope is 70 mm long which is similar to the one in a ureteroscope; (2) an actuation robotic platform to bend in two directions and to insert the robotic endoscope. In total, there are 3 Degrees of Freedom (DoFs) in Visual-Servoing Module. An Arduino Mega is used to implement two PID controllers and served as a bridge between high-level commands and all the actuators. A deep learning based visual servoing high-level controller is used to autonomously segment the lumen from the camera images, as presented in [6], and compute the center of the ureter. The information from the detected center is used to calculate the error and bring the tip of the endoscope towards the detected center point. The average center detection time is 0.15 s deployed on a NVIDIA GeForce RTX 2080 GPU.

- A *Shape Sensing Module* adapted from [7] includes a multicore Fiber Bragg Grating (FBG) (FBGS, Belgium) embedded in the center channel to sense the 3D shape of the soft robotic endoscope. Two electromagnetic (EM) tracking sensors (NDI, Canada) are attached to the tip and to the base of the robotic scope to localize the 3D reconstructed shape in the EM coordinate frame.
- A multi-level autonomy *HMI* including a Graphical User Interface (GUI) implemented in Unity3D (Unity Technologies, USA) and a joystick for controlling the endoscope in manual or assisted control. The GUI shows: (1) the endoscopic image recorded by the camera, (2) the position and deformation of the endoscope registered in the pre-operative and intra-operative images, (3) the image with higher level features such as computed by the visual-servoing module, and (4) the phase of the procedure.

Three different levels of assistance are considered.

- 1) *Manual* - The operator can see the endoscopic images recorded by the camera and the information regarding the position and deformation of the endoscope obtained from the shape sensing module. The operator controls the 3 DoFs of the endoscope through a joystick or keyboard.
- 2) *Visual* assistance - In addition to the information mentioned in the manual scenario, information regarding the detected center of the lumen and the clinical phase is shown to the operator. In this operation mode, the navigation is still performed by the user using the joystick.
- 3) *Autonomous* - During fully automated intraluminal navigation the visual servoing module drives the endoscope inside the lumen, i.e. the 2 DoF bending and the insertion/retraction, following the center-line detected with the computer vision module. The clinician supervises the procedure and in case there is any concern they can halt immediately the process by pressing the emergency stop button and recover full manual control over the endoscope. In case that the computer vision system fails on detecting the lumen, the robot halts its movement.

As initial test, we performed a set of experiments to compare the capabilities of the system, at its different levels of assistance, for the task of centering the endoscope in the lumen. In the *Manual* and *Visual* scenarios, all participants (10 individuals with non-medical background) allowed to get familiar with the system for five minutes before performing the task in order to exclude possible learning effects. The participants had only one chance and were asked to command the robot, i.e. moving only 2 DoFs for the bending of the tip, to reach the center of the lumen. In the *Autonomous* scenario, the robot performs the task with aid of the visual servoing high-level controller.

The performance metrics taken into account are the settling time and the Steady State Error (SSE). The settling time is defined as the first time the endoscope reaches a distance within less than 20% of the initial distance with respect to the center of the lumen. A condition was herein that this position is maintained for more than one second. The SSE is defined as the distance between the theoretical detected lumen center and the center of the camera frame when the participants or the robot finished the tasks.

RESULTS AND CONCLUSION

A total of 10 participants were considered for the *Manual* and *Visual* feedback experiments. In the case of the *Autonomous* scenario 10 experiments were carried out. Boxplots comparing the results between each of the modalities are shown in Fig. 2.

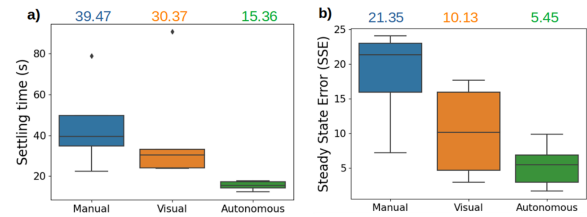


Fig. 2. Boxplots comparison of a) Steady State Error (pixels) and b) settling time between the three modalities of the system (manual control, visual feedback, autonomous) tested for the lumen centering task. The median value for each setting is presented on the top.

In the case of settling time the median values obtained were 39.47, 30.37 and 15.36 s for the manual, visual, and autonomous respectively and the values obtained for SSE were 21.35, 30.37 and 15.36 pixels, respectively. In both metrics the autonomous modality obtains the best performance. In the case of settling time, it reaches the goal in half the time that is required with visual feedback and is 2.5 faster than the case when there is no feedback. For the case of SSE metric the values obtained with visual feedback and manual mode are twice and four times higher than the autonomous mode.

In conclusion, robotic platforms represent an opportunity to reduce the risks and difficulties related to ureteroscopy. This work successfully demonstrated autonomous navigation in a simple ureter phantom using a robotic catheter endowed with real-time shape sensing and localization and a multi-level autonomy HMI. Future work includes (1) integrate FBG/EM trackers data to improve the autonomous control algorithm of the catheter (now used only for shape-sensing and localization), (2) test the whole system in a multi-organ phantom, (3) conduct user tests with expert clinicians.

ACKNOWLEDGEMENT

This work was supported by the ATLAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813782. This work was also partially supported by French State Funds managed by the "Agence Nationale de la Recherche (ANR)" through the "Investissements d'Avenir" (Investments for the Future) Program under Grant ANR-10-IAHU-02 (IHU-Strasbourg).

REFERENCES

- [1] V. De Coninck, E. X. Keller, B. Somani, G. Giusti, S. Proietti, M. Rodríguez-Socarras, M. Rodríguez-Monsalve, S. Doizi, E. Ventimiglia, and O. Traxer, "Complications of ureteroscopy: a complete overview," *World journal of urology*, vol. 38, no. 9, 2020.
- [2] J. Rassweiler, M. Fiedler, N. Charalampogiannis, A. S. Kabacki, R. Saglam, and J.-T. Klein, "Robot-assisted flexible ureteroscopy: an update," *Urolithiasis*, vol. 46, no. 1, 2018.
- [3] M. M. Desai, R. Grover, M. Aron, A. Ganpule, S. S. Joshi, M. R. Desai, and I. S. Gill, "Robotic flexible ureteroscopy for renal calculi: initial clinical experience," *The Journal of Urology*, vol. 186, no. 2, 2011.
- [4] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastri, "Autonomy in surgical robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 651–679, 2021.
- [5] J. F. Lazo, C.-F. Lai, S. Moccia, B. Rosa, M. Catellani, M. de Mathelin, G. Ferrigno, P. Breedveld, J. Dankelman, and E. De Momi, "Autonomous intraluminal navigation of a soft robot using deep-learning-based visual servoing," 2021.
- [6] J. F. Lazo, A. Marzullo, S. Moccia, M. Catellani, B. Rosa, M. de Mathelin, and E. De Momi, "Using spatial-temporal ensembles of convolutional neural networks for lumen segmentation in ureteroscopy," *International Journal of Computer Assisted Radiology and Surgery*, 2021.
- [7] X. T. Ha, M. Ourak, O. Al-Ahmad, D. Wu, G. Borghesan, A. Menclassi, and E. Vander Poorten, "Robust catheter tracking by fusing electromagnetic tracking, fiber bragg grating and sparse fluoroscopic images," *IEEE Sensors Journal*, vol. 21, no. 20, 2021.

Real-Time Detection of Laryngeal Squamous Cell Carcinoma via Video Laryngoscopy with WL and NBI using YOLO

Muhammad Adeel Azam, Shunlei Li, Ajay Gunalan, and Leonardo S. Mattos

Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy,

adeel.azam@iit.it, adeelazam70@gmail.com

INTRODUCTION

Laryngeal squamous cell carcinoma (LSCC) is inflammatory cancer that develops in the mucosa layer of the throat. To limit the morbidity caused by surgery or non-surgical therapies, early diagnosis of malignant tumors is essential. LSCCs are diagnosed in the clinic using an adjustable trans-nasal laryngoscopy. Currently, due to the limited resolution of the analogic optical images generated by standard fiber optic trans-nasal laryngoscopy, it has intrinsic disadvantages.

High-definition video laryngoscopy has recently supplanted traditional fiber optic endoscopy. High-definition optical laryngoscopy has progressively overtaken traditional fiber optic imaging since it identifies laryngeal malignancies with much more precision [1]. However, with white light (WL) laryngoscopy, the morphology of larynx mucosal abnormalities can be very non-specific, and preoperative clinical examination is sometimes not reliable in evaluating the real eventual malignancy. Narrow Band Imaging (NBI) is a well-established optical technology that uses narrow spectrum filters to improve imaging methods for effective tissues detection in the larynx.

Unfortunately, there are several limits to the current use of this technique, particularly in low experience clinics: analyzing NBI endothelial patterns is difficult and largely relies on the clinician's competence, and therefore as a result, it requires excessive time. Computer-aided detection (CAD) uses deep learning (DL) approaches including Convolutional Neural Networks (CNNs) for automatic human diseases assessments. Furthermore, CAD driven by strong and reliable CNNs has the potential to significantly boost the effect of video laryngoscopy along with NBI in massive LSCC monitoring in the future.

The previous study has attempted the prospect of using CAD to identify laryngeal disease, and the findings have been impressive [2], [3]. These studies, on the other hand, only focused on how well-trained CNNs could identify the presence of a specific lesion on static images. Instead, ongoing research in this domain must focus on applying CAD to real-time video detection. The prospect of using DL for real-time detection and localization of LSCC on video laryngoscopy WL and NBI recordings is investigated in this work. We emphasized assessing the accuracy of our trained DL model and its applicability for real-time video endoscopic LSCC automated detection. The proposed Ensemble You-Only-Look-Once (YOLO) framework

[4], which is built on V5s and V5m with pre-trained weights and TTA, performs best for identifying LSCC. A few sample LSCC images and enhanced images are shown in figure 1.

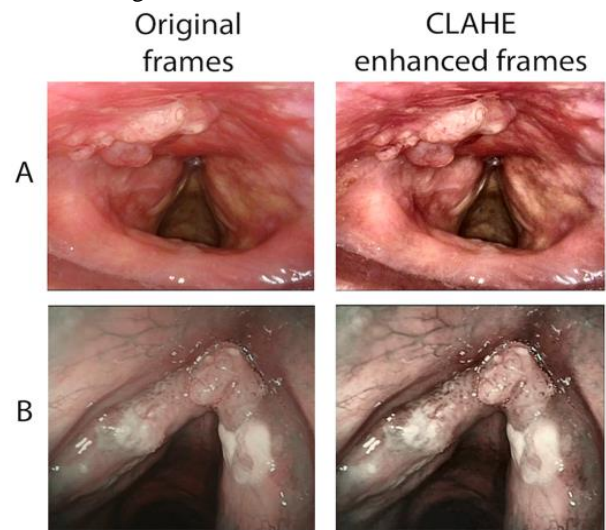


Figure. 1. WL and NBI LSCC images (A) WL; (B) NBI

MATERIALS AND METHODS

Glenn Jocher first introduced YOLOv5 on GitHub in the year 2020 [4]. YOLO utilizes a compact neural network architecture to provide bounding boxes and class labels in real-time. YOLO surpassed other state-of-the-art algorithms, including such faster-RCNN detection. Yolo v5 has several advanced improvements that seek to improve test time and activation functionality. Among the most remarkable extra features is the distinctive mosaic data augmentation.

YOLOv5 trained model is suitable for real-time detection on low computation devices [5] since it has faster inference and fewer model weight files than YOLOv4. The three major blocks that make up the YOLO deep learning architecture are the backbone, neck, and head as shown in figure 2.

Yolov5s, Yolov5m, Yolov5l, and Yolov5x are the four various model versions of the YOLOv5 that range in size from tiny to extra-large. The tiny model comprises 7.3 million while the extra-large model comprises 87.7 million parameters. We incorporated two new Yolo variations, Yolov5s6 and Yolov5m6, in our research and findings. The COCO pre-trained weights are used in the Yolov5 model for transfer learning features.

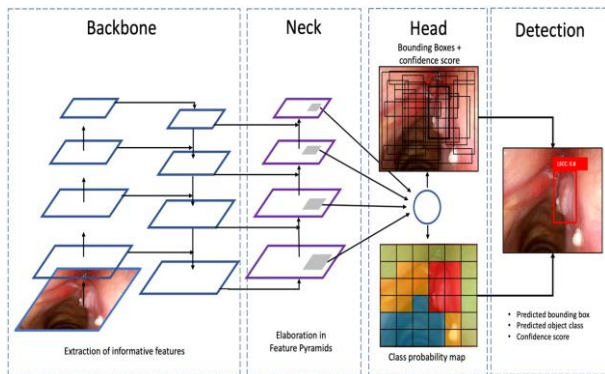


Figure. 2. YOLOv5 framework architecture for detection.

EXPERIMENTAL ANALYSIS

The IRCCS Ospedale Policlinico San Martino organizational ethics board accepted the retrospective analysis (CER Liguria: 230/2019), which adhered to the Helsinki Declaration requirements. This prospective study included patients at IRCCS Ospedale Policlinico San Martino in Genoa, Italy, between 2014 and 2019. More than 200 patients are considered in this study, with the majority of those being male, with 196 of those being male, and 23 of those being female, with an average age of 67.9 years.

The current state-of-the-art YOLOv5 algorithm was tested on 657 images (318 WL, 306 NBI, and 33 unlabeled). The model trained with 543 images, of which 256 were WLs, 254 were NBIs, and 33 were left unlabeled. Validation contained 54 images, 32 of which were in WL and 22 in NBI. Finally, 60 randomized images were evaluated, 30 from the WL and 30 from the NBI. Data augmentation approaches contrast limited adaptive histogram equalization (CLAHE), and test time augmentation (TTA) applied to increase model performance. Six pre-operative laryngoscopy videos were chosen to evaluate the trained DL model's real-time LSCC detection performance.

RESULTS

The tiny YOLOv5 model variant performed very well on a limited and diverse dataset. Due to the small dataset, the model with the highest parameter started overfitting (unnecessary prediction) early in training and validation. The Ensemble approach presented here diagnosed laryngeal carcinoma in images that were similar to the ground truth. The Ensemble YOLO model shown here has the best performance for diagnosing LSCC, with a positive predictive value (PPV) of 0.664, a sensitivity of 0.621, a mean average precision (mAP) of 0.627, and a true positive (TP) of 82 percent. Figure 3 illustrates both the actual video frames and the model's expected frames.

In video-stream testing phase, we focused only on inference times for the real-time approach excluding image acquisition time. The Ensemble model (Yolo5s plus Yolo5m – TTA) was chosen as the principal framework in this context due to its superior performance. The model detected LSCC with an average latency per frame of 0.023 to 0.034 milliseconds.

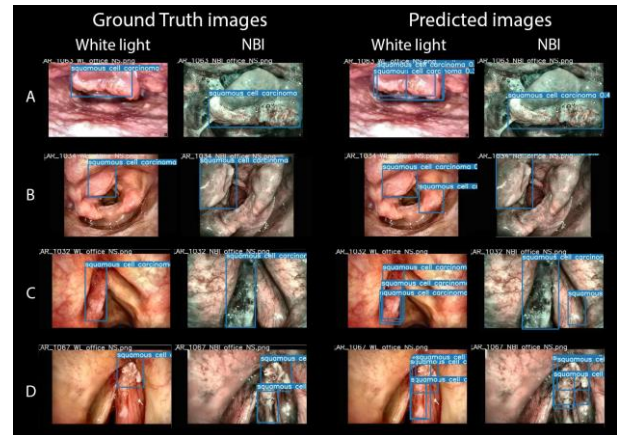


Figure. 3. Proposed Ensemble YOLOv5 prediction results on static images.

CONCLUSION AND DISCUSSION

This is the first study using YOLO to detect LSCC cancer as per our knowledge. This DL model can detect LSCC in images and videos in real-time. Our ensemble model had the best LSCC detection precision, recall, and mAP@.5. Technically, we emphasize the importance of using data augmentation to help the model learn how to identify LSCCs. Finally, overlapping bounding boxes improved the inference strength of the original YOLO model.

We found the Ensemble YOLOv5 model suitable to be used in WL and NBI video laryngoscopies. This model is ideal for real-time processing owing to its efficiency and accuracy. Even with the small training set used in this study, the availability of NBI images to feed the algorithm was critical to the reported detection performances. If this model will be trained on a larger image dataset, we expect it to perform faster and more effectively in the future.

REFERENCES

- [1] Scholman C, Westra JM, Zwakenberg MA, et al. Differences in the diagnostic value between fiberoptic and high definition laryngoscopy for the characterisation of pharyngeal and laryngeal lesions: A multi-observer paired analysis of videos. *Clin Otolaryngol.* 2020; 45(1):119-125
- [2] Vilaseca I, Valls-Mateus M, Nogués A, et al. Usefulness of office examination with narrow band imaging for the diagnosis of head and neck squamous cell carcinoma and follow-up of premalignant lesions. *Head Neck.* 2017; 39 (9):1854-1863.
- [3] Tama BA, Kim DH, Kim G, Kim SW, Lee S. Recent advances in the application of artificial intelligence in otorhinolaryngology-head and neck surgery. *Clin Exp Otorhinolaryngol.* 2020; 13 (4):326-339.
- [4] GitHub - ultralytics/yolov5: YOLOv5. Published 2021. Accessed July 27, 2021. <https://github.com/ultralytics/yolov5>.
- [5] Bisio I, Haleem H, Garibotto C, Lavagetto F, Sciarrone A. Performance Evaluation and Analysis of Drone-based Vehicle Detection Techniques From Deep Learning Perspective. *IEEE Internet of Things Journal.* 2021 Nov 15.

Toward automatic bone breach detection for spine surgery using tissue bio-electrical conductivity sensing

Elie Saghbiny^{1,2}, Jimmy Da Silva^{2,3}, Célia Chaimi², Thibault Chandanson³,
Guillaume Morel², and Raphaël Vialle¹

¹ISIR, UMR 7222 Sorbonne University, CNRS, U1150 INSERM, Paris, France. ²Armand Trousseau hospital, APHP, Paris, France. ³SpineGuard, Vincennes, France.

INTRODUCTION

Scoliosis is a three-dimensional deformation of the spine [1]. The treatment consists in correcting the deformity using a rigid construct until the fusion of the spine vertebrae. The construct is formed mainly by screws inserted into the vertebrae's pedicles and two rods (see Fig. 1a).

The first step for pedicle screw placement is preparation using a pedicle probe. Pedicles are small and deformed in patients having scoliosis, while being close to vital structures such as the spinal cord. If pedicle preparation, followed by screw placement, is not optimally performed, it may cause severe complications for the patient. For instance, medial and lateral cortical perforation may cause vascular and neurological damage (see Fig. 1b). The conventional free-hand non-assisted pedicle screw placement can lead to a misplacement rate for idiopathic scoliosis surgery as high as 14% [2].

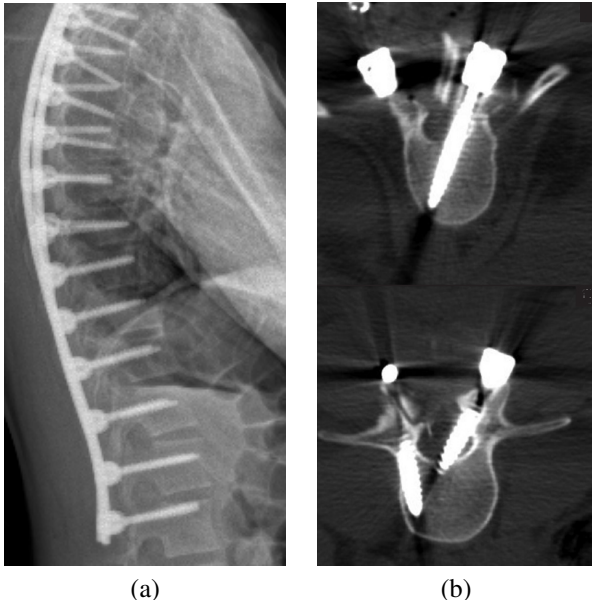


Fig. 1: (a) Sagittal X-ray of the spine after extended spinal fusion for a scoliosis patient. (b) Axial scan views of a misplaced pedicle screw [3].

Navigation systems for guiding pedicle preparation are based on irradiating imaging techniques. To reduce irra-

diation, the SpineGuard company develops surgical instruments for pedicle preparation, such as the PediGuard, which measures the tissue's electrical conductivity in contact with the instrument's tip and sends auditory and visual feedback to the surgeon [4]. Using the PediGuard instrument for spine surgery has been proven to reduce the need for fluoroscopy use by 30% while maintaining a 97.5% accuracy [5].

Several studies proved that measuring electrical conductivity and sending audio feedback to the surgeon increases accuracy for optimal pedicle screw placement [4], [6]. However, no study has directly investigated electrical conductivity values and compared them to clinical evaluation. The present study aims to develop an automatic bone breach detection system, which could be integrated into surgical robotics for automated pedicle preparation.

MATERIALS AND METHODS

Data are collected during surgeries performed on idiopathic scoliosis patients, all operated at the same university hospital. Pedicle preparation is done using the Pediguard (instrument developed by SpineGuard) that measures the electrical conductivity signal $DSG(t)$ at its tip at 5 Hz. Values are recorded through a wireless application on a laptop. The surgeries are video recorded with a 10 mm diameter endoscope with zero-degree angulation linked to an arthroscopy column (Fig. 2). All the data are recorded on the same laptop for time synchronization.

For each pedicle, the surgeon evaluates the prepared trajectory as usual, *i.e.* through palpation with a probe or a ball-tip "feeler". The palpated trajectory is defined as "optimal" if it feels like a bone tunnel without any cortical perforation; defined as "breached" if there is a bone perforation, and modification of the aiming direction is needed. Trajectories that do not meet these above definitions, such as voluntary "in-out-in" trajectories, have been classified as "undetermined" and eliminated from the statistical analysis.

Then, for data post-processing, at each time sample, the insertion depth of the PediGuard's tip inside the vertebra is calculated using image processing techniques thanks to the recorded videos using the endoscope. The electrical conductivity values are only taken into account when measured after 10 mm of depth. Before 10 mm, the data might not be relevant because of the blood in



Fig. 2: Setup used in the operating room: A: endoscope, B: Pediguard, C: Arthroscopy column, D: Laptop recording synchronously video and electrical conductivity data.

the pre-drilled entrypoint. Finally, the electrical conductivity difference ΔDSG_{1sec} is computed between values separated by 1-second time interval, *i.e.*, $\Delta DSG_{1sec}(t) = DSG_t - DSG_{t-1}$.

RESULTS & DISCUSSION

A total of 457 pedicle preparations are recorded on 21 patients. 266 are classified as "optimal", 40 as "breached", and 151 as "undetermined". The maximum value ΔDSG_{max} of the signal $\Delta DSG_{1sec}(t)$ is computed for each pedicle. The median of ΔDSG_{max} across all the data are 159 mV for the "optimal" group, and 579.5 mV for the "breached" group ($p < 0.05$) (see Fig. 3).

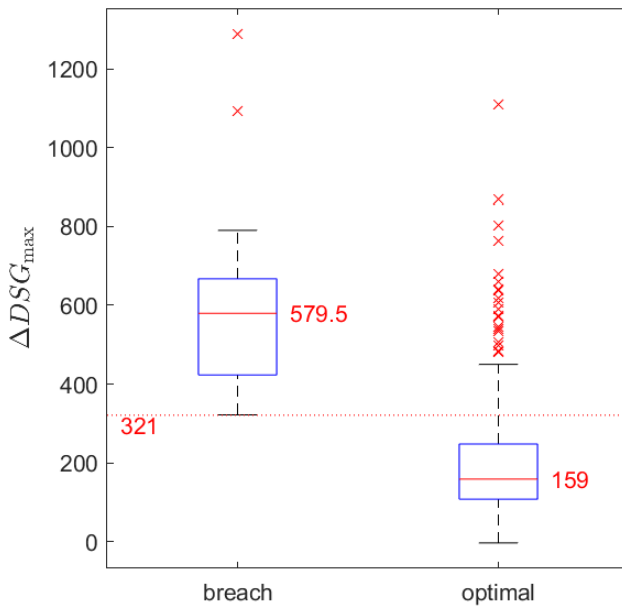


Fig. 3: Distribution of ΔDSG_{max} according to the quality of the pedicle aiming.

No pedicle preparation in the "breached" group has a maximum value less than 321 mV ($\Delta DSG_{max} < 321$ mV), and only 7.9% in the "optimal" group have $\Delta DSG_{max} > 321$ mV. Therefore, this shows that ΔDSG_{max} can be used to detect if a pedicle preparation is optimal in

92.1% of cases. The presence of intraspinal hemangioma, a benign vascular tumor, or remnants of growth plates in the vertebral body might explain the lack of detection in 7.9% of cases; the incidence of intraspinal hemangioma is between 10-27% [7].

This study demonstrated that an increase in the electrical conductivity measurement of more than 321 mV in a 1-second time interval evokes the possibility of a breach with a 92.1% chance. The 321 mV corresponds to the minimum value of ΔDSG_{max} of the "breached" group. Therefore, this information could be valuable to inform surgeons that a change in the direction of the trajectory is necessary before inserting the screw. Also, the obtained results show that a breach detection model could be used to increase the safety of robot-assisted pedicle screw placement preparation [8].

ACKNOWLEDGMENTS

This research has received funding from the CAMI labex and ANRT through the CIFRE program.

REFERENCES

- [1] W. Skalli *et al.*, "Quantification of three-dimensional vertebral rotations in scoliosis: what are the true values?" *Spine*, vol. 20, no. 5, pp. 546–553, 1995.
- [2] K. Abul-Kasim *et al.*, "The rate of screw misplacement in segmental pedicle screw fixation in adolescent idiopathic scoliosis," *Acta Orthop.*, vol. 82, no. 1, pp. 50–55, feb 2011.
- [3] T. Orief *et al.*, "Accuracy of percutaneous pedicle screw insertion in spinal fixation of traumatic thoracic and lumbar spine fractures," *Surgical neurology international*, 2018.
- [4] C. Bolger *et al.*, "A preliminary study of reliability of impedance measurement to detect iatrogenic initial pedicle perforation (in the porcine model)," *Eur. Spine J.*, 2006.
- [5] C. D. Chaput, K. George, A. F. Samdani, J. I. Williams, J. Gaughan, and R. R. Betz, "Reduction in radiation (fluoroscopy) while maintaining safe placement of pedicle screws during lumbar spine fusion," *Spine*, vol. 37, no. 21, pp. E1305–E1309, 2012.
- [6] M. Allaoui *et al.*, "Contribution of Dynamic Surgical Guidance to the Accurate Placement of Pedicle Screws in Deformity Surgery: A Retrospective Case Series," *World Neurosurg.*, 2018.
- [7] M. Barzin *et al.*, "Incidence of Vertebral Hemangioma on Spinal Magnetic Resonance Imaging in Northern Iran," *Pakistan J. Biol. Sci.*, 2009.
- [8] J. D. Silva *et al.*, "Robot-assisted spine surgery guided by conductivity sensing: first preclinical experiments demonstrate X-ray free breach detection," in *Hamlyn Symp. Med. Robot.* The Hamlyn Centre, Faculty of Engineering, Imperial College London, 2019.

Augmented Reality and Robotic Navigation System for Spinal Surgery

Matteo Boles¹, Junling Fu¹, Elisa Iovene¹, Francesco Costa², Giancarlo Ferrigno¹, Elena De Momi¹

¹Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

²Department of Neurosurgery, Fondazione Istituto Nazionale Neurologico, "C. Besta", Milan, Italy

INTRODUCTION

Robotic spinal surgery requires the surgeon to drill a hole inside the vertebral body of the patient at the level of the pedicle and insert a screw afterward while ensuring to avoid causing damage to surrounding vulnerable structures, such as nerve roots, spinal cord, and blood vessels. To achieve this, computer-assisted navigation (CAN) and robotic guided techniques are frequently adopted to assist surgeons. The CAN-based method allows surgeons to see the driller's position on a screen in real-time, however, requires great hand-eye coordination. In the latter one, robots are usually adopted as tool guidance systems. Operations such as the drilling, are still manually performed by surgeons to control the driller position accurately and stop timely, which may introduce lesions into sensitive tissues.

To tackle the above-mentioned problems, augmented reality (AR) can cooperate with the robotics prototype to achieve real-time navigation and control of the path and depth during surgical operations. Besides, different techniques have been investigated in AR applications to align the virtual model with the real one [1]. Manual-based approaches showed the difficulty in assessing relative depths of physical and virtual models while external tracking system-based method could inevitably introduce errors when computing the transform matrix from the markers to its view origin.

Besides, automatic drilling also has been explored and various kinds of sensors utilized to acquire signals such as force [2], sound [3] and vibration [4]. However, these signals can be easily affected by external noise. Also, the sensor-based technique utilizes the threshold algorithms, which may result in erroneous transition detection with inappropriate threshold values.

MATERIALS AND METHODS

In this paper, an AR-based robotic navigation system for spinal surgery was developed. CT dataset of a vertebral column (UWSpineCT dataset, Imperial College of London) was segmented and processed using 3D Slicer. The real vertebral model is made of Polylactic Acid (PLA) material using a 3D printer, with 20-micron layer resolution ultra-smooth surfaces.

As shown in Fig.1, the real model is positioned inside a sponge and covered with another one to resemble skin, fat and muscle of human body. Fiducial-based registration approach was adopted to match the virtual model with the

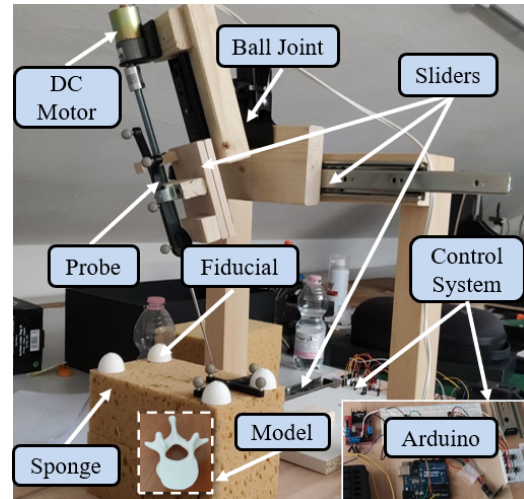


Fig. 1: Prototype of the developed robotic system.

real one using four 3D printed fiducial markers which are glued on the sponge. As shown in Fig.2, fiducials were located in the real world using a 3D printed pen equipped with a QR code to track the position of its tip and automatically matched with their virtual counterparts by computing the translation and rotation matrix through the Singular Value Decomposition (SVD) algorithm.

The robotic prototype shown in Fig. 1 is composed of three parts, namely the mechanical structure, the motor drilling system, and the control system. The mechanical structure has 6 Degree-of-Freedom and is composed of three sliders and a ball joint. The motor drilling system is driven by a DC motor with the probe mounted to represent the automatic drilling process. Besides, an Arduino board is adopted as the control system and connected to the wireless network using an ESP-8266 Wi-Fi module. The NDI Polaris Vicra (Northern Digital, Inc, Ontario, Canada) is used to track the position of the mounted probe.

In Fig. 2, the pre-planned path of drilling the obtained vertebra model was manually defined in 3D Slicer and then it is integrated into the hologram of the vertebra for visualization. Fig. 3 illustrates the communication diagram of the whole system. The position of the probe was integrated on the CT scans using the Plus server in real-time and displayed on a virtual screen in Fig. 2. Command will be sent from MATLAB to ROS (Robot Operating System) to stop the movement of the probe when it reaches

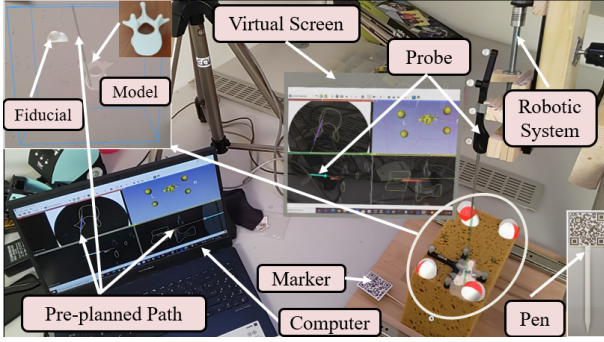


Fig. 2: Experimental setup of the proposed AR-based robotic spinal surgery navigation system.

the pre-planned position.

To verify the feasibility of the proposed navigation system, nine subjects were invited to perform the real and virtual model alignment task and position the robotic prototype in line with the entry point of the 3D printed vertebra inside the sponge looking to the virtual path displayed on the hologram and to the virtual screen in the AR application.

In this paper, performance metrics are chosen to quantitatively and qualitatively evaluate the experimental results, including both the objective and subjective metrics:

- The Fiducial Registration Error (FRE), used to define the distance between the hologram and the real model, which calculates the Root-Mean-Square distance between two fiducials point clouds.
- The target and angular deviation, between the desired path and the actual path, computed in the frontal, sagittal, and transversal planes.
- Questionnaire-based subjective method to evaluate the usability of the designed system.

RESULTS AND DISCUSSION

The experiment of aligning the real model and the virtual one is first implemented. Experimental results showed an alignment error between the real and virtual model of 4.07 ± 0.48 mm reaching a maximum of 4.85 mm and starting from a minimum of 3.37 mm.

The system positioning accuracy was then evaluated through the angular and target deviation defined in the three planes. As shown in TABLE I, experimental results showed the target deviation errors of 1.34 ± 0.59 mm in the frontal plane while 0.91 ± 0.40 mm and 0.88 ± 0.65 mm in the transversal and sagittal plane, respectively. Considering the thickness of the cortical bone at which the driller during spinal surgery should stop has a 2 mm thickness, the error results were acceptable. In terms of angular deviation we got instead in the frontal, transversal and sagittal plane, errors $1.31^\circ \pm 0.87^\circ$, $0.93^\circ \pm 1.55^\circ$, $6.38^\circ \pm 6.10^\circ$ respectively.

Furthermore, subjective experimental results demonstrated that all the subjects involved in the experiment found the system easy to use without too much information to be known. Also, the questionnaire results illustrate that most of the participants preferred to rely on the virtual

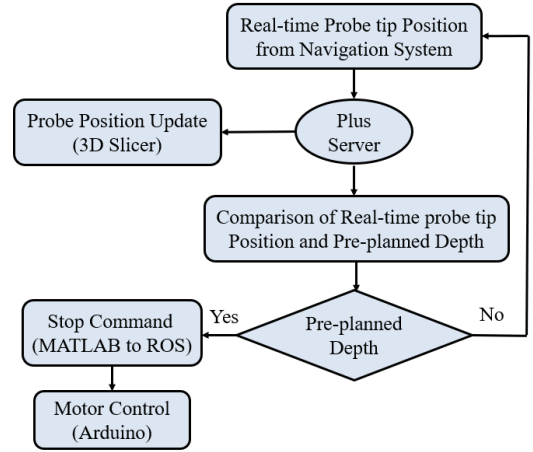


Fig. 3: Diagram of the communication of the proposed navigation system.

TABLE I: Experimental results of the target deviation and angular deviation errors.

	Frontal	Transversal	Sagittal
Target deviation (mm)	1.34 (± 0.59)	0.91 (± 0.40)	0.88 (± 0.65)
Angular deviation ($^\circ$)	1.31 (± 0.87)	0.93 (± 1.55)	6.38 (± 6.10)

display to position the probe and not on the path displayed on the hologram of the vertebra probably considering the difficulty in assessing the depth of the virtual model.

CONCLUSIONS

In this paper, an AR-based robotic navigation system is developed for spinal surgery. Experimental results demonstrated the possibility of integrating the AR-based instrumentation into operating rooms for spinal surgery navigation. Participants can position correctly the robotic prototype by looking at the virtual display shown in the AR application and at the path seen on the hologram. Also, in all the trials the automatic drilling can activate and stop successfully which suggests a possible solution that could be used to automatize processes such as the drilling one in spine surgery.

REFERENCES

- [1] C. M. Andrews, A. B. Henry, I. M. Soriano, M. K. Southworth, and J. R. Silva, "Registration techniques for clinical applications of three-dimensional augmented reality devices," *IEEE journal of translational engineering in health and medicine*, vol. 9, pp. 1–14, 2020.
- [2] Y. Hu, H. Jin, L. Zhang, P. Zhang, and J. Zhang, "State recognition of pedicle drilling with force sensing in a robotic spinal surgical system," *IEEE/ASME Transactions On Mechatronics*, vol. 19, no. 1, pp. 357–365, 2013.
- [3] Y. Torun and Ö. Pazarci, "Parametric power spectral density estimation-based breakthrough detection for orthopedic bone drilling with acoustic emission signal analysis," *Acoustics Australia*, vol. 48, no. 2, 2020.
- [4] Y. Dai, Y. Xue, and J. Zhang, "Noncontact vibration measurement based thoracic spine condition monitoring during pedicle drilling," *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 5, pp. 1532–1540, 2013.

Surgical image generation for binary tool segmentation

Emanuele Colleoni and Danail Stoyanov

Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London (UCL), UK

INTRODUCTION

Introducing artificial data to train deep learning models is common practice among fields that have limited access to information resources such as the medical and surgical fields [1] due to legal and ethical constraints. To this end, image-to-image translation (*I2I* [2]) models demonstrated high suitability for this task. *I2I* is the process of translating data between different domains: this includes a variety of tasks such as image colorization or simulation to real image adaptation. In this abstract we focus on the latter task to produce synthetic frames starting from a robot simulation environment to train surgical tool segmentation models. Simulation environments are indeed one of the preferred choices as a starting domain for the application of *I2I* thanks to their ability in producing automatic labels [3]. Differently from previous works [1], [4], [5], our model include attention modules [6] to support feature translation between domains and it is able to optimally blend the simulation tools on the surgical background, thus increasing their realistic appearance. Validation of the proposed method was conducted through evaluation of segmentation models trained on synthetic images produced using state-of-the-art methods and the proposed one. Here we show that performances close to training on real data can be achieved with the proposed method on 3 different datasets, with improvements from 2% to 10% intersection-over-union (*IoU*) over other image generation pipelines.

MATERIALS AND METHODS

A. Model architecture

Based on recently developed attention mechanisms, we built our *I2I* architecture similar to [6], where segmentation models are trained along each generator to help the model focus on domain-specific features. The architecture is composed of two generators and two attention modules generate synthetic frames from simulation and real images and two discriminators for to distinguish between real and fake frames. Details of our architecture's components are presented as follows:

1) *Generators*: Each generator was built following Cycle-GAN architecture[2]: a 7x7 convolution layer followed by two down-sampling layers with kernel size $K=3$ and stride $S=2$ reduce the image dimensions and increase the number of channels. Then a series of 8 residual blocks followed by two up-sampling convolutional layers further processes the image and return it with its original dimension.

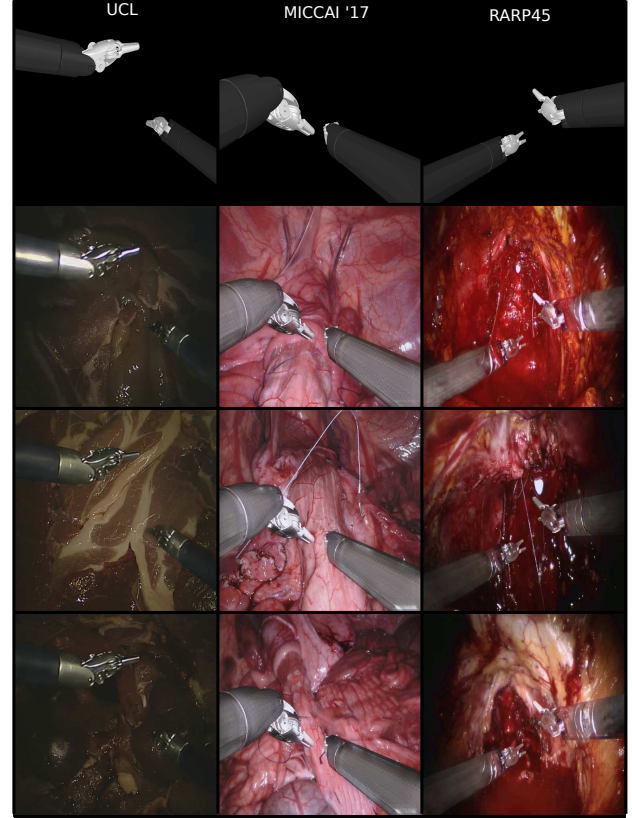


Fig. 1: Sample images from the proposed method generated starting from MICCAI '17, UCL and RARP45.

2) *Attention modules*: Our attention modules were built using a U-net architecture with a ResNet34 backbone pre-trained on Imagenet. We choose a very deep model compared to the original formulation [6] since pre-trained backbones have shown to improve results in segmentation procedures compared to basic U-net structure, thus we believe better image understanding can derive in the same way for the proposed attention module.

3) *Discriminators*: Our discriminators follow the PatchGAN [7] paradigm: the model is fed with one frame and returns a 30x30 gray-scale matrix where each pixel evaluates a 70x70 patch of the input frame to be either fake or real. This method has been proved to allow a lighter computation without any important drop in performances.

B. Domain sets

Domain sets refer to the source and target images used to train our *I2I* model. Since we aim to transform labelled simulation frames to have realistic stylish appearance, we

TABLE I: Segmentation performances for binary segmentation over *UCL*, *MICCAI '17* and *RARP45* test sets. Results are shown in the following form: Best *IoU* (%) / Average *IoU* (%) over three different training procedures for each model.

	<i>UCL</i>	<i>MICCAI '17</i>	<i>RARP45</i>
<i>No I2I</i>	39.52 / 36.25	5.23 / 3.41	0.0 / 0.0
<i>CoSegGAN</i>	84.04 / 75.76	89.33 / 85.21	54.46 / 52.81
<i>Ozawa et al.</i>	76.37 / 72.94	91.83 / 90.34	33.81 / 30.69
<i>Colleoni et al.</i>	85.67 / 84.85	91.15 / 90.16	-
<i>Proposed</i>	86.33 / 85.26	93.32 / 93.02	64.39 / 57.72
<i>Real</i>	85.93 / 84.06	96.09 / 96.01	-

use a simulation and a real datasets as source and target domains respectively:

1) *Real domain set*: Differently from [1], we built our real domain set considering frames from ex-vivo and in-vivo surgical procedures without further pre-processing. The employed datasets are *MICCAI '17* from Endovis Challenge [8], *UCL* from [9] and *RARP45* an in-vivo surgical dataset from real prostatectomies.

2) *Simulation domain set*: We composed simulation domain frames by blending simulated surgical tools from a Coppeliasim¹ [10] da Vinci robot onto real surgical background images obtained from real surgical videos. We considered only EndoWrist® Large Needle Drivers (LNDs) due to the limited availability of 3D simulation surgical tool models in Coppeliasim environment, though we believe that the procedure could be extended to other instrument shapes.

C. Experiments

In order to evaluate our model against state-of-the-art image generation works, we trained our and other three models from literature to produce images using *MICCAI '17*, *UCL* and *RARP45* datasets as real domain set. Then, for each model, we generated a segmentation dataset (using segmentation labels from simulation) and trained a Unet on that. Finally, we compared their segmentation performances, with the assumption that higher quality images lead to better segmentation. Each model was trained three times on each dataset and results will be provided in terms of mean and best *IoU* over the three models. We considered [4], [1] and [5] for our comparison.

RESULTS

Some qualitative results from the proposed workflow can be appreciated in Fig.1 while segmentation performances for all the considered methods are shown in Table I. The segmentation models trained on images produced with our workflow outperformed all models trained on frames generated using other pipelines, with the exception for *Real*, i.e. the model trained on real images from the datasets. The best difference was achieved on

RARP45, with an improvement of 10% best *IoU* and on *MICCAI '17*, with an improvement of 3% mean *IoU*.

CONCLUSIONS AND DISCUSSION

Fig.1 shows that our *I2I* model was able to translate features on simulation tools and blend them on the provided surgical background. Moreover, the luminance conditions on the tools vary according to the background, thus producing more realistic images. Moving to Table I and following experiments carried out in [1], the obtained results suggest that, with the proposed method, it is possible to produce surgical images that better mirror real surgical frames. This affects the segmentation performances of the model that uses these frames during training, leading to enhanced segmentation capabilities. For future research in this field, the potential of surgical simulators could be exploited to produce different sources of labels, e.g. surgical tool pose or depth maps. This is not possible in other methods [5] that are limited to surgical image segmentation. *I2I* paradigm could also be translated over surgeon's training environments to improve simulation realism. However, in this scenario, further effort should be put to ensure spatial and temporal consistency of the anatomical structures and the surgical tools.

REFERENCES

- [1] E. Colleoni and D. Stoyanov, "Robotic instrument segmentation with image-to-image translation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 935–942, 2021.
- [2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [3] M. Denninger, M. Sundermeyer, D. Winkelbauer, Y. Zidan, D. Olefir, M. Elbadrawy, A. Lodhi, and H. Katam, "Blenderproc," *arXiv preprint arXiv:1911.01911*, 2019.
- [4] T. Ozawa, Y. Hayashi, H. Oda, M. Oda, T. Kitasaka, N. Takeshita, M. Ito, and K. Mori, "Synthetic laparoscopic video generation from machine learning-based surgical instrument segmentation from real laparoscopic video and virtual surgical instruments," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 9, no. 3, pp. 225–232, 2021.
- [5] M. Kalia, T. A. Aleef, N. Navab, P. Black, and S. E. Salcudean, "Co-generation and segmentation for generalized surgical instrument segmentation on unlabelled data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 403–412.
- [6] Y. Alami Mejjati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, "Unsupervised attention-guided image-to-image translation," *Advances in Neural Information Processing Systems*, vol. 31, pp. 3693–3703, 2018.
- [7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [8] M. Allan, A. Shvets, T. Kurmann, Z. Zhang, R. Duggal, Y.-H. Su, N. Rieke, I. Laina, N. Kalavakonda, S. Bodenstedt *et al.*, "2017 robotic instrument segmentation challenge," *arXiv preprint arXiv:1902.06426*, 2019.
- [9] E. Colleoni, P. Edwards, and D. Stoyanov, "Synthetic and real inputs for tool segmentation in robotic surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 700–710.
- [10] G. A. Fontanelli, M. Selvaggio, M. Ferro, F. Ficuciello, M. Venditelli, and B. Siciliano, "A v-rep simulator for the da vinci research kit robotic platform," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 2018, pp. 1056–1061.

¹<https://www.coppeliarobotics.com/>

Spatio-temporal models for online activity recognition in cataract surgical videos

Sanat Ramesh^{1,2}, Diego Dall'Alba¹, Paolo Fiorini¹, and Nicolas Padoy^{2,3}

¹University of Verona, Verona, Italy, ²ICube, University of Strasbourg, CNRS, France, ³IHU Strasbourg, France

INTRODUCTION

Automatic recognition of surgical activity could play a key role in providing advanced support in computer-assisted interventions and autonomous functionalities in robot-assisted surgeries. There are different types of activities that are defined based on the level of detail, that could describe a surgical procedure. Katic et al. [1] decomposed a surgical workflow into different types of activities: whole procedure, phases, stages, steps, and actions. Phase recognition is a very active research area in the medical computer vision community and has received significant attention [2]. Recently, there has been a surge of research works focusing specifically on step recognition [3], [4]. This trend can be related to the fact that steps define a surgical workflow at a more fine-grained level than phases which could be beneficial in developing more precise and advanced monitoring and support systems.

Most recent works in step recognition from endoscopic surgical videos [5], [4] proposed a deep learning model consisting of a ResNet-50 backbone to learn spatial information and a Temporal Convolutional Network (TCN) to learn the temporal information from spatial features extracted from the backbone. Ramesh et al. [4] trained the spatial and temporal model separately for the recognition of surgical phases and steps from videos of Laparoscopic Roux-en-Y gastric bypass (LRYGB) procedures.

In this work, inspired by the work described in [4], we propose an end-to-end Spatio-temporal model combining ResNet-50 and TCN models for step recognition in cataract procedures. We evaluate the proposed model for step recognition on the public benchmark CATARACTS containing 50 videos of cataract surgeries.

MATERIALS AND METHODS

The proposed Spatio-temporal model consists of a ResNet-50 [6] backbone for learning spatial features and a single-stage TCN (SS-TCN) for temporal modeling from the spatial features. The model is shown in Fig. 1.

We use SS-TCN that is a single-stage variant of Multi-Stage TCN (MS-TCN) [5] to learn the spatial coherence across video frames. The choice of SS-TCN was motivated by the work of [4] where MS-TCN did not provide a significant improvement over SS-TCN for both the step and phase recognition. The SS-TCN, following the same design of MS-TCN, contains neither pooling layers nor fully connected layers and is constructed with only temporal convolutional layers, specifically dilated residual layers performing dilated convolutions. With the aim of online activity segmentation, we perform at each layer causal

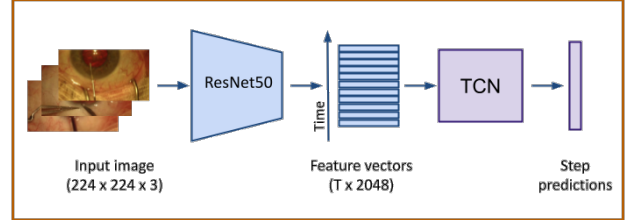


Fig. 1. The architecture of the proposed spatio-temporal model consisting of a ResNet-50 backbone followed by a TCN for temporal modeling.

convolutions that depend only on the current frame and n previous frames. The proposed model is trained end-to-end as opposed to two stage approach proposed in [4].

The model takes an input video consisting of T frames $x_{1:T}$. The ResNet-50 maps $224 \times 224 \times 3$ RGB images to a feature space of size $N_f = 2048$. The frame-wise features, collected over time, are input to the TCN model that predicts $\hat{y}_{1:T}^s$ where \hat{y}_t^s is the class label for the current timestamp t , $t \in [1, T]$. Step recognition being a multi-class classification problem that exhibits imbalance in the class distribution, softmax activation and class-weighted cross-entropy loss are utilized. The class weights are calculated using the median frequency balancing [7].

Training: The ResNet-50 model is initialized with ImageNet pre-trained weights. The end-to-end ResNet-50 + SS-TCN model is then trained for the task of step recognition. Since TCN models capture the temporal information in an online setup, features of past frames of a video are cached using a feature buffer which is reset at the end of the video. The model in all the experiments is trained for 50 epochs with a learning rate of $1e-5$ and a batch size of 64. The models were implemented in PyTorch.

Dataset: We evaluate our method on the public CATARACTS dataset proposed in [8]. The dataset consists of 50 videos of cataract surgery with step annotations. In total, there are 19 steps defined for this procedure¹. The list of steps is presented in Tab. I. We split the 50 videos into 25, 5, and 20 videos for training, validation, and test sets, respectively. The images of resolution 1920×1080 are extracted at 1 fps from the videos. The frames are resized to 224×224 and the training set is augmented with horizontal flip, saturation, and rotation.

Evaluation metrics: We follow the same evaluation metrics used in other related publications [5], [4], where

¹<https://www.synapse.org/#!Synapse:syn21680292/wiki/601563>

TABLE I

THE LIST OF STEPS DEFINED FOR CATARACT PROCEDURES.

ID	Step	ID	Step
0	Idle	10	Irrigation/Aspiration
1	Toric Marking	11	Preparing Implant
2	Implant Ejection	12	Manual Aspiration
3	Incision	13	Implantation
4	Viscodilatation	14	Positioning
5	Capsulorhexis	15	OVD Aspiration
6	Hydrodissection	16	Suturing
7	Nucleus Breaking	17	Sealing Control
8	Phacoemulsification	18	Wound Hydration
9	Vitrectomy		

accuracy (ACC), precision (PR), recall(RE), and F1 score (F1) are used to effectively compare the results.

DISCUSSION AND CONCLUSION

Although we use a fully annotated dataset for evaluating our proposed method, obtaining large-scale annotated dataset is very expensive and time-consuming. Especially in the medical domain as it requires annotators with expert domain-specific knowledge. To show the effectiveness of our method at different labeled data regimes, we test our Spatio-temporal model on the CATARACTS dataset with varying number of labeled videos (3, 6, 12, 18, and 25 videos) in the training set. The results on the test set are reported in Tab. II. We notice that with 100% of labeled dataset the model achieves 79.23% and 56.61% in accuracy and F1, respectively. In Fig. 2 we see that having more labeled data is beneficial for the model to recognize steps. Particularly with just six annotated videos, we achieve an accuracy of 64.48% which is 15% below 100% of the dataset annotated. Furthermore, with just 50% of the labels, we can achieve a result that is just 5-6% below 100% labeled dataset in all the evaluation metrics. The results show that the proposed Spatio-temporal model works well in recognizing activities from cataract surgical videos that are different compared to LRYGB procedures.

Although the proposed method shows good performance at different percentages of available labeled data, there's plenty of room for improvement in the lower data regime. Creating huge labeled datasets, which is crucial for training deep learning models, is very expensive and time-consuming. Hence, we would like to tackle this problem, as part of future work, by developing semi-/self-supervised learning methods that work with few labeled videos.

ACKNOWLEDGEMENT

This work was supported by the ATLAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813782. This work was also partially supported by French State Funds managed by the "Agence Nationale de la Recherche (ANR)" through the "Investissements d'Avenir" (Investments for the Future) Program under Grant ANR-10-IAHU-02 (IHU-Strasbourg).

TABLE II

SPATIO-TEMPORAL MODEL PERFORMANCE WITH VARYING AMOUNT OF LABELED VIDEOS. ACCURACY (ACC), PRECISION (PR), RECALL (RE), AND F1-SCORE (F1) (%) ARE REPORTED.

# Videos	ACC	PR	RE	F1
3 (12%)	43.55	35.9	23.63	24.67
6 (25%)	64.48	47.87	37.19	40.08
12 (50%)	74.62	59.19	47.49	50.6
18 (75%)	78.94	63.62	54.86	57.22
25 (100%)	79.23	67.33	54.39	56.61

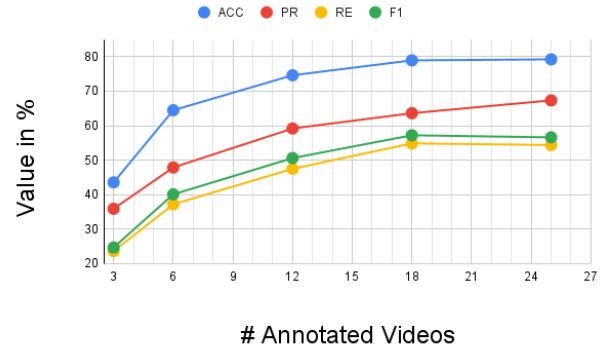


Fig. 2. Model evaluation on varying amount of step annotated videos.

REFERENCES

- [1] D. Katić, C. Julliard, A.-L. Wekerle, H. Kenngott, B. P. Müller-Stich, R. Dillmann, S. Speidel, P. Jannin, and B. Gibaud, "LapOntoSPM: an ontology for laparoscopic surgeries and its application to surgical phase recognition," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 9, pp. 1427–1434, Jun. 2015.
- [2] I. Rivas-Blanco, C. J. Pérez-Del-Pulgar, I. García-Morales, and V. F. Muñoz, "A review on deep learning in minimally invasive surgery," *IEEE Access*, vol. 9, pp. 48 658–48 678, 2021.
- [3] K. Charrière, G. Quellec, M. Lamard, D. Martiano, G. Cazuguel, G. Coatrieux, and B. Cochener, "Real-time analysis of cataract surgery videos using statistical models," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22 473–22 491, May 2017.
- [4] S. Ramesh, D. Dall'Alba, C. Gonzalez, T. Yu, P. Mascagni, D. Mutter, J. Marescaux, P. Fiorini, and N. Padoy, "Multi-task temporal convolutional networks for joint recognition of surgical phases and steps in gastric bypass procedures," *International Journal of Computer Assisted Radiology and Surgery*, May 2021.
- [5] T. Czempel, M. Paschali, M. Keicher, W. Simson, H. Feussner, S. T. Kim, and N. Navab, "TeCNO: Surgical phase recognition with multi-stage temporal convolutional networks," in *MICCAI*, 2020.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision – ECCV 2016*. Springer International Publishing, 2016.
- [7] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2650–2658.
- [8] H. A. Hajj, M. Lamard, P.-H. Conze, S. Roychowdhury, X. Hu, G. Maršalkaitė, O. Zisimopoulos, M. A. Dedmari, F. Zhao, J. Prellberg, M. Sahu, A. Galdan, T. Araújo, D. M. Vo, C. Panda, N. Dahiya, S. Kondo, Z. Bian, A. Vahdat, J. Bialopetravičius, E. Flouty, C. Qiu, S. Dill, A. Mukhopadhyay, P. Costa, G. Aresta, S. Ramamurthy, S.-W. Lee, A. Campilho, S. Zachow, S. Xia, S. Conjeti, D. Stoyanov, J. Armatiz, P.-A. Heng, W. G. Macready, B. Cochener, and G. Quellec, "CATARACTS: Challenge on automatic tool annotation for cataract surgery," *Medical Image Analysis*, vol. 52, pp. 24–41, Feb. 2019.

Prototyping a Laparoscopic Smart Trocar for Semi-Autonomous Surgical Robotics

Nicola Piccinelli*, Eros Ghignoni*, Matteo Piano,
Giacomo De Rossi*, Francesco Setti*, and Riccardo Muradore*

* *Department of Computer Science, University of Verona, Italy*

INTRODUCTION

The development of advanced robotic-assisted platforms for laparoscopic surgery necessarily requires advanced sensing techniques to provide data during surgery. In current laparoscopic interventions, accurate and precise measurements can be found primarily in complete robotic platforms that integrate actuated arms control and endoscopic camera positioning. These platforms, however, are expensive in both acquisition costs and maintenance and, as such, still represent a minority in all laparoscopy operations performed globally. In this paper, we present a solution to push forward the improved care offered by surgical robotics to laparoscopy with a *smart trocar* concept that allows accurate tracking of the instruments in minimally-invasive scenarios, like the one presented in [1].

Previous prototypes of a smart trocar either rely on only an accelerometer to integrate the orientation [2], with the latter operating only during the trocar placement and not for measurements during the procedure, or adopt a fixed-star navigation solution with a camera mounted on the trocar and targets scattered around the operating room [3].

The proposed smart trocar consists of two potentiometers with an RGB camera for estimating the locale position and orientation of the tip of the laparoscopic instruments. They have been developed under the EU funded SARAS project (<https://saras-project.eu/>) and have been tested on a scenario involving robot-driven instruments and realistic manikins to simulate a radical nephrectomy procedure. This solution requires no modifications to commercial trocars and needs minimal calibrations to perform accurately, thus offering a quick integration into traditional laparoscopic procedures.

MATERIALS AND METHODS

The design of the smart trocar, shown in Figure 1, integrates a sensorized platform with a standard set of laparoscopic trocars and tools. The tool diameter considered in this design is 5 mm, but it can be easily adapted to any trocar size. It is built around a gimbal and bridge component that transfers the pivoting motion of the trocar cannula holder to two potentiometers. A PCB is attached to the side of the cannula to host a microcontroller and a camera. This mounting position

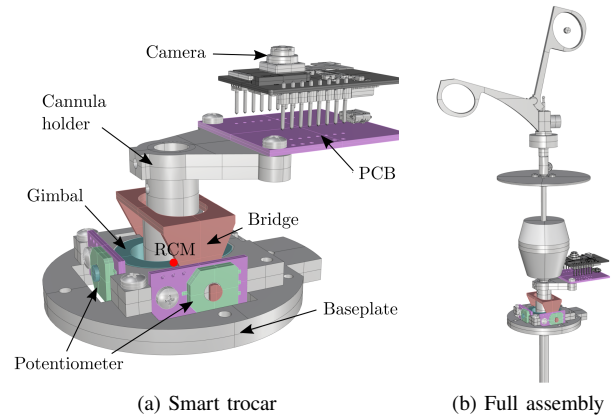


Fig. 1. CAD of the smart trocar final prototype. a) a close view of the base plate, b) a complete view with the tool and trocar mounted on it.

allows the camera to track a set of markers painted on a disk on the laparoscopic instrument which does not interfere with the trocar's motion. The camera is used to measure the rotation and insertion of the tool. Assuming a mounting position in contact with the patient epidermis, the effective Remote Center of Motion (RCM) is 9 mm above. This is not a major concern as the maximum radial deformation at the insertion point is ≈ 10 mm given the maximum longitudinal and transversal rotation of the gimbal and bridge of 40° and 64° , respectively. For this reason, the forces acting on the epidermis are negligible.

The microcontroller (an ESP32-CAM model) handles the raw measurements from the potentiometers via the PCB and synchronizes them with the camera frame acquisition. It communicates this information through WiFi via the *rosserial* protocol to a computer that estimates the pose of the laparoscopic tool tip (Tool Center Point, TCP). The proposed method computes the position and the orientation of the TCP T_e^r with respect to the mechanical RCM placed in the middle of the smart trocar base plate.

Let T_r^w be the transformation of the RCM with respect to a global reference frame; in our use-case we attached the smart trocar to a rigid stand as shown in Figure 2 and the transformation is estimated by using the algorithm proposed in [4]. T_e^r can be written as

$$T_e^r = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (1)$$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779813 (SARAS).

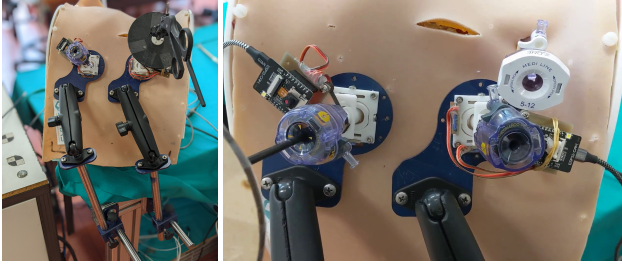


Fig. 2. The 3D printed final prototype of the smart trocar used during the experimental validation; the trocars are mounted on a pair of adjustable arms to adapt to the anatomy; the diameter of the smart-trocar cannula is 5 mm.

where $R = R(\psi)R(\theta)R(\phi)$ and $t = [0 \ 0 \ t_z]$. The position of the TCP T_e^w with respect to the global reference frame is then

$$T_e^w = T_r^w T_e^r. \quad (2)$$

The Euler angles ψ and θ are the values of the two potentiometers mounted on the board. We aligned the rotation axis of the sensors with a set of Cartesian axis having their z -axes aligned with the instrument shaft. To measure both the distance and the ϕ rotation we used a set of eight ArUcO markers [5] placed on a support disk anchored to the laparoscopic tool. The support has been designed in such a way to always have at least one of these markers visible by the camera mounted on the support. The distance is estimated by assuming that the camera and the disk remain parallel to each other which allows to average the depth of all identified ArUcO markers. Finally, the rotation ϕ is estimated by mapping each markers unique id to its position on the face-plate by design to zero the rotation and, then, measure the variation of the x and y axis angles relative to the camera. The ArUcO-based insertion measurements were found to be more precise and reliable than the early design that used a *time of flight* (ToF) sensor, with the added bonus of being able to also estimate the tool's rotation along its main axis.

EXPERIMENTS

The proposed smart trocar prototype, as shown in Figure 2, has been tested on an experimental setup with a Franka Emika[®] Panda arm acting as the ground truth position reference and a Storz[®] grasper laparoscopic tool as the surgeon instrument. We registered the robot and the smart trocar prototype in a shared reference frame [4] to later compare their measurements. We performed a sequence of motion towards a fixed position in an anatomical manikin having an impedance similar to the human skin at the RCM. Then, we recorded the target positions using the robot kinematics and we compared them with the measured positions of the laparoscopic TCP over multiple tests. Figure 3 shows the prototype performing in a set of three puncturing. The contacts happen around 10 s, 18 s and 26 s and the maximum error is about 4 mm. Such error is deemed sufficient for the intended application. It is the authors' opinion that such error can be mostly attributed

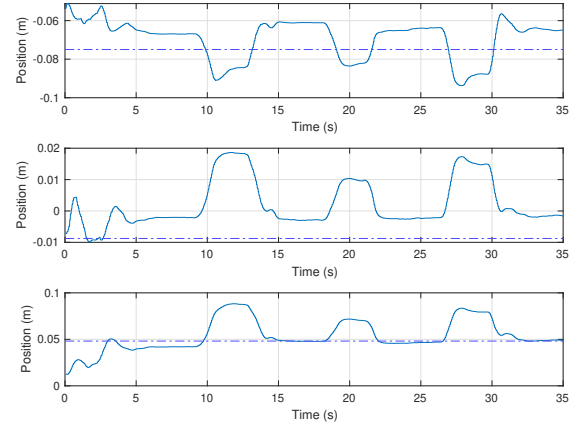


Fig. 3. The experimental validation of the proposed prototype: the solid line shows the laparoscopic tool position and the dashed line shows the reference position of the landmarks.

to the mechanical play related to the materials used in the prototype.

CONCLUSIONS AND DISCUSSION

The design of the smart trocar proved to be simple yet efficient in providing a reliable pose estimation despite the low cost construction of the prototype version. The validation on the radical nephrectomy procedure carried out with expert surgeons were successful with no reported obstruction on the nominal procedure flow. The position measurements as shown in the experiments appear to be stable and sufficiently precise for controlling robotic manipulators in semi-autonomous surgery. The adopted mounting solution of the smart trocar is stable and reliable, but the design ergonomics should be further improved.

REFERENCES

- [1] F. Falezza, N. Piccinelli, G. De Rossi, A. Roberti, G. Kronreif, F. Setti, P. Fiorini, and R. Muradore, "Modeling of surgical procedures using statecharts for semi-autonomous robotic surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 4, pp. 888–899, 2021.
- [2] J. Schroppe, B. Olmanson, C. Fick, C. Motameni, T. Viratyosin, Z. D. Miller, J. Harmon, and P. Emerson, "The smart trocar: Force, deviation, and impedance sensing trocar for enhanced laparoscopic surgery," in *Frontiers in Biomedical Devices*, vol. 41037. American Society of Mechanical Engineers, 2019, p. V001T06A002.
- [3] G. Joerger, A. Huang, B. Bass, B. Dunkin, and M. Garbey, "Global laparoscopy positioning system with a smart trocar," in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2017, pp. 359–366.
- [4] A. Roberti, N. Piccinelli, D. Meli, R. Muradore, and P. Fiorini, "Improving rigid 3-d calibration for robotic surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 4, pp. 569–573, 2020.
- [5] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

Design, realization and preliminary validation of an active physical simulator for the study of pelvic floor damages during childbirth

G. Modarelli¹, S. Maglio², S. Tognarelli², A. Menciasci²

¹*University of Pisa, Pisa, Italy*

²*The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa*
giuliam93@live.it

INTRODUCTION

Pelvic floor (PF) is a muscle-fascial system that inferiorly closes the pelvis. It is really important in woman's life because it guarantees the correct positioning of the pelvic organs, the urines and feces continence and evacuation, and it allows sexual activity and childbirth [1]. Dysfunctions of PF involve important physical and psychological consequences in the daily life of a woman and significant expenses for their treatment.

Vaginal delivery represents one of the main risk factor for the onset of PF disorders. Pelvic muscles in pregnant women show very different elastic characteristics if compared with non-pregnant women's muscles. Moreover, they undergo considerable stretching during the passage of the foetus, that often results in a medium and/or long-term tissues injury. Thus, dedicated studies could both pave the way to a better comprehension of this phenomena, and improve prevention clinical techniques. Physical simulators deepen the study of the factors characterizing a clinical event and allow the doctor to have a physical support, which is essential for technique refinement and knowledge transfer in the clinical practice. Currently, there are only birth simulators available on the market, in which the PF is merely an additional element recreated at a low-fidelity level in terms of anatomical and physiological features. In addition, commercial simulators equipped with PF are limited in number and entirely passive, hence they don't provide any feedback to the clinician.

In this framework, the aim of this paper is to realize a sensorized physical simulator of maternal PF that can be used both as a teaching and/or a training system for gynaecologists and obstetrics. The innovative features of the proposed simulator are the following: *i*) high-fidelity reconstruction of the maternal PF anatomy; *ii*) use of soft materials able to replicate the biomechanical properties of human tissues; *iii*) active evaluation of the muscle deformation induced by the foetal head (FH) passage through the PF structure.

MATERIALS AND METHODS

By combining anatomical data with literature information, a 3D model of the maternal PF was reconstructed and utilized for muscle fabrication through moulding technique (Figure 1). Ecoflex 0030 silicone (Smooth-On, PA, USA) was used for reproducing the biomechanical behaviour of the PF muscles. Its elastic modulus of $E=0.068$ MPa is close to the pregnant woman muscles modulus, i.e., $E=0.06$ MPa [2, 3]. The tendons

structures of the PF are then replicated for guaranteeing the right positioning of the muscle component (Figure 1 for details). Finally, a female pelvis model with standard dimensions was found for bones replication [4].

In order to record the muscle deformation during the delivery simulation, elastic sensors capable to follow the elongation of the PF are required. Commercial sensors were discarded because commonly they are rigid and thus they do not meet the silicone elasticity and integration requirements. As an alternative, a specific elastic conductive fabric (Med-tex P130, Shieldex, Germany) was investigated for fabricating a custom home-made resistive sensor. Conductive fabric has been already demonstrated as a valid solution for soft applications [5], allowing to evaluate the elongation by changing its resistance. Finally, signal acquisition is ensured by firm electrical connections made of conductive wires properly hand-stitched on the fabric mesh. Electronic components for signal conditioning and acquisition are required. To evaluate the sensors resistance-elongation relation, several tensile tests were performed by using the Instron 5965 machine (Instron, USA). Rectangular shaped textile samples were stretched from 0% to 100% - in order to cover the entire human tissue stretch range - for 5 cycles at 400 mm/min speed. Test was repeated 3 times.

The components were integrated and the final device was validated both on the bench and with clinicians. In order to calibrate the simulator and to collect data that can be used for comparing the obtained system with literature, a descent of the FH through the PF was simulated by using the Instron machine. A vertical descent of 100 mm at 150 mm/min speed was repeated 3 times; applied forces and sensor elongation were acquired from the Instron and simulator electronics, respectively. As in the literature, it was used a rigid FH approximating a 10 cm sphere made of PETG realized by means of the FDM technique. Ultimately, to complete the preliminary system validation phase, a pre-clinical testing protocol was realized by involving expert gynaecologists at Azienda Ospedaliero Universitaria Pisana. The tests were performed by combining our simulator with a commonly used FH simulator. The descent of the FH through the PF was reproduced. Five different positions that the FH might assume during the delivery were simulated by the clinicians. Tests were repeated 3 times.

RESULTS

The integrated system is reported in Figure 1. Pelvis bones were made of PETG through a FDM printer (i3 MK3S+, Prusa Research, Czech Republic).

Three different sensors were made in order to cover the muscular areas mostly stressed during childbirth: 2 rectangles of 51mmx6.5mm in size were positioned between the pubic symphysis and the perineal body; 1 smaller rectangle of 44mmx6.5mm was integrated near the perineal body (Figure 1-bottom view). Sensors signal was acquired using a Wheatstone bridge. The electronic circuit was powered and managed through the Arduino Mega 2560 board (Arduino, Italy).

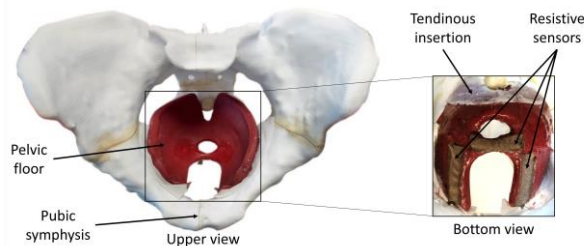


Figure 1. Complete simulator. On the left upper view and on the right bottom view.

By combining the elongation applied by the Instron machine on the sensors samples with their resistance values, the sensor calibration curve was obtained (Figure 2). As shown in the figure, the resistance increases linearly up to 60% of the elongation and decreases linearly from 60% to 100%. An equal and opposite behaviour was observed during the shortening phase of the sensor. The acquired signal proved to be accurate, showing an identical behaviour over time and during multiples tests.

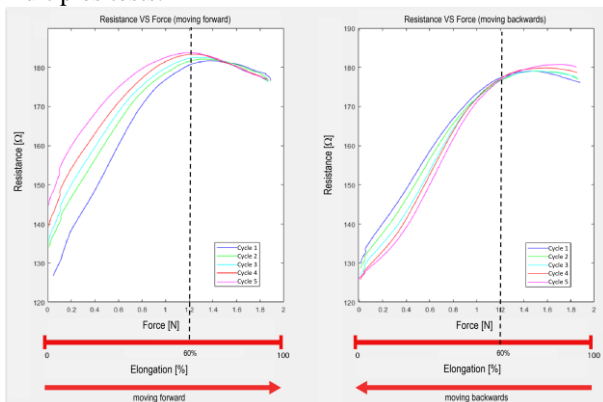


Figure 2. Resistance VS Force for sensor force calibration curves obtained moving forward (on the left) and moving backward (on the right).

The resistive sensors, made with Med-tex P130 fabric, were integrated on the PF structure using Ecoflex 0010 silicone. Muscles and bones were glued together through a thin inextensible fabric replicating PF tendons. The forces acquired during the test simulating FH descent with the Instron machine (i.e., 45N as peak force) are similar to those found in literature (i.e., 30 N as peak force) [6]. Moreover, the sensor resistance-elongation curve acquired with the final system, show the same behaviour obtained for the single sensor, demonstrating the system robustness and the analysis repeatability. From the tests with three expert gynaecologists (Figure 3), it was possible to evaluate which one of the

reproduced foetal positions (e.g., occiput anterior (OA) position - baby facing towards mamma's spine; occiput posterior (OP) position - baby's head is entering face-up) causes greater muscle distension, resulting in an increased possibility of PF injury and/or the onset of PF dysfunction in the postpartum period.

The obtained results (Figure 3A) confirmed that all the integrated sensors worked correctly during the tests. In addition, sensors provided different output signals, showing different peak values (point of instantaneous main contact between FH and PF) at different descent instants.

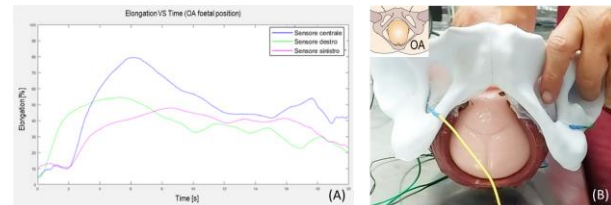


Figure 3. Test conducted with expert gynaecologist. A) Elongation VS Time curves extracted from the integrated sensors; (B) Testing set-up.

CONCLUSION AND DISCUSSION

The described active high-fidelity simulator showed good potentialities and it could be considered a valid system for PF damages evaluation. Clinicians confirmed the simulator is anatomically and physiologically faithful to human anatomy. Resistive textile sensors allowed to monitor the elongation of the PF tissues during the FH descent. The device, therefore, represents a promising tool that can be included both in training courses for experienced and non-experienced clinicians, and in gynaecological education lessons. Future efforts will be dedicated to a structured test protocol with at least 40 trainees (expert and residents). Additionally, a simple solution for fixing the system to the gynecological table during the test protocol will be identified for avoiding to affect the experimental data with drawbacks due to a not-stable manual held.

REFERENCES

- [1] G. Bolis, "Manuale di Ginecologia e Ostetricia," II. EdiSeS.
- [2] S. Badir *et al.*, "A novel procedure for the mechanical characterization of the uterine cervix during pregnancy," *J. Mech. Behav. Biomed. Mater.*, vol. 27, pp. 143–153, 2013.
- [3] J. Lepage *et al.*, "Biomechanical pregnant pelvic system model and numerical simulation of childbirth: impact of delivery on the uterosacral ligaments, preliminary results," *Int. J. Urogynecology J.*, vol. 26, no. 4, pp. 497–504, 2015.
- [4] "Sketchfab." <https://sketchfab.com/3d-models/female-human-pelvis-9c4180fe752e4340a2ea90d5a8ad2bf3>.
- [5] A. Atalay *et al.*, "Batch Fabrication of Customizable Silicone-Textile Composite Capacitive Strain Sensors for Human Motion Tracking," *Adv. Mater. Technol.*, vol. 2, no. 9, pp. 1–8, 2017.
- [6] M. E. T. Silva *et al.*, "Study on the influence of the fetus head molding on the biomechanical behavior of the pelvic floor muscles, during vaginal delivery," *J. Biomech.*, vol. 48, no. 9, pp. 1600–1605, 2015.

End-of-instrument Stiffness Perception Device Based on Sensory Augmentation for Laparoscopic Surgery

Charl  lie Saudrais¹, Bernard Bayle², Marie-Aude Vitrani¹, and Fabien Verit  ¹

¹ ISIR, UMR 7222 Sorbonne University, CNRS, U1150 INSERM, Paris, France,

² ICube, UMR 7357 University of Strasbourg, CNRS, INSA, Strasbourg, France
fabien.verite@isir.upmc.fr

INTRODUCTION

Laparoscopic surgery brings a lot of advantages for the patient [1] counterbalanced by an increased difficulty for the surgeon, caused in part by limitations in perception [2]. On the one hand, the visual perception is degraded (loss of depth perception, reduced field of view, etc.), on the other hand, due, among other, to the friction of the trocar, the stiffness of the abdominal wall and the lever effect, the haptic perception of forces is also highly degraded [3]. Since both haptic and visual sensory modalities are degraded, the overall force perception, resulting in the integration of both, is insufficient [4]. It leads to tissue damages [5], application of excessive force [6] and poor palpation precision (in terms of hardness and size discrimination for instance) [7].

Therefore, it appears of great interest to reconstitute pertinent information to the surgeon, to improve the perception of forces. We propose here to use sensory augmentation to assist the degraded natural haptic feedback with a sensory substitution approach to reconstitute the effort measured at the tool-tip. This feedback has to respect several constraints. First, it should not interfere with the rest of the operating room. We therefore opt for the tactile modality so as not to increase the load on the already highly solicited auditory and visual modalities [8]. In particular, a tactile feedback based on tangential skin deformation – skin stretch – appears as a suitable candidate; Indeed, it has been suggested [9] that as it is the same sensory channel as for the perception of an effort, it should then be intuitive with no increase of the cognitive load. Finally, in order to avoid disrupting the medical gesture, we decide to apply the skin deformation feedback on the arm instead of the fingertip, as usually done in teleoperation [10] or haptic interaction with virtual environments [9].

To summarize, the present paper focuses on sensory augmentation based on tangential skin deformation on the arm to reconstitute the force applied at the end of the instrument to the surgeon. The design of a first prototype and a pilot study aiming at experimentally assessing the validity of the proposed approach in a 1-Dof context, focusing only on the axial force, are presented hereinafter.

MATERIALS AND METHODS

A. Prototype: Design & Control

We developed a skin-stretch wristband, shown in Fig. 1, in order to evaluate the proposed feedback strategy. The

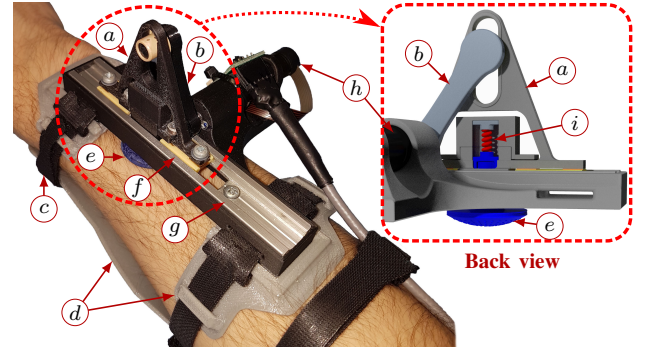


Fig. 1. Skin-stretch wristband: (a) slider, (b) crank, (c) hook-and-loop fasteners, (d) soft TPU pads, (e) tactor, (f) low-friction carriage, (g) guide rail, (h) motor, (i) compression spring

device consists of a 1DoF mechanism based on the Scotch-Yoke variant of the crank-slider, as this variant is more reliable, decreases the dimensional requirements [11] and presents less friction [12]. The element in contact with the skin providing the friction-induced skin-stretch is the blue 25 mm diameter spherical convex tactor. It is made out of high friction silicone to avoid slippage and it is connected to the slider with a spring-based mechanism to ensure the contact with the skin along the normal direction. The slider, hence the tactor, features a stroke of ± 30 mm from the center position. The wristband is positioned and maintained on the arm using soft TPU pads strapped together. All the control and instrumentation signals are connected to a single board computer, running a real-time controller. The latter is performing the position control of the motor and acquiring the force measured by a 6-axis F/T sensor.

The amount of skin stretch to be applied is determined similarly to [9], [10] as follows:

$$x = r_{ss} f, \quad (1)$$

where x [m] is the desired displacement of the tactor, r_{ss} [m/N] is the skin stretch ratio and f [N] is the force to be rendered. The desired displacement of the tactor is then converted into a desired motor angle using the kinematic model.

B. Protocol: Design & Experimental Setup

We performed an experimental pilot study in order to evaluate our approach and compare it to a visual feedback and no feedback, in a stiffness discrimination palpation task, mimicking a laparoscopic surgery context. A total

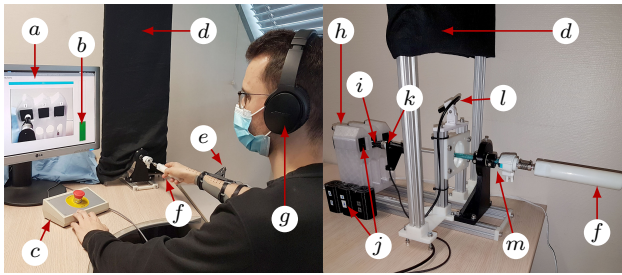


Fig. 2. Experimental setup of the pilot study

Left: (a) view of the scene, (b) force gauge, (c) emergency stop, (d) opaque sheet, (e) wristband, (f) tool, (g) noise-canceling headphone
Right: (h) stiffness sample holder, (i) spherical palpation probe, (j) stiffness samples, (k) F/T sensor, (l) camera, (m) fulcrum + trocar

of 5 subjects between the ages of 24 and 32 participated in this study, which was performed in accordance with the ethical principles provided by the declaration of Helsinki.

The experimental setup is presented in Fig. 2. The participant is holding a palpation tool inserted in a medical trocar, itself placed in a fulcrum. The end of the instrument consists of a 8mm spherical probe, mounted on a 6 axis F/T sensor, allowing to measure the applied efforts. However, only the force along the tool axis is considered here. Three stiffness samples are placed in a sample holder. The vision of the scene, blocked by an opaque black sheet, is transferred on a screen.

The elementary task is a palpation task in which the participant is presented with three stiffness samples where one, the *odd-one-out*, is stiffer than the other two, the *references*, which are identical. The samples are masked on the screen when the tool-tip is in contact, so as not to influence the perception. The subject is given 15s to identify the stiffest sample.

This elementary task (ET) is performed under three experimental conditions: “*natural*” where there is no additional feedback than the distorted natural haptic feedback, “*visual*” where a gauge whose filling is proportional to the applied effort is displayed on the screen, “*skin stretch*” where the proposed feedback is applied.

For each experimental condition, there is one experimental block subdivided into two parts: a training part of 7 ET and a main part of 21 ET. The main part is composed of three difficulty levels (*easy*, *medium*, *hard*), with 7 ET per level. An increase in difficulty is characterized by a decrease in the difference in stiffness between the *reference* and the *odd-one-out*.

RESULTS

The performance in term of stiffness discrimination precision is evaluated using the success rate, that is, the percentage of correct identification of the stiffest sample. Fig 3 shows the success rate for all participants, for every difficulty level under each experimental condition.

These preliminary results seems to indicate that the participants are performing better with a feedback than with no feedback. Furthermore, if the *visual* and *skin-stretch* feedback appear to perform similarly for the *easy* level, it seems that, as the difficulty increases, the *skin-stretch* feedback is detaching from the *visual* feedback and yields a better success rate.

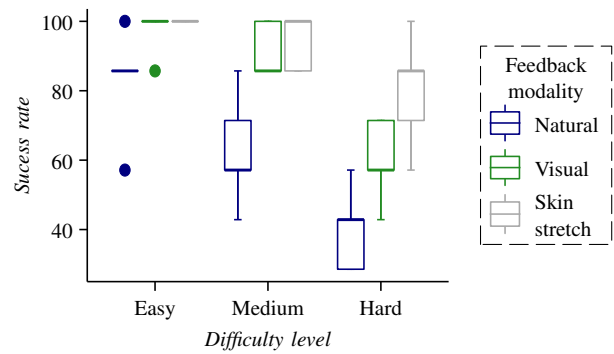


Fig. 3. Results in term of success rate for all participants

CONCLUSION AND DISCUSSION

This paper proposed a feedback modality based on tangential skin-stretch on the arm in order to improve the degraded perception of forces in laparoscopic surgery. The accent was put on the design of a prototype and on a pilot study that showed promising results, confirming the interest of the presented approach.

A complete study is therefore required and will be conducted in future work to confirm the emerging trend from these preliminary results using statistical tests on a larger population. The impact on the cognitive load also needs to be addressed to verify the assumptions established in introduction.

ACKNOWLEDGMENT

This work was supported by the *Investissements d’Avenir* program of the French Government, more specifically by Labex CAMI ANR-11-LABX-0004.

REFERENCES

- [1] V. Velanovich, “Laparoscopic vs open surgery: A preliminary comparison of quality-of- life outcomes,” *Surg. Endosc.*, vol. 14, no. 1, pp. 16–21, Jan. 2000.
- [2] J. Dankelman *et al.*, “Current technology on minimally invasive surgery and interventional techniques,” in *2011 2nd Int. Conf. Instrum. Control Autom.*, Nov. 2011, pp. 12–15.
- [3] E. P. Westebring – van der Putten *et al.*, “Haptics in minimally invasive surgery – a review,” *Minim. Invasive Ther. Allied Technol.*, vol. 17, no. 1, pp. 3–16, Jan. 2008.
- [4] H. Xin *et al.*, “Laparoscopic surgery, perceptual limitations and force: A review,” in *First Can. Stud. Conf. Biomed. Comput.*, 2006, pp. 44–46.
- [5] D. D. Marucci *et al.*, “Grasper Trauma During Laparoscopic Cholecystectomy,” *Aust. N.Z. J. Surg.*, vol. 70, no. 8, pp. 578–581, Aug. 2000.
- [6] P. Joice *et al.*, “Errors enacted during endoscopic surgery - A human reliability analysis,” *Appl. Ergon.*, vol. 29, no. 6, pp. 409–414, Dec. 1998.
- [7] M. V. Ottermo *et al.*, “The role of tactile feedback in laparoscopic surgery,” *Surg. Laparosc. Endosc. Percutan. Tech.*, vol. 16, no. 6, pp. 390–400, Dec. 2006.
- [8] S. Schostek *et al.*, “Review on aspects of artificial tactile feedback in laparoscopic surgery,” *Med. Eng. Phys.*, vol. 31, no. 8, pp. 887–898, Oct. 2009.
- [9] S. B. Schorr *et al.*, “Sensory substitution via cutaneous skin stretch feedback,” in *2013 IEEE Int. Conf. Robotics and Autom.*, May 2013, pp. 2341–2346.
- [10] Z. F. Quek *et al.*, “Evaluation of Skin Deformation Tactile Feedback for Teleoperated Surgical Tasks,” *IEEE Trans. Haptics*, vol. 12, no. 2, pp. 102–113, Apr. 2019.
- [11] M. Pramoth Kumar *et al.*, “Scotch-Yoke mechanism for a syringe pump - A case study,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 149, no. 1, p. 012221, Sep. 2016.
- [12] V. Arakelian *et al.*, “Design of Scotch yoke mechanisms with improved driving dynamics,” *Proc. Inst. Mech. Eng., Pt. K. J. Multi-body Dyn.*, vol. 230, no. 4, pp. 379–386, Dec. 2016.

Online detection of disturbances in model-based weight compensation for robotic endoscope holders

^{1,2}Jesus Mago, ¹Megane Millan, ³Constance Jacques, and ¹Marie-Aude Vitrani

¹ISIR F-75005, ²Haption SARL F-53210, ³SUMMIT F-75005

INTRODUCTION

Classic laparoscopy is an ergonomic burden for both assistants and surgeons. To ease this aspect, a robotic endoscope holder that features a model-based weight compensation was introduced in [1]. This feature allows the user to place the camera at a desired position in a comanipulation fashion. Once in place, the camera can be released, and the robot will hold it steady.

However, endoscopes feature difficult-to-model elements. For instance, the optic cable shown in Fig. 1 is a non-rigid element whose gravity model depends on the pose of the endoscope. If this variability is not accounted for, the user can no longer release the camera safely as it will present a fall/rise.

In the literature several approaches have been proposed to deal with variable payloads of robots [2], [3], [4]. However, they present drawbacks: reliance on additional hardware and/or on predefined end-effector trajectories.

This article proposes an online method to detect disturbances in the robot gravity model. By using Neural Network, the method can differentiate human manipulation from fall/rises due to poor weight compensations. This is a first step towards a complete approach that aims at an online adaptation of the endoscope's gravity model.

MATERIALS AND METHODS

A. Experiment

The overall setup used is the “Surgical Cockpit” [5]. This platform reproduces a realistic situation of robot-assisted minimally invasive surgery. For the experiment, a participant is asked to reproduce the target image shown on the small screen by manipulating the endoscope attached to the robotic arm (Fig. 1 on the left). The endoscope video is displayed on the large screen. Once the target image is reached, the endoscope can be released. A total of 6 objects are placed inside a laparotrainer. Target images are randomly drawn.

a) *Robot actuation*: The endoscope holder is Achilles, characterized in [6]. This robot features a passive gravity compensation of its links. For this experiment it also features a model-based algorithm, proposed by [6], to compensate endoscope weight. This algorithm consists in mapping a Cartesian force \mathbf{F}_g at the robot's end effector (Point W in Fig. 1) respecting (1).

$$\mathbf{F}_g = -m(\beta(\mathbf{I} - \mathbf{z}_I \mathbf{z}_I^T) + \mathbf{z}_I \mathbf{z}_I^T)g, \quad \beta = \frac{\overline{CG}}{\overline{TW}} \quad (1)$$

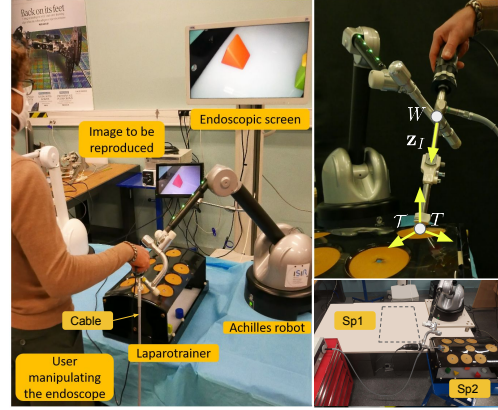


Fig. 1: Overall set-up. On the left is the general view, on the top right a focus on the endoscope, on the bottom right spatial configurations Sp1 and Sp2.

Where \overline{CG} is the distance from the camera tip C to its center of mass G , \overline{TW} the distance from the fulcrum point T to the wrist W , (using [6]), m the mass of the endoscope and \mathbf{z}_I a unit vector parallel to the endoscope's axis.

b) *Weight configurations*: Four configurations - A, B, C, and D - are set in which different values of mass m and percentages of optic cable hanging from the endoscope are tested. They are presented in Table I.

Values 0.5 kg and 0.65 kg correspond to the correct model mass to compensate for the endoscope alone and with the entire cable's length attached, respectively. Each configuration has compensation errors which will result in the endoscope either falling or rising after release.

c) *Datasets*: Two spatial configurations Sp. 1 and Sp. 2 (Fig 1, on the bottom right) were tested to evaluate how sensible the algorithm is to changes in the robot-patient configuration. A total of 200 releases in the configuration Sp.1 are performed and used as training set. The testing set comprises 60 releases for both spatial configurations (Sp.1 and Sp.2). For every set, weights configurations were randomly drawn before each release.

B. Detection Algorithm

In order to detect perturbations in the gravity model, several algorithms have been tested and Neural Networks have shown the best results. Furthermore, many features were investigated for the input, and only the following are kept, as they showed the best performance (Section -C):

- Position of W w.r.t frame \mathcal{T} - $\mathbf{x} \in \mathbb{R}^{3 \times 1}$ (See Fig. 1).
- Velocity of W w.r.t trocar frame \mathcal{T} - $\mathbf{v} \in \mathbb{R}^{3 \times 1}$
- Differentiation of joint torques - $d\boldsymbol{\tau} \in \mathbb{R}^{3 \times 1}$

	A	B	C	D
Mass [kg]	0.5	0.5	0.65	0.65
Cable (%)	100	50	50	0

TABLE I: Weight configurations. Values of model mass m in (1) in first row, percentage of hanging optic cable in second row.

Configuration	Precision	Recall	Delay [ms]
Sp 1	98.1	93.6	81.3
Sp 2	82.3	86.1	134.6

TABLE II: Results for both spatial configurations. The last column shows mean values.

During surgery, both the robot's base and the patient can move. For this reason, input features were calculated w.r.t the trocar frame \mathcal{T} to be robust to spatial changes.

It was found that the differentiated joint torques presented a different behavior between human manipulations and poor compensations, as illustrated in Fig. 2.

As stated before, a Neural Network is used to detect disturbances. Since data are temporal, Recurrent Neural Network (RNN), more specifically Long Short-Term Memory Network (LSTM), were chosen [7]. Indeed RNNs seem to be the most suitable choice because they are capable of remembering previous decisions and are influenced by what happened in the past. This is possible since information is passed from timestep to timestep, thanks to recurrent connections [8], [9].

C. Results

Input data were recorded at 20Hz, in order to detect significant changes between human manipulation and poor compensations. To obtain a fast behavior, the network is composed of a single layer of LSTM cells of size 18, followed by an output layer. To train it, data were cut into 25-second windows. Using the Adam algorithm [10], the LSTM network is trained during 50 epochs with a learning rate of 0.001. During the testing phase, probabilities are computed for each sample, and states are propagated in order to keep the temporal aspect of the network. This method can therefore be used in an online fashion.

Three metrics were chosen to quantify results: 1) Precision, 2) Recall, 3) Delay of detection in milliseconds (See "Delay" in Fig. 2). Both precision and recall are calculated for each sample.

As shown in Table II, the proposed method is able to correctly detect poor compensations, in experimental condition. Furthermore, detection delay is relatively short, compared to the average total duration of a fall or rise (~ 600 ms). Results for Sp2 are significantly lower than for Sp1, but that was expected, since training data were exclusively from Sp1.

To put results in perspective, it is important to compare them with existing methods. Previous work aiming at detecting such disturbances was presented in [11] with mean detection delays of up to 1.5s. Compared to this work, the detection delay has been reduced by more than 90%, which is a considerable improvement. The method

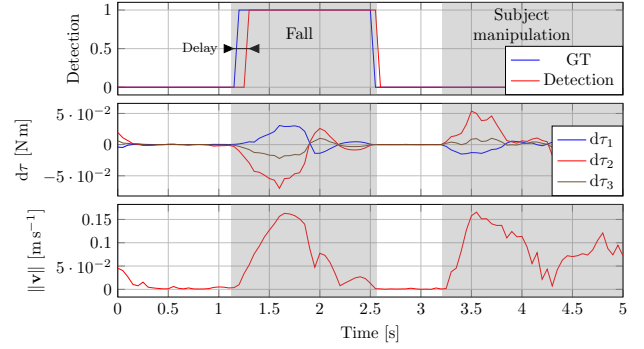


Fig. 2: The top plot shows the output of the model as well as the ground truth. The two bottom graphs show the chosen features.

from [11] was only tested in Sp1, so we cannot compare results when the spatial configuration is changed.

CONCLUSIONS AND DISCUSSION

In this paper, a new approach to detect disturbances in the gravity model is implemented. Using position, speed, and joint torques differentiation as input, a Neural Network can detect camera motions due to poor weight compensations. When compared with another approach presented in [11], this method showed better performances, reducing the mean detection delay by $\sim 90\%$. Obtained results are encouraging, and we propose to use this model that differentiates human movements from poor weight compensations to achieve an adaptive behavior. For example, this model can be used to trigger a correction of the gravity model when a poor compensation happens, e.g., updating the mass m in (1).

REFERENCES

- [1] J. Mago *et al.*, "Safe teleoperation of a laparoscope holder with dynamic precision but low stiffness," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2693–2699.
- [2] L. Ciupitu, "Adaptive balancing of robots and mechatronic systems," *Robotics*, vol. 7, no. 4, p. 68, 2018.
- [3] S. Beyhan *et al.*, "Adaptive fuzzy and sliding-mode control of a robot manipulator with varying payload," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, 2011, pp. 8291–8296.
- [4] R. Sharma *et al.*, "Design of two-layered fractional order fuzzy logic controllers applied to robotic manipulator with variable payload," *Applied soft computing*, vol. 47, pp. 565–576, 2016.
- [5] M. Aricò *et al.*, "A surgical cockpit for minimally invasive surgery," 2021.
- [6] L. Dong, "Assistance to laparoscopic surgery through comanipulation," Ph.D. dissertation, 2017.
- [7] S. Hochreiter *et al.*, "Long short-term memory," *Neural Computation*, 1997.
- [8] B. Singh *et al.*, "A multi-stream bi-directional recurrent neural network for fine-grained action detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1961–1970, 2016.
- [9] G. Zhang *et al.*, "Classification of hand movements from EEG using a deep attention-based LSTM network," *IEEE Sensors Journal*, vol. 20, pp. 3113–3122, 2020.
- [10] D. Kingma *et al.*, "Adam: A method for stochastic optimization," *CoRR*, 2015.
- [11] J. Mago *et al.*, "Fall detection for robotic endoscope holders in minimally invasive surgery," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4032–4039, 2021.

Towards an autonomous robotic platform for percutaneous procedures

Valentina Corbetta, Alice Segato, Andrea Fortuna, Tommaso Magni, Mattia Magro, Xiu Zhang
and Elena De Momi

Department of Electronics, Information and Bioengineering, Politecnico di Milano

INTRODUCTION

Structural Intervention Cardiology (SIC) procedures still present limitations (e.g. use of fluoroscopy with consequent radiation exposure, lack of force or haptic feedback, risk of embolisation or perforation) [1]. The ARTERY proposes an autonomous robotic platform to overcome these drawbacks.

MATERIALS AND METHODS

A. Real environment: actuating the ARTERY catheter

The MitraClip (MC) system is composed of two catheters, the sheath catheter and the delivery catheter; the latter is inserted in the first. Studying how the components of the MC system are handled by the surgeon during procedures [2] led to choose a motorize stabilizer for actuation, as seen in Figure 1B. The sheath catheter has two DOF, including bending and translational insertion (Figure 1B.1) and the delivery catheter has three DOF, which are two perpendicular in plane bending and one translational movement (Figure 1B.2).

The two main elements for actuation are the following:

- Stepper motors (1.a, 2.b, 2.c) to rotate the pulleys attached by the tendons, which guarantee a maximum holding torque of 0.6 Nm;
- Linear actuators (1.b, 2.a) that carry the catheters and perform a translation movement.

These components are designed in Solidworks and printed in Polylactic acid (PLA) material via the S3 3D printer (Ultimaker). This design was properly connected to the MC system to obtain the first prototype of the experimental setup. The first attempt to control the MC is performed using an Arduino One, exploiting a joystick (Figure 1B.7): different buttons actuate different motors.

B. Virtual environment

Starting from a Computed Tomography (CT) scan (dimension $512 \times 512 \times 347$) provided by IRCCS Ospedale San Raffaele, we reconstructed in Unity the anatomical environment in which the agent (i. e. the catheter) moves. In 3D Slicer, the original scan has been manually segmented to obtain the anatomical structures of interest (i. e. right and left atria and ventricles, inferior and superior vena cava, femoral veins, pulmonary artery) and subsequently smoothed, filtered with a gaussian filter. The final result is shown in Figure 1A. The purpose of the simulation environment is to allow pre- and intra-operative path planning. To this end, we have created two different

environments: pre-operative, which is static, and intra-operative, which is dynamic, as it simulates the pulsing of the heart and the vessels. We implemented a manual path planner controlled using a joystick. The surgeon can thus design and visualize all the possible trajectories in real time in the specif patient's anatomy and select the best one.

C. Bridging the real and virtual environments: Cosserat rod theory

In order to have an accurate simulation of the physical catheter in the virtual environment, it is paramount that the virtual catheter behaves as the real one in terms of geometry and kinematics. For this reason we have applied Cosserat rod theory to represent the MC geometry [3]. Cosserat rod theory employs a robot-attached reference frame composed by a matrix \mathbf{R} (orientation) and a vector \mathbf{p} (position). The robot configuration is evaluated in a fixed number of points, named nodes, placed along the rod. Solving the system for the nodes allows to obtain a reliable representation of the body. The frame is attached to the nodes and its evolution along the body length s is described by means of a system of differential equations:

$$\begin{aligned}\dot{\mathbf{R}}(s) &= \mathbf{R}(s)\hat{\mathbf{u}}(s) \\ \dot{\mathbf{p}}(s) &= \mathbf{R}(s)\mathbf{v}(s)\end{aligned}\quad (1)$$

We solve the equilibrium equations between internal forces and moments, $\mathbf{n}(s)$ and $\mathbf{m}(s)$, and external forces and moments, $\mathbf{f}(s)$ and $\mathbf{l}(s)$, for each node to obtain $\mathbf{u}(s)$ and $\mathbf{v}(s)$, which are the values of angular and linear rate of change of each node.

$$\dot{\mathbf{n}}(s) + \mathbf{f}(s) = 0 \quad (2)$$

$$\dot{\mathbf{m}}(s) + \dot{\mathbf{p}}(s) \times \mathbf{n}(s) + \mathbf{l}(s) = 0$$

At last, the internal force and moment are related with \mathbf{v} and \mathbf{u} , exploiting constitutive material laws:

$$\begin{aligned}\mathbf{n}(s) &= R(s)K_{se}(s)(\mathbf{v}(s) - \mathbf{v}^*(s)) \\ \mathbf{m}(s) &= R(s)K_{bt}(s)(\mathbf{u}(s) - \mathbf{u}^*(s))\end{aligned}\quad (3)$$

Combining (1), (2), (3), we obtain the complete set of Cosserat rod model derivative equations to be solved. Evaluating the system in the nodes gives a mechanics-based representation of the robot shape.

RESULTS

D. Real environment

In Table I are reported the specifications of the main components chosen for the experimental setup.

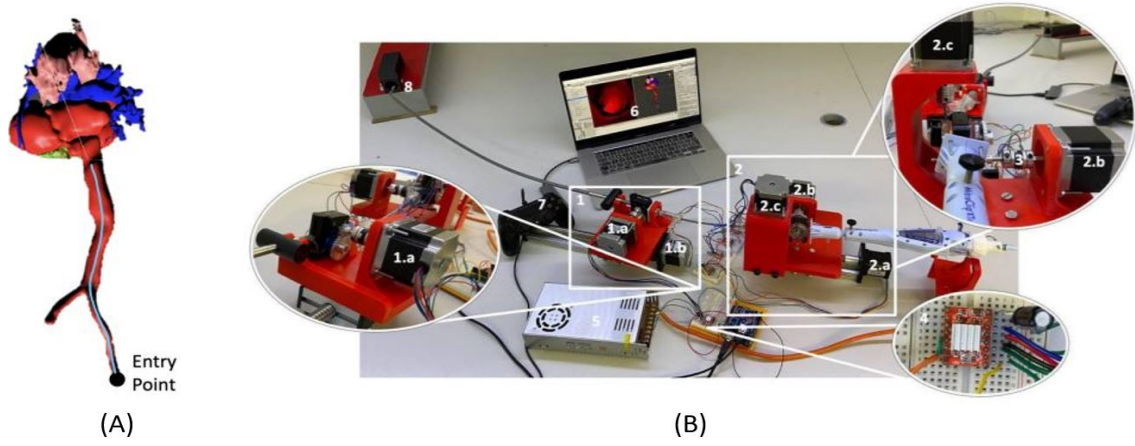


Fig. 1: (A) 3D model of cardiovascular environment and path planning results (light blue path); (B) ARTERY set up. Two main structures: (1) sheath catheter's actuation and (2) delivery catheter's actuation. Motor (1.a) allows the rotation of the external guide (i.e. sheath catheter); linear actuator (1.b) permits its insertion; linear actuator (2.a) works on the insertion of the delivery catheter and on the best positioning of the clip, using motors (2.b) and (2.c). The shafts of the motors and of the catheter are connected through oldham adapters (3). A4988 driver (4) controls the motor through the Arduino One and (5) is the 24V power supply. (6) shows the virtual environment and (8) emulates the femoral vein Entry Point.

TABLE I: Experimental setup's components

Components	Number	Weight [kg]	Torque [Nm]
Nema 17 Stepper Motor	3	0.350	1.26
Oldham adapter	3	0.05	1.7
Linear Actuators	2	1.36	2

E. Virtual environment

The trajectories designed in the simulation environment are evaluated in terms of minimum and average distance kept from the delicate structures (e.g. wall of vessels) and time to reach the target. Table II reports the averaged results on 10 manually executed trajectories.

TABLE II: Trajectories' parameters

Results	Min Distance [mm]	Mean Distance [mm]	Time [s]
Value	2.74	3.99	47.0

Figure 2 shows the output of the kinematic model when the catheter is subject to the gravity force along the z-axis and the displacements of the tendons. In particular, the first tendon is subjected to a 2.5 cm displacement, accordingly the third one is released by the same amount, we don't have any displacement for the tendon on the top. The resulting position of the end effector is equal to $p = [0.3, -0.33, 0.24]$.

CONCLUSIONS AND DISCUSSION

This work presents the first steps towards the development of an autonomous robotic platform for percutaneous procedures, exploiting a digital twin approach. Indeed, the actuation of the physical catheter shows promising results, as we are now able to control the catheter via a joystick. The kinematic behaviour of the real catheter is well

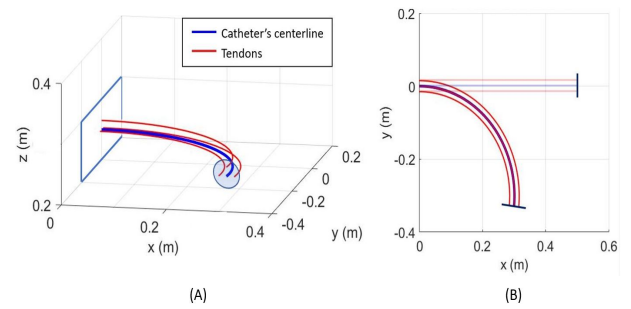


Fig. 2: (A) Catheter's kinematic model with gravity force and tendons displacements; (B) The top view shows the deformation of the catheter with respect to the original position.

predicted by the kinematic model, that will be the basis of our final control strategy. Finally, our virtual environment allows us to find the best possible path in simulation. We will use the manual virtual planner to collect a series of demonstrations performed by surgeons that will be used to train a Deep Reinforcement Learning Model based on the principal of learning from demonstrations. Integrating all these components will lead towards an autonomous platform.

REFERENCES

- [1] P. Legeza, G. W. Britz, T. Loh, and A. Lumsden, "Current utilization and future directions of robotic-assisted endovascular surgery," *Expert Review of Medical Devices*, vol. 17, no. 9, pp. 919–927, 2020.
- [2] J. Hasan, H. Asma, and K. Saibal, "Mitralclip: a novel percutaneous approach to mitral valve repair," *Journal of Zhejiang University-SCIENCE B (Biomedicine Biotechnology)*, vol. 12,(8):633–637, 2011.
- [3] A. Mousa, S. Khoo, and M. Norton, "Robust control of tendon driven continuum robots," in *2018 15th International Workshop on Variable Structure Systems (VSS)*. IEEE, 2018, pp. 49–54.

Hybrid controllers for middle ear surgery: towards efficient cholesteatoma removal

Jae-Hun So¹, Stéphane Sobucki¹, Jérôme Szewczyk¹, Naresh Marturi², and Brahim Tamadazte¹

¹ISIR, UMR 7222 Sorbonne University, CNRS, ERL AGATHE, U1150 INSERM, Paris, France

²Extreme Robotics Lab, University of Birmingham, Birmingham B15 2TT, United Kingdom

INTRODUCTION

Modern-day robots are required to operate autonomously in unknown, unstructured and dynamic environments with sufficient accuracy and repeatability. They are also expected to share the workspace with humans or other robots in a safe and secure manner. To this extent, external sensors like force, position, vision etc. play a prominent role. For instance, force sensors provide local contact information while vision sensors such as cameras provide rich and global information of the environment. This brings up to the question, whether it is possible to effectively combine several modes of perception (e.g., force and vision) within the same control scheme to benefit from their complementary advantages? Of course, this does not concern the sequential controllers that are usually reported in the literature [1]. Combining several information of different nature and geometry in the same control scheme can benefit many robotic applications, including surgical robotics to which the work described in this paper belongs. This work is part of an overall effort to improve the current surgical procedure, which consists of removing a pathological tissue that develops in the middle ear cavity, called cholesteatoma [2], [3]. This will involve the development of a minimally invasive robotic solution and adapted control modes to ensure exhaustive, accurate and safe cholesteatoma resection from the middle ear cavity.

In this paper, we investigate new control modes for middle ear surgery. The main contributions lie in the formulation of new hybrid controllers by consistently combining two physical quantities in the control loop. This allows separating the linear degrees of freedom (DoF), which are controlled in tele-operation (respectively, comanipulation) while the rotational DoFs are intuitively controlled by a vision-based controller.

MATERIALS AND METHODS

As depicted in Fig. 1, the developed robotic setup consists of a 7 DoF cobot from FRANKA EMIKA on which a 6 DoF Force-Torque (F/T) sensor (ATI MINI-40) is attached at the end-effector. Additionally, a 3D printed tool of 2mm diameter mimicking a typical surgical instrument used for surgery inside the middle ear cavity is fixed on the robot's end-effector. A AVT Guppy PRO F033b camera (with a frame rate of 25 Hz) is mounted in an *eye-in-hand* configuration on the robot end-effector. A head phantom at scale 1:1 is positioned in order to

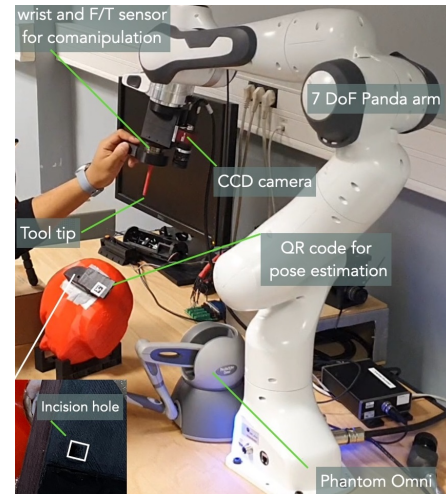


Fig. 1: Overview of the developed robotic setup.

simulate the position of a patient on the operating table during intervention. Finally, a tunnel (having the shape of the 3D tool) is drilled on the head representing the operator channel. The size of the drilled hole is slightly larger than that of the tool to have a tolerance of 0.5 mm during the tool insertion (Fig. 1).

Different control methods are developed and evaluated using the robotic setup. Firstly, a classical end-frame tele-operation mode using position-based controller with a Sensable Phantom Omni, as well as a comanipulation mode using the integrated F/T sensor are implemented on the robotic setup [4]. Secondly, to improve the peg-in-hole style insertion task of the surgical tool into the middle ear cavity and through the operating channel with respect to the ergonomics, intuitiveness, precision and safety of the procedure, particularly in minimally invasive surgery, we have developed a new generation of controllers. These controllers consist of two shared control laws, which combine proximal (local) and global measurements (Fig. 2). More specifically, local measurements provided by position or force sensors are combined with globally observed visual information, which is acquired from the camera attached to the robot end-effector in an *eye-in-hand* setting.

The first shared controller we developed in this work is called hybrid parallel (Fig. 2(a)). It consists of the parallel juxtaposition of two internal loops [5]. The first is a vision feedback loop to automatically manage the angular motion of the robot. The second is a position-based (in case of tele-operation mode) or force-based (in

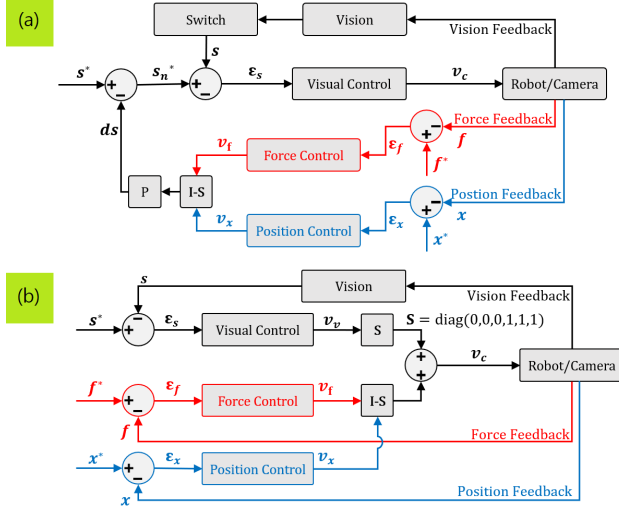


Fig. 2: Diagram of proposed shared control loops: (a) Parallel Hybrid force/vision comanipulation (in red) and position/vision teleoperation (in blue) control scheme (b) External Hybrid force/vision comanipulation (in red) and position/vision teleoperation (in blue)

case of comanipulation) loop to control the linear motion.

When both the force/position and vision-based controllers work in parallel, there is a risk that the tracked visual features (i.e., the *AprilTag*) are lost (i.e., goes out of the camera field of view (FoV)). This could jeopardize the accuracy of final positioning task. To tackle this issue, we have designed a new hybrid controller called external hybrid, as depicted in Fig. 2(b). The underlying idea is to express the control task as two hierarchical sub-tasks. The first task (priority sub-task) deals with maintaining the *AprilTag* at the center of the camera FoV, while the second one (secondary sub-task) is devoted to the regulation of the error between the current and the desired poses.

RESULTS

The proposed controllers were validated and compared with each other by analysing the scenario where a 3D printed tool is inserted into an incision hole. Initially, the robot is placed in an arbitrary position, then the operator has to jog it in front of the incision and to insert the tool. A group of 5 participants (2 experts and 3 novices) carried out the positioning and insertion tasks using different teleoperation (classic, parallel hybrid and external hybrid) and comanipulation (classic, parallel hybrid and external hybrid) control modes. For each of the performed tasks, Cartesian errors (along each DoF) as well as the time required to achieve the tasks are recorded and analysed.

Fig. 3 summarizes the translation and rotation errors. For comanipulation modes, it can be noticed that the classical mode is slightly more accurate than the external hybrid one. The parallel hybrid is the most accurate with the mean linear error $e_t = 0.84$ mm (2.15 mm and 2.17 mm for the classical and external hybrid methods, respectively). The mean angular error clearly shows that the parallel hybrid comanipulation gives the best result with $e_r = 0.45^\circ$ (6.67° and 3.07° for the classical and external hybrid methods, respectively).

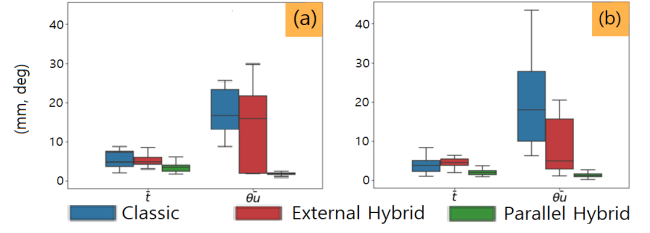


Fig. 3: Mean translation errors and rotation errors for the evaluated control laws. (a) tele-operation modes (b) comanipulation modes

For tele-operation modes, the result is similar to comanipulation modes, parallel hybrid is the most accurate. The average linear error is $e_t = 0.67$ mm (1.32 mm and 1.49 mm for the classical and external hybrid methods, respectively). Finally, one can point out that both shared control methods outperform the classical control laws.

Considering the required time to complete the defined insertion task, it appears that the hybrid parallel teleoperation approach needs in average 40.53 ± 10.09 seconds, which is 50% faster than the classical and external hybrid ones, when the hybrid parallel controller requires on average 29.03 ± 6.94 seconds to achieve the task, which is approximately 25% faster than the others control schemes.

CONCLUSIONS AND DISCUSSION

The objective of this study was to provide surgeon, ergonomic, intuitive and accurate control method for the tool positioning at the early stage of a cholesteatoma surgery. To do so, a new generation of hybrid position/vision and force/vision were proposed and evaluated using a robotic setup under a scenario of inserting a surgical tool through a millimetric incision hole performed behind the ear lobe. This study demonstrated that the shared control laws outperform the classical tele-operation and comanipulation modes, both in accuracy and time required to perform the predefined task.

Future work will focus on the implementation of the proposed controllers in clinical situation. Senior and junior surgeons will be recruited to evaluate the benefit of such approaches in the operating room.

ACKNOWLEDGMENT

This work has been supported by French ANR μ RoCS Project no ANR-17-CE19-0005-04.

REFERENCES

- [1] J. Baeten, H. Bruyninckx, and J. De Schutter, "Integrated vision/force robotic servoing in the task frame formalism," *The Int. J. of Rob. Res.*, vol. 22, pp. 941–954, 2003.
- [2] B. Dahrour, B. Tamadazte, and e. a. Tavernier, "Review on otological robotic systems: Toward micro-robot assisted cholesteatoma surgery," *IEEE R. in Biom. Eng.*, vol. 11, pp. 125 – 142, 2018.
- [3] J. J. Holt, "Ear canal cholesteatoma," *The Laryngoscope*, vol. 102, pp. 608–613, 1992.
- [4] J. So, B. Tamadazte, and J. Szewczyk, "Micro/macro-scale robotic approach for middle ear surgery," *IEEE T. on Med. Rob. and Bion.*, vol. 2, pp. 533–536, 2020.
- [5] J.-H. So, S. Sobucki, J. Szewczyk, N. Marturi, and B. Tamadazte, "Shared control schemes for middle ear surgery," *Frontiers on Robotics and AI*, pp. 1–12, 2022.

A posterior landmark-based registration-free method to identify pedicle screw trajectories for robot-based navigation: A proof-of-concept

D. Suter^{1,2}, A. Massalimova¹, C. Laux², F. Carrillo¹, M. Farshad², P. Frnstahl¹

¹Research in Orthopedic Computer Science, University Hospital Balgrist, University of Zurich,

²Department of Orthopedic Surgery, University Hospital Balgrist, University of Zurich

daniel.suter@balgrist.ch

INTRODUCTION

Pedicle screw placement (PSP) is a frequent intervention in spine surgery. The use of robotic systems for PSP gains increasing attention, as they allow more accurate screw implantation. In current approaches, screw trajectories are predefined in preoperative computed tomography (CT) data and adapted to intraoperative anatomy via registration. However, this approach is increasingly being questioned because the patient is exposed to radiation, the registration remains error-prone, and the screw trajectories cannot be adjusted intraoperatively. The EU Horizon 2020 project FAROS aims to develop a radiation- and registration-free robotic system for spine surgery. FAROS foresees using robotic ultrasound (US) to reconstruct a 3D model of the spine based on the preliminary work of Ruixuan Li et al. [1] (see figure 1). However, the existing approaches for navigation rely on CT scans where the full vertebra is visible to define screw trajectories. As US only allows to reconstruct the posterior surface of the spine (Fig.1) and not entire vertebrae a new method is required. In this work, we present a new approach capable of defining screw trajectories based on two points (p1 and p2) which can be defined on posterior surface models alone. P1 thereby is equivalent to the conventional entry point and p2 to the point of the screw tip.

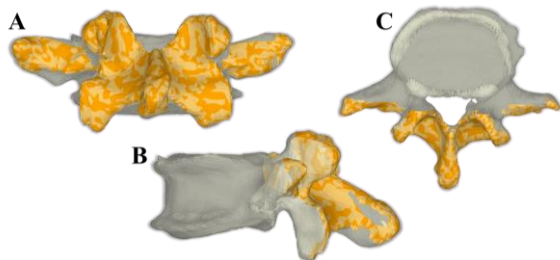


Figure 1. 3D model of the vertebrae (gray) and posterior surface of the spine visible in 3D US (orange) in posterior (A); sagittal (B); axial (C) views.

MATERIALS AND METHODS

In freehand screw placement the EP in the lumbar spine is located at the junction between the pars interarticularis and the transverse process immediately lateral to the mammillary process, or at the bisection of a vertical line through the facet joints and a horizontal line through the transverse process [2]. An optimal screw trajectory should be parallel to the upper-end plate in the sagittal plane and should not cross the midline in the axial plane. The screw is inserted perpendicular to the tangent of the

spinous processes. There are rough reference values for the convergence of the screw (axial plane). The convergence of the screw increases from L1 to L5 and ranges from 5° (L1) to 45° (L5). Our method can be divided into the following steps: definition and adjustment of the sagittal, coronal, and axial plane (step 1), definition of p1 (step 2), and definition of the p2 and thus determination of the trajectory (step 3).

Step 1. Definition and fitting of anatomical planes

Three orthogonal planes corresponding to the sagittal, coronal, and axial planes of the human body are created. Those are hereafter aligned to anatomical landmarks of the 3D model as follows. First, the sagittal plane is aligned in the direction of the spinous process (Fig. 2 A, B). Alignment of the axial plane is first performed in the posterior and then in the lateral view as follows. In the craniocaudal direction, the axial plane is aligned to the junction of the middle to caudal third of the transverse process (Fig. 2 C).

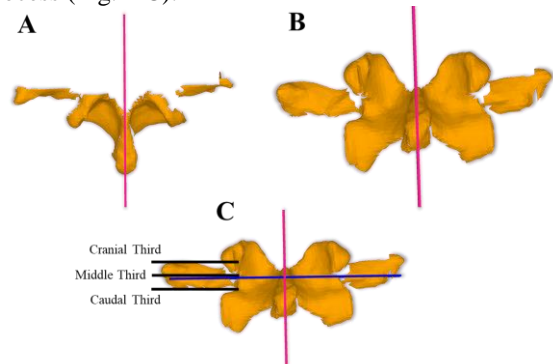


Figure 2. A and B: Alignment of sagittal plane (pink) to spinous process on posterior mesh (orange) in axial (A) and posterior (B) views. C: Orientation of axial plane (blue) based on transverse process. Black lines indicate the border of cranial, middle and caudal third of transverse process.

Subsequently, the sagittal angulation of the axial plane is defined on the lateral view. To this end, the coronal plane is first translated to be tangent to the transverse processes just laterally of the mammillary processes. Afterwards, a copy of the coronal plane (hereinafter defined as tangent-plane, yellow in Fig. 3A) is parallelly moved posteriorly until it touches the most posterior aspect of the spinous process. The axial plane is then rotated until the tangent plan is tangent to the lordosis at the most posterior aspect of the spinous process.

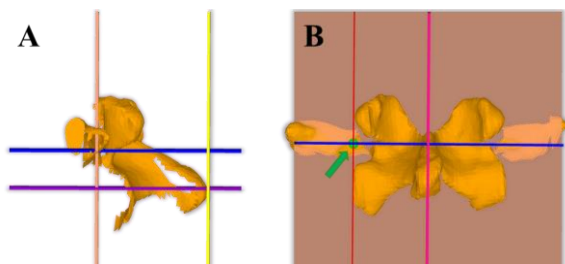


Figure 3. A: (sagittal view) Orientation of axial plane (blue) in lateral view based on the tangent plane (yellow). Purple plane : parallel to axial plane (blue) through main axis of spinous process. Light brown plane: coronal plane. **B:** (posterior view) definition of p1 (green sphere and arrow). Red plane= parasagittal plane.

Step 2. Definition of p1

P1 is defined as the intersection of the coronal plane, the axial plane, and a right/left parasagittal plane (red in fig. 3B) lateral to the mammillary processes if prominent or at the transition between the transverse process to the pars interarticularis.

Step 3. Definition of p2

To identify p2, additional planes are defined. Due to anatomical differences between levels, the definition varies for levels L1-L2 and L3-L5. First, a parasagittal plane is chosen tangential to the most lateral aspect of the facet joint (facet-plane, red in Fig. 4). For L1-L2, a parasagittal plane midway between the sagittal and facet planes (screw-tip plane, green in Fig. 4) is selected. For L3-L5, the screw-tip plane lies medially at one-quarter of the distance between the facet plane and the sagittal plane. Finally, a additional plane parallel to the tangent plane is created with the distance tangent plane to at the distance from the tip of the spinous process to the most anterior fifth of the vertebral body according to the axial CT image (anterior plane, black in Fig. 4). P2 is defined as the intersection of the anterior plane, the screw-tip plane and the axial plane. Lastly, the screw trajectory is defined as the direction vector from p1 to p2.

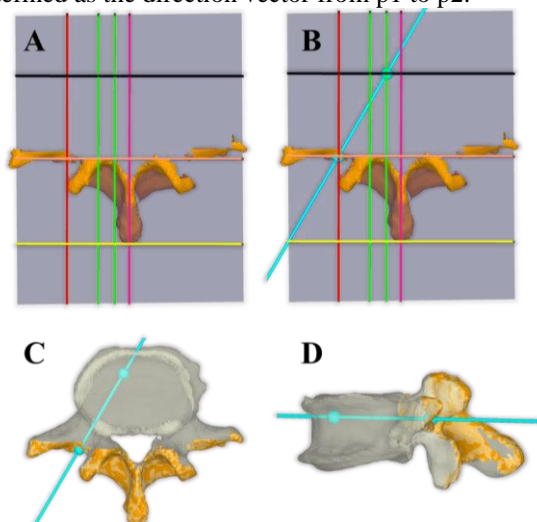


Figure 4. A: axial view with anterior plane (black), facet plane (red), screw tip planes (green, lateral for L1-2 and medial for L3-5), tangent plane (yellow), sagittal plane (pink), coronal plane (light brown) and axial plane (blue transparent).

B: axial view with defined p1, p2 and trajectory. **C, D:** defined trajectory on axial and sagittal view, respectively.

Evaluation

We used retrospective CT data from 9 human lumbar spine cadavers for evaluation of the proposed method. For each lumbar vertebra (L1 to L5), *ground-truth* screws were planned using a surgery planning software (CASPA, Balgrist, Switzerland). In a next step, the posterior surface was made in Materialise 3-Matic using the brush function to mimic the reconstructed 3D US data (Fig. 1). The proposed method was then manually applied on the posterior surfaces using CASPA. Subsequently, the trajectory position and direction error was evaluated by comparing the trajectory from the proposed method with the ground-truth trajectories.

RESULTS AND CONCLUSION

The average trajectory position error in 3D (eucliden distance) between the new method and the *ground truth* was 4.06 mm (± 2.62 mm) in our analysis. The average trajectory direction error (euler angle) was 4.74° ($\pm 3.48^\circ$). However, these data are not fully conclusive as to whether a clinically relevant perforation would result of our screw placement. Concerning the angular deviation, our values lie within the range of the reported literature. The position error in our analysis was determined for p2 in all instances, which corresponds to the tip of the screw. Up to date, there are no reported values for this measurement. However, the deviation at the tip point of 4.06 mm (± 2.62 mm) is just below the difference between the conventionally available screw lengths. Thus, the deviation would be correctable by a different screw length.

In this study, we proposed a novel method for defining the screw trajectory during spinal surgery only requiring a 3D-reconstruction of the posterior surface of the spine. The preliminary results proved feasibility on simulated data based on CT. The evaluation on real data obtained from robotic ultrasound will be addressed in future work.

Table .1. Accuracy of our methods compared to ground truth data

	Position error [mm]	Direction error [°]
overall	4.06 (± 2.62)	4.74 (± 3.48)
L1	6.09 (± 1.59)	7.44 (± 4.50)
L2	4.90 (± 1.68)	5.91 (± 3.74)
L3	1.76 (± 1.64)	2.58 (± 1.34)
L4	2.67 (± 2.09)	3.03 (± 0.92)
L5	4.88 (± 3.76)	4.74 (± 4.26)

ACKNOWLEDGEMENT

FAROS is a project of the EU's Horizon 2020 research and innovation program under grant agreement No 101016985.

REFERENCES

- [1] Li R, Niu K, Wu D, Vander Poorten E. A Framework of Real-time Freehand Ultrasound Reconstruction based on Deep Learning for Spine Surgery. In 10th Conference on New Technologies for Computer and Robot Assisted Surgery, Barcelona, Spain 2020.
- [2] Perna, F., Borghi, R., Pilla, F. et al. Pedicle screw insertion techniques: an update and review of the literature. *Musculoskelet Surg* 100, 165–169 (2016). <https://doi.org/10.1007/s12306-016-0438-8>

Intra-Domain Adaptation for Robust Visual Guidance in Intratympanic Injections*

Luke M. Shepherd¹, Sophia Bano¹, Joseph G. Manjaly², Lukas Lindenroth¹, and Danail Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London

²University College London Hospitals Biomedical Research Centre, National Institute for Health Research

INTRODUCTION

Intratympanic steroid injections are commonly used for the treatment of ear diseases. During this treatment, an expert Ear, Nose Throat (ENT) clinician deliver the drug by viewing through a large microscope that provides a close-up view of the anatomical landmarks on the middle ear. A steady hand and swift response to any patient movement is required to avoid improper placement of the needle. To assist the clinician during this treatment, a fluidic soft robot is proposed in [1] that can steer inside a lumen for providing steady guidance for the drug delivery. For robust visual guidance, stable anatomical landmarks (tympanic membrane, malleus, umbo) segmentation is required.

The current method [1] uses a clinical data pixel-wise annotated for the segmentation model training, which does not generalise to the phantom ear data currently use for the in-lab validation of the soft robot. The clinical data is taken from a high-resolution optical microscope from a fixed camera perspective with diffuse, even lighting. While the phantom images are recorded on a miniature digital camera endoscopically, from multiple perspectives and often with uneven, highly local illumination. The phantom is 3d-printed from patient scans [2] and coated with a pigmented silicone rubber (Dragon Skin™ 30, Smooth-On Inc., Easton, PA, US) to create a skin-like surface texture. Transparent silicone rubber is employed to create a membrane to resemble the tympanic membrane. Whilst being close to real patient anatomy, the phantom exhibits visual differences in terms of the coloring of the tissues and tympanic membrane, the translucence of the tympanic membrane as well as the overall visibility of the middle ear structure. This means that labelled clinical data and the previously labelled phantom data are not representative of the images that will be passed to the model during deployment (Fig. 1).

Due to the difference in training and deployment data, model predictions can be both inaccurate (incorrectly identifying the structures of the tympanic membrane) and unstable (predictions are discontinuous and noisy). Both are serious challenges for using such models as part of a robot control system – especially, the lack of prediction

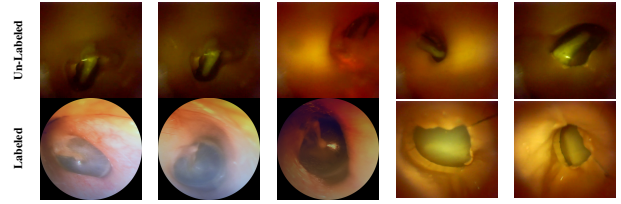


Fig. 1: Target domain (top) source domain (bottom)

stability. In this work, we perform intra-domain adaptation to learn a generalised model that provides stable and consistent segmentation on unseen phantom data.

MATERIALS AND METHODS

We propose using three segmentation models with DeeplabV3+ [3] architecture and three different backbones. We use transfer learning, initialising the models with ImageNet weights, and high amounts of data augmentation while using training techniques that encourage reduced convergence times for domain adaptation. These models are ensembled, with the output taking predictions of the composed logits for each class; taking advantage of each model whilst filtering out uncorroborated predictions. This results in both the stable and generalised predictions on unseen phantom data. Our approach takes advantage of smaller models' ability to learn smaller class features in fewer epochs and larger encoder's ability to output more stable predictions over a wider array of novel inputs.

Transfer Learning and Learning rate schemes: To compensate for the limited amount of labelled data and the need for the model to generalise to out of domain images, we used encoders pre-trained on Imagenet. To preserve useful, generalisable pattern and shape identifying convolutions shallower in the network, we utilise *discriminative learning rates* throughout the training - training shallower layers at a lower learning rate and deeper layers at a higher learning rate to increase the performance of the task specific segmentation. We also utilise *scheduled learning rates* – specifically the fit one cycle [4] which has been shown to both improve model performance and reducing convergence time. This approach increases the learning rate of the model for the first 30% of the total epochs; the learning rate is then annealed, returning to its original base value at the end of the training. Additionally, momentum is also adjusted to help with regularisation and ameliorate the

*This research was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z]; the Engineering and Physical Sciences Research Council (EPSRC) [EP/P027938/1, EP/R004080/1, EP/P012841/1]; and the Royal Academy of Engineering Chair in Emerging Technologies Scheme.

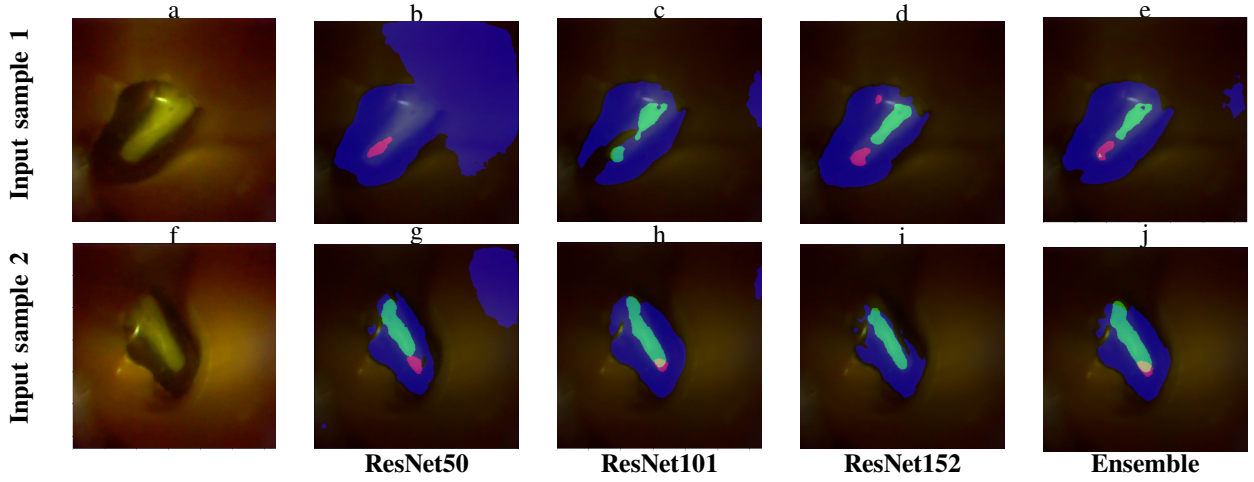


Fig. 2: Visual results showing unseen phantom input images and the model output when using ResNet50, ResNet101, ResNet152 as backbones for DeepLabV3+ [3] and the ensemble of all these models.

TABLE I: Qualitative comparison of different models using mean intersection over union (IoU) and F1 score.

Architecture	Encoder	IoU	F1 score
DeepLabV3+	ResNet50	0.8323	0.9085
DeepLabV3+	ResNet101	0.8277	0.9057
DeepLabV3+	ResNet152	0.8322	0.9084
Ensemble (proposed)	ResNet50/101	0.8353	0.9089

effects of the higher learning rate. Following the inverse of the learning rate schedule, momentum is reduced to its minimum value at 30% of total epochs, returning to its original value at the end of training. Learning rates were set using the *learning rate finder approach* [4].

Starting with a very low initial learning rate many batches are drawn, the loss calculated and recorded then optimised. With each mini batch the learning rate is incrementally increased until the loss explodes. This process allows us to efficiently and quickly estimate optimal learning rates without having to conduct large, slow and computationally expensive hyperparameter sweeps.

Data Augmentation: Data augmentation is an important part in stabilising and generalising model predictions. To help the model generalise to the unseen Phantom data we augment the data across three categories: image position (dihedral flips, rotation) perspective (warping, zooming) and lighting (brightness, saturation). Additionally, we also employed random resize crop which has been shown to both assist with generalisation and model performance.

Ensemble of Models: The ensemble mask prediction was created by composing the three models logits for each class; The sum of the logits were taken, averaged and then passed through a threshold to create the ensemble's class prediction. For the malleus and umbo layers, the magnitude of the negative logits were reduced by 50% for the ResNet152 and ResNet101 encoders and 80% for the ResNet50 - This helps to reduce the effect of large negative logits from an unsure model cancelling out correct predictions. This is particularly necessary for small class features

that can be easily obscured by false negative logits; conversely, the Tympanic membrane Benefits from taking the sum of (un-reduced) logits helping to stabilise boundary edges, especially with directional or off axis lighting.

RESULTS, DISCUSSION, AND CONCLUSIONS

We observe from qualitative comparison that the ensemble model's prediction are both more stable and more robust even on off axis out of domain sequential images (videos). However, the improvement reported in Table I is marginal which is mainly because limited labelled data was available for the quantitative analysis that didnot capture variabilities that are observed in sequential data. Ensembling model of different backbone sizes creates outputs that have significantly more stable anatomical landmarks - specifically with the umbo which often is either not predicted (Fig. 2 (b), (c), (g)) or predicted in multiple locations (Fig. 2(d)). This can cause problems If the landmark is used As part of a positioning or typing system. Ensembling multiple different models has several advantages over alternative methods such as test time augmentation which while can effectively stabilise predictions severe off axis or novel lighting can insufficiently cause smaller, more marginal class features to be lost. However, increase in predictive power comes at the cost of greater prediction latency memory requirements.

REFERENCES

- [1] L. Lindenroth, S. Bano, A. Stilli, J. G. Manjaly, and D. Stoyanov, "A fluidic soft robot for needle guidance and motion compensation in intratympanic steroid injections," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 871–878, 2021.
- [2] D. Sieber, P. Erfurt, S. John, G. R. Dos Santos, D. Schurzig, M. S. Sørensen, and T. Lenarz, "Data descriptor: The openEar library of 3D models of the human temporal bone based on computed tomography micro-slicing," *Scientific Data*, vol. 6, pp. 1–9, 2019.
- [3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [4] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006, 2019, p. 1100612.

A Fabric Soft Robotic Exoskeleton with Novel Elastic Band Integrated Actuators for Hand Rehabilitation

Cem Suulker, Sophie Skach, and Kaspar Althoefer

Centre for Advanced Robotics
School of Engineering and Materials Science
Queen Mary University of London, United Kingdom

INTRODUCTION

Common disabilities like stroke and spinal cord injuries may cause loss of motor function in hands [1]. They can be treated with robot assisted rehabilitation techniques, like continuously opening and closing the hand with help of a robot, in a cheaper, and less time consuming manner than traditional methods [2]. Hand exoskeletons are developed to assist rehabilitation [3], but their bulky nature brings with it certain challenges. As soft robots use elastomeric and fabric elements rather than heavy links, and operate with pneumatic, hydraulic or tendon based rather than traditional rotary or linear motors, soft hand exoskeletons are deemed a better option in relation to rehabilitation [4].

A variety of soft actuator types can be used to achieve hand rehabilitation tasks. Silicone [5], fabric [6] and tendon [7] based actuators are the most common ones. In this work we present a soft hand exoskeleton introducing elastic band integrated fabric based actuators (Fig. 1). The glove helps the user perform hand flexion and extension motions, making it usable for rehabilitation and assistive purposes. Furthermore, thanks to expert textile manufacturing techniques and the exploiting of soft textile structures, it enhances wearability comfort and is closer than previous incarnations to a clothing accessory.



Fig. 1: Our novel fabric based soft hand exoskeleton robot, worn relaxed (left) and actuated (right).

MATERIALS AND METHODS

Our fabric exoskeleton glove is developed following iterative prototyping and experimenting with various textile materials. To optimise the design, we have tested a wide selection of fabric types and manufacturing techniques that show potential synergy for a soft exoskeleton finger. Due to required stretch capabilities, we focused on structures with a high elasticity as a key characteristic. Moreover, we chose materials that accommodate ubiquitous design

approaches to rehabilitation equipment and a better user-acceptance: soft and smooth fabrics similar to our own clothing.

Among the other key features of fabric actuators for soft hand exoskeletons are a large curvature and constant force across the surface of the actuator. This can be achieved by creating an imbalance between the top and bottom layer of the fabrics. The greater the imbalance, the better the flexion as well as force capability.

After preliminary sampling, we selected 3 actuators, shown in Fig. 2, that combined the most successful actuator characteristics: a woven stretch fabric (Fig. 2 c); a non-stretch woven cotton fabric with an added elastic band along the sides (Fig. 2 b); and the combination of the two, a woven stretch fabric with elastic band (Fig. 2 a). To actuate them, a latex bladder is inserted to guarantee airtightness.

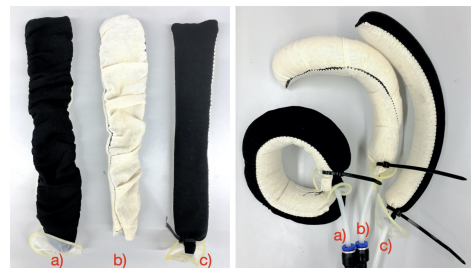


Fig. 2: Best performing designs non actuated (left) and actuated (right): a) woven stretch fabric, elastic band ; b) non-elastic woven, elastic band; c) woven stretch, no band.

RESULTS

To identify the best of the 3 actuators, we ran tests to measure their flexion angle and force capabilities. Flexion angle is a critical parameter for assistive gloves, it is imperative that each finger is unrestricted in relation to its maximum angle [1]. Inflatable actuators therefore need large bending capabilities. Equally important is its force capability. For achieving activities of daily living an acceptable force capability is around 10-15 N [1].

Figure 3(a) shows that bending angle capability of the 'stretch with band' actuator is superior to the others. Just 30 kPa pressure is enough for it to bend a full cycle of 360 degrees. Since it is geometrically impossible to bend further, the actuator maintains its position after this point.

'Stretch only' and 'cotton with band' actuators can bend approximately 190 and 210 degrees respectively under 70 kPa pressure. While these values are also acceptable from a rehabilitation perspective, the 'stretch with band' actuator is clearly a better option.

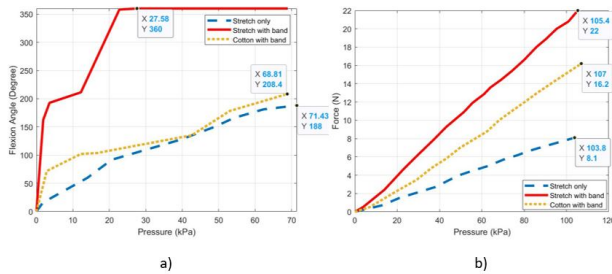


Fig. 3: Flexion angle versus pressure (a) and force versus pressure (b) plots for three selected actuators.

From Figure 3(b) it is seen that the 'stretch with band' actuator can apply 22 N force at about 100 kPa pressure. This is outperforming any other fabric or silicone based soft hand exoskeleton bending actuator. The 'stretch only' actuator can go up to 8 N force and the 'cotton with elastic band' actuator up to 18 N. It is also seen that our latex bladders can maintain more than 100 kPa pressure.

To unleash the maximum potential of the actuators, they need to be inflated to 100 kPa. However to prevent the latex bladder to fail (between 110 and 120 kPa) running them at 100 kPa for a long period might reduce the lifetime of the bladder. Operating them at 70 kPa might represent a sensible compromise, but that pressure would not be enough for the 'stretch only' actuator to carry out activities of daily living. The 'stretch with band' and 'cotton with elastic band', however, would be able to do so [1].

These tests confirm that, the actuator using stretch fabric with integrated elastic bands shows excellent performance both in relation to bending angle potential and force capability. The findings of the tests can now be used to lead the design of a new exoskeleton prototype.

The Prototype

The fabric based soft hand exoskeleton is manufactured using our 'stretch with band' actuator (Figure 1) which was found to be the best candidate in light of test results. The base glove, which the actuators are mounted on, is made from four-way stretch fabric: a black viscose jersey knit to guarantee soft feel on the skin.

The five inflatable fingers were sewn together with an overlock machine and hand-stitched onto the glove, reaching from the fingertip down to the centre of the back of the palm. For this prototype, each finger was fabricated from the same pattern and with the same dimensions - something that can be simply changed to accommodate different finger sizes. For the glove to be able to assist with the extension motion of the hand as well, another simple two layer fabric actuator is hand stitched between the base glove and the bending actuators.

In Figure 4 the exoskeleton glove can be seen supporting the hand while grasping objects.

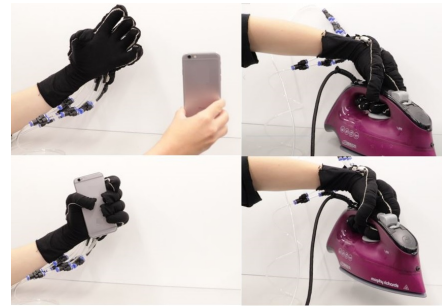


Fig. 4: The final hand exoskeleton glove "in action" grasping and releasing a phone and an iron.

CONCLUSION AND DISCUSSION

In this work, we present a novel actuation method for a fabric-based soft robotic exoskeleton glove. Integrating elastic band to the stretch fabric based actuator enhancing the capability of the exoskeleton, making it able to bend high angles in low pressure values, and boosting its force exertion capability to the fingers. Furthermore, our 'stretch with band' actuator can produce higher forces than any other available silicone or fabric based alternative. This improves grasping ability for the user that is a critical element in the rehabilitation process.

Assistive wearable devices should be designed with consideration given to the needs of the physically vulnerable. Being lightweight, a characteristic inherent in fabric-based devices, is therefore clearly a potential advantage textiles bear. We are accustomed to feel textile on our skin, which makes this material promising with regards to rehabilitation. Last but not least, this makes the robot also more acceptable to users.

REFERENCES

- [1] C. D. Takahashi, L. Der-Yeghiaian, V. Le, R. R. Motiwala, and S. C. Cramer, "Robot-based hand motor therapy after stroke," *Brain*, vol. 131, no. 2, pp. 425–437, 2008.
- [2] S. L. Wolf, C. J. Winstein, J. P. Miller, E. Taub, G. Uswatte, D. Morris, C. Giuliani, K. E. Light, D. Nichols-Larsen, f. t. EXCITE Investigators *et al.*, "Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the excite randomized clinical trial," *Jama*, vol. 296, no. 17, pp. 2095–2104, 2006.
- [3] H. Kawasaki, S. Ito, Y. Ishigure, Y. Nishimoto, T. Aoki, T. Mouri, H. Sakaeda, and M. Abe, "Development of a hand motion assist robot for rehabilitation therapy by patient self-motion control," in *2007 IEEE 10th International Conference on Rehabilitation Robotics*. IEEE, 2007, pp. 234–240.
- [4] B. Radder, G. B. Prange-Lasonder, A. I. Kottink, A. Melendez-Calderon, J. H. Buurke, and J. S. Rietman, "Feasibility of a wearable soft-robotic glove to support impaired hand function in stroke patients," *Journal of rehabilitation medicine*, vol. 50, no. 7, pp. 598–606, 2018.
- [5] P. Polygerinos, Z. Wang, K. C. Galloway, R. J. Wood, and C. J. Walsh, "Soft robotic glove for combined assistance and at-home rehabilitation," *Robotics and Autonomous Systems*, vol. 73, pp. 135–143, 2015.
- [6] H. K. Yap, P. M. Khin, T. H. Koh, Y. Sun, X. Liang, J. H. Lim, and C.-H. Yeow, "A fully fabric-based bidirectional soft robotic glove for assistance and rehabilitation of hand impaired patients," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1383–1390, 2017.
- [7] H. In, B. B. Kang, M. Sin, and K.-J. Cho, "Exo-glove: A wearable robot for the hand with a soft tendon routing system," *IEEE Robotics & Automation Magazine*, vol. 22, no. 1, pp. 97–105, 2015.

Electrical Impedance Tomography Sensing for Soft Continuum Surgical Manipulators

A. Alian, E. Zari, J. Avery, E. Franco, M. Runciman, F. Rodriguez y Baena, and G. Mylonas

The Hamlyn Centre for Robotic Surgery, Imperial College London

INTRODUCTION

Soft robotic manipulators are attracting increasing interest among researchers, thanks to their ability to achieve structural deflection upon fluidic actuation and adapt to unconstrained environments. In particular, in the field of minimally invasive surgery (MIS), where a soft interaction with the human body is required, such systems made of low-stiffness materials have enabled new possibilities.

Despite noticeable advantages that using soft matter might bring in designing interventional surgical devices, a number of impediments are constituted to the development of a fully autonomous system. Designing a stable and real-time control system along with developing a robust sensing scheme are open problems in soft robotics. The engineering challenges related to the control of soft robots usually stem from the fact that soft continuum manipulators tend to be underactuated mechanical systems. Another contributing factor to control challenges are the unstructured areas with which soft robots interact, especially those in surgical applications. However, accurate measurement of the deformation of soft robots in these environments facilitates the design of a reliable control system by providing correct feedback signal. Given the highly compliant characteristic of continuum robots, a sensor should be able to stretch enough to track and detect slight deflections in the soft actuator. Aside from being flexible enough, the sensing of a surgical tool should be biocompatible, robust to the existing electromagnetic field in the operation room and radiation-free [1].

Here, we propose a promising method for proprioceptive shape reconstruction of soft continuum manipulators based on electrical impedance tomography (EIT), which is a compelling candidate for measuring deformation in soft surgical instruments [2]. Additionally, the EIT performance for different number of electrodes is evaluated through numerical simulations.

MATERIALS AND METHODS

In this study we simulate the planar deformations of a soft manipulator made of silicone elastomer and reinforced with nylon fiber. Figure 1.c shows the soft manipulator in a cardiovascular phantom. Hydraulic actuation is preferred over pneumatic actuation as it shows several benefits, including lower friction and the possibility to exert higher forces at the tip of the manipulator, as well as facilitating sensor integration. Using EIT in the proposed soft continuum manipulator of Figure 1 renders

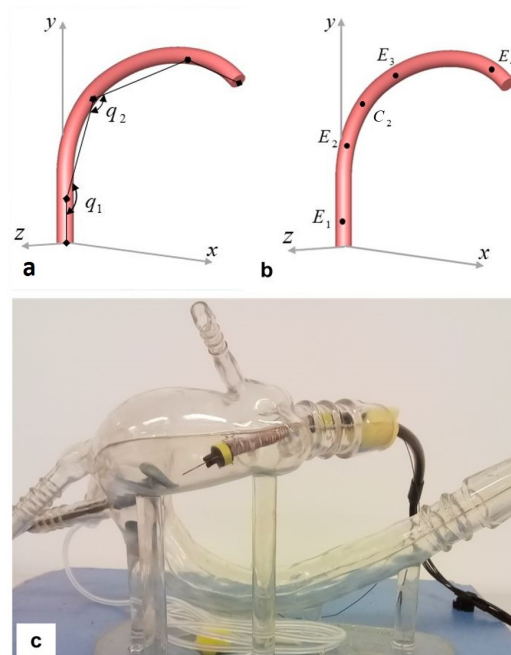
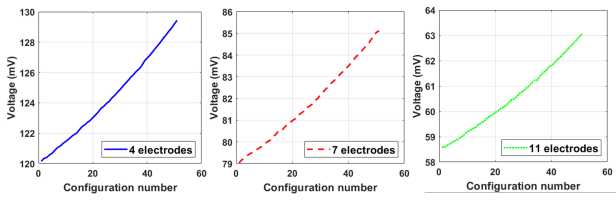
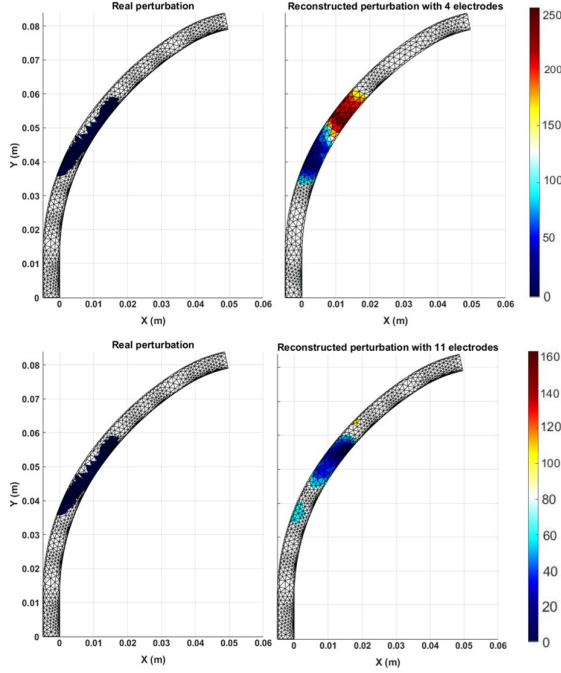


Fig. 1: a) Equivalent rigid kinematic model defined by variables q_1 and q_2 . b) Location of EIT electrodes on the proposed soft manipulator. c) Soft manipulator for MIS in a cardiovascular phantom. [3]

a proprioceptive actuator that exploits a single hydraulic input as both sensing and actuation unit of the system. The proposed soft actuator in this article is pressurized by means of 0.9 % saline solution. Saline is a biocompatible substance and using it in MIS applications minimizes risks to the patient in case of any damage to the soft body or leakage. Electrical current is passed through saline and, by measuring the resultant localized potential differences from particular sections along the robot's axis, the corresponding shape of the systems can be estimated in response to the deformations. This process of current injection and voltage recording is carried out via an array of electrodes that are inserted and spaced at specific intervals along the axis of the soft manipulator. After obtaining the voltages in a sequential manner and with various current injection patterns, a series of partial differential equations are employed for solving the forward problem in EIT [4]. Employing the geometry of the robot as boundary conditions, solving the forward problem using the finite element method (FEM) leads to the calculation of a sensitivity matrix. The inverse of the sensitivity



(a) Voltage measurements for different numbers of electrodes. Configuration number 1: $q_1 = 5$, $q_2 = 10$ degrees, Configuration number 2: $q_1 = 30$, $q_2 = 35$ degrees.



(b) Estimation of real perturbation value by 4 and 11 electrodes with corresponding conductivity changes.

Fig. 2: EIT performance for different number of electrodes

matrix is utilized for solving the inverse problem in EIT which yields the reconstruction of the soft manipulator's shape [2].

RESULTS

Current injection and voltage recordings are carried out by pairs of electrodes and through repeated measurement patterns. For example, as depicted in Figure 1.b, the current is injected in 1-4, and the voltage is measured from 1-2, 2-3, and 3-4, which represent the measurement protocol. In the next iteration, the injection pair shifts to 2-3, for instance, and the voltage is measured from the same three pairs. This protocol of voltage measurement is then repeated for the other injection pairs. One of the design parameters that can significantly contribute to the improvement of the accuracy in EIT sensing is the number of integrated electrodes.

To investigate the effect of the number of electrodes, several deformations are simulated using an equivalent rigid-link kinematic model [3] shown in Figure 1.a. Thus, the curvature of the soft manipulator is defined by joint angles q_1 and q_2 . In particular, 50 deformations were simulated with q_1 from 5 to 30 degrees, and q_2 from 10

to 35 degrees. To investigate the effect of the number of electrodes, the current was injected between the first and last electrodes at the distal ends of the soft manipulator. As voltage measurement pairs, the electrodes on the both sides of C_2 were chosen for each electrode count. For instance, in the case of 4 electrodes, electrode 2 and 3 were chosen (see Figure 1.b). As shown in Figure 2.a, by increasing the number of electrodes, the sensitivity of EIT decreases. This can be seen in the different voltage ranges for each electrode counts. This sensitivity reduction is due to the smaller volume of saline between electrodes, which is in direct correlation with the increase in the measured voltage. However, increasing the number of electrodes results in better spatial resolution. In particular, the deflection caused by the interaction of the manipulator with obstacles are interpreted as perturbations in EIT and higher spatial resolution results in a more accurate shape reconstruction of the soft manipulator. It is shown in Figure 2.b that EIT with 11 electrodes renders a better approximation for the value of the real perturbation than the one with 4. As a result, when choosing to include more electrodes in EIT, a balance between high sensing sensitivity and accurate shape reconstruction should be maintained.

CONCLUSIONS AND DISCUSSION

A promising strategy to design a sensing scheme for soft continuum manipulators hydraulically actuated by a conductive fluid was presented that may enable a new generation of soft manipulators with real-time proprioceptive sensing.

In this work, an optimization of the number of electrodes was performed. The results indicate that increasing the number of electrodes yields a decreased sensitivity but also a more accurate shape reconstruction. Ongoing work focuses on the manufacturing of soft manipulators with integrated EIT sensing in order to validate the simulations with experimental data.

ACKNOWLEDGMENTS

This work is partially supported by a research collaboration with the Multi-Scale Medical Robotics Centre, The Chinese University of Hong Kong.

REFERENCES

- [1] C. Shi, X. Luo, P. Qi, T. Li, S. Song, Z. Najdovski, T. Fukuda, and H. Ren, "Shape sensing techniques for continuum robots in minimally invasive surgery: A survey," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1665–1678, 2017.
- [2] J. Avery, M. Runciman, A. Darzi, and G. P. Mylonas, "Shape sensing of variable stiffness soft robots using electrical impedance tomography," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9066–9072.
- [3] E. Franco and A. Garriga-Casanovas, "Energy Shaping Control of Soft Continuum Manipulators with in-plane Disturbances," *The International Journal of Robotics Research*, vol. 40, no. 1, pp. 236–255, 3 2021.
- [4] K. Y. Aristovich, G. S. D. Santos, B. C. Packham, and D. S. Holder, "A method for reconstructing tomographic images of evoked neural activity with electrical impedance tomography using intracranial planar arrays," *Physiological Measurement*, vol. 35, no. 6, pp. 1095–1109, 2014.

A Linguistic Comparison Between Textual Datasets to Assess the Complexity of Surgical Robotic Procedural Descriptions

Marco Bombieri¹, Diego Dall’Alba¹, and Paolo Fiorini¹

¹Department of Computer Science, University of Verona, Italy
Contact: marco.bombieri_01@univr.it

INTRODUCTION

Natural language processing (NLP) has rapidly expanded over the last 20 years thanks to increasing computational resources and high investment from both many research groups and private companies, with the aim of automating various tasks previously performed manually [1]. A notable rise has occurred recently also in the medical domain, with the introduction of deep learning technologies and pre-trained language models, such as BERT [2]. Despite this rapid growth, only few studies have attempted to apply deep learning NLP methods to the surgical domain, which is a sub-category of the medical one. The majority of the studies in surgery deals with the extraction of outcomes from textual records and with other text classification applications [3].

Robotic devices, also known as Surgical Robotic Systems (SRSs), are increasingly pervasive in operating theaters. Together with the introduction of surgical robots, numerous books, manuals, academic papers and online-resources about robot-assisted surgery have been published. They are used by medical trainees to acquire or refine their knowledge.

Extracting structured workflows from surgical textual resources would be beneficial both for the development of autonomous SRSs and for the intra-operative assistance to the surgeon. However, the *manual* extraction of these workflows requires substantial human effort and expertise. NLP mining technologies for robot-assisted surgical texts will open new frontiers of research. In particular, they would enable:

- the extraction of high-level information from surgical text-books (e.g. phases, steps or actions). This would improve computer-assisted planning of surgical procedures.
- the development of cognitive SRSs, that recognize surgical entities (e.g. anatomical parts, tools, actions) in the texts and link them with existing a-priori knowledge represented in ontologies. They would allow to infer knowledge not explicitly stated in the texts improving situation awareness algorithms and robotic assistance;
- the use of a-priori knowledge by the SRS. This will be an essential step toward cognitive surgical robotics [4].

From surgical textbooks, we can extract two types of knowledge:

- *Declarative knowledge*, usually formalized into ontologies. In surgery, declarative knowledge describes for example the characteristics of anatomical structures (e.g., the tissue color or the organ size), the technical specifications of the robotic tools and exceptional events that can happen both intra- or post-operatively.
- *Procedural Knowledge*, possessed by an intelligent agent that is able to perform a task. In surgery, the intelligent agent is a surgeon or an autonomous SRS, and procedural knowledge describes how surgical interventions should be performed.

A limited number of studies (in non-medical domains) have addressed the problem of *Procedural Knowledge* extraction, from annotated instructional texts of recipes or maintenance/repair manuals [5]. The number of published works is even lower in the biomedical field, and only one preliminary work has been published in the robotic surgery domain [6]. This is due to the lack of annotated *publicly* available data in the procedural surgical literature that could be used to develop supervised deep-learning methods. Furthermore, annotating datasets for NLP is a very expensive task that requires expert personnel. A possible solution could be to use *procedural* datasets belonging to other domains. This work therefore wants to investigate the linguistic differences between robot-assisted surgery books and other domains, such as cooking recipes or repair manuals. This is an essential information for enabling the direct application of knowledge extraction algorithms designed and developed for other domains to surgical robotic.

I. MATERIALS AND METHODS

In this work, we propose the first linguistic comparison between the procedural language used in robotic-surgery textbooks and in other domains. We have compared three representative datasets, SPKS dataset¹ and the two procedural datasets used by [5]. The first consists of 20 descriptions of robot-assisted surgical procedures belonging to the urological, gynecological, gastrointestinal and thoracic domains [6]. [5] instead deals with 30 cooking recipes from the BBC recipe website² (referred as "BBC" later on) and 30 "How-Tos" from eHow.com website referred as ("E-HOW" later on).

¹<https://gitlab.com/altairLab/spks-dataset>

²<https://www.bbc.co.uk/food/recipes>

TABLE I: Values of the metrics considered in this work for the 3 analyzed datasets.

	SPKS	BBC	E-HOW
α	13.8	5.6	6.4
β	3.64	2.0	2.35
γ	1.74	1.75	1.79
δ	> 30%	<5%	<5%

The comparison is based on the following metrics:

- (α): Flesh-Kincaid’s readability score³:

$$FKGL = 0.39(ASL) + 11.8(ASW - 15.59)$$

where *ASL* (Average Sentence Length) is the number of words divided by the number of sentences, *ASW* (Average Number of Syllables per Word) is the number of syllables divided by the number of words. The higher the score, the harder the text.

- (β) average number of verbs per sentence identified with POS-Tagging and lemmatization techniques [7].
- (γ) number of average semantic roles per verb using state-of-the-art transformer-based semantic role detection algorithms [2].
- (δ) Percentage of sentences using domain verbs not found in [8]. The coverage is evaluated in terms of presence of the lemma and of an appropriate frame in the resource.

The first metric is designed to measure the difficulty of an English sentence for an human reader. The second metric instead investigates the syntactic and structural aspect, analyzing how many tokens with the role of verb (i.e. possible actions) are present in the sentence. It gives clues on the procedural complexity of a sentence. The third metric investigates how complex it could be for an information extraction algorithm to understand and disambiguate the sentence and link the lexical tokens semantically. The fourth metric measures how much state-of-the-art semantic propositional banks, built manually using statistical analysis of general-English texts, cover the lemmas and frames often used in procedural languages.

RESULTS AND DISCUSSIONS

Table I shows metrics described in the previous section for the 3 datasets considered.

The SPKS dataset has a much higher FKGL score than the other two. A value of 13.8 means *difficult to read: college graduate expertise*. The other two datasets have scores below 7, which correspond to *fairly easy to read: primary school expertise*. The implication is that a domain expert (doctor or linguist) is needed to annotate the surgical dataset for supervised algorithms. SPKS also has a higher number of verbs per sentence (1.83 times more than BBC and 1.54 times more than E-HOW). This indicates that the procedural sentences written in the surgical manuals are significantly more complex than those of the other domains. All 3 datasets have an almost

³This formula has obvious limitations because considers only structural features: for example, the pangram "Cwm fjord-bank glyphs vext quiz" is considered *very easy to read*. It is, however, representative, since it captures useful information.

equal number of semantic roles per verb. This is expected, since the languages are procedural and the number of elements involved (e.g action, agent, target or tool) is almost the same, regardless of the domain. Finally, the surgical domain makes use of a larger number of domain verbs (for instance *kocherize*, *extraperitonealize*, *grasp*) not covered by semantic textual resources. This suggests that current semantic banks developed for general-English are unable to fully represent the semantic complexity of procedural surgical language and should therefore be extended.

CONCLUSIONS

The linguistic comparison presented in this work confirms that the descriptions of robotic-assisted surgical procedures are much more demanding to interpret: (i) they use more verbs per sentence; (ii) they require expert personnel to annotate dataset; (iii) they make use of terms that are not covered by existing semantic banks.

This preliminary investigation suggests that, in order to obtain suitable performance in mining textual descriptions for surgical robotics, modifications to existing semantic banks and the development of specialized processing methods are required.

ACKNOWLEDGEMENTS

This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 742671 “ARS”).

REFERENCES

- [1] J. Wang, H. Deng, B. Liu, A. Hu, J. Liang, L. Fan, X. Zheng, T. Wang, and J. Lei, “Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: Bibliometric study on pubmed,” *J Med Internet Res*, 2020.
- [2] P. Shi and J. Lin, “Simple BERT models for relation extraction and semantic role labeling,” *CoRR*, vol. abs/1904.05255, 2019.
- [3] J. Mellia, M. Basta, Y. Toyoda, S. Othman, O. Elfanagely, M. Morris, L. Torre-Healy, L. Ungar, and J. Fischer, “Natural language processing in surgery: A systematic review and meta-analysis,” *Ann Surg*, 2021.
- [4] M.-C. Fiazza and P. Fiorini, “Design for interpretability: Meeting the certification challenge for surgical robots,” in *2021 IEEE Int. Conf. on ISR*, 2021.
- [5] Z. Zhang, P. Webster, V. S. Uren, A. Varga, and F. Ciravegna, “Automatically extracting procedural knowledge from instructional texts using natural language processing,” in *Proceedings of LREC 2012, Istanbul, Turkey*, 2012.
- [6] M. Bombieri, M. Rospocher, D. Dall’Alba, and P. Fiorini, “Automatic detection of procedural knowledge in robotic-assisted surgical texts,” *Int. J. CARS*, vol. 16, 04 2021.
- [7] D. Jurafsky and J. H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, 2nd Edition, ser. Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education International, 2009. [Online]. Available: <https://www.worldcat.org/oclc/315913020>
- [8] M. Palmer, P. R. Kingsbury, and D. Gildea, “The proposition bank: An annotated corpus of semantic roles,” *Comput. Linguistics*, 2005.
- [9] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 NAACL-HLT Conference, Minneapolis, MN, USA*, 2019.

Deep learning-based localization of the biliary tract on white-light images acquired during laparoscopic cholecystectomy

Cristina Iacono¹, Sara Moccia², Aldo Marzullo³,
Elena De Momi⁴, Fanny Ficuciello¹, and Umberto Bracale¹

¹Department of Electrical Engineering and Information Technology, Università degli Studi di Napoli Federico II, Napoli, Italy

²The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy

³Department of Mathematics and Computer Science, Università della Calabria, Rende, Italy

⁴Department of Electrical Engineering and Information Technology, Università degli Studi di Napoli Federico II, Napoli, Italy

⁵Department of General Vascular and Thoracic Surgery, Università degli Studi di Napoli Federico II, Napoli, Italy

INTRODUCTION

Cholecystectomy is one of the most frequently performed procedures in gastrointestinal surgery, and the laparoscopic approach is the gold standard for symptomatic cholelithiasis and acute cholecystitis. Since the introduction of Laparoscopic cholecystectomy (LC), surgeons have prioritized preventing complications.

Besides the advantages of a distinctly faster recovery and better cosmetic results, the laparoscopic approach bears a higher risk for bile duct injury with an incidence in the range 0.3–1.5%. Bile duct injury has a significant impact on quality of life and survival. Many measures have been implemented to reduce the risk of bile duct injury during laparoscopic cholecystectomy.

To avoid bile duct injury, intraoperative visualization of the bile duct using near-infrared light and the fluorescent dye indocyanine green (ICG) were introduced during cholecystectomy [1]. ICG is metabolized by the liver and excreted in bile, making it an excellent medium for biliary imaging. The problem in using ICG is that, while enhancing the bile duct, it makes it challenging to see all the other anatomical structures, see Fig. 1.

This work aims to address this problem, helping the surgeon to identify the biliary tract without the use of ICG. To this end, a deep-learning algorithm for the localization of the biliary tract from white-light images acquired during the surgical procedures has been implemented. This work also includes the construction and annotation of an image database to train the deep learning algorithm.

MATERIALS AND METHODS

The method directly uses laparoscopic images to localize the biliary duct. To this end, You Only Look Once (YOLO), a state-of-the-art convolutional neural network, has been used [2]. YOLO is a regression-based object detector that looks at the whole image once to perform the detection. It consists of a single CNN that simultaneously predicts bounding boxes and the class probabilities for the boxes. In addition, YOLO looks at the entire image to encode contextual information during prediction, and thus, it is extremely fast and found suitable to detect or localize objects in real-time.

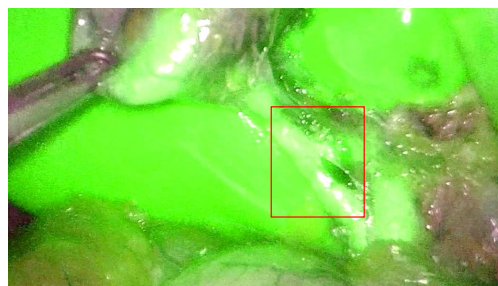


Fig. 1: Intraoperative visualization of the bile duct using indocyanine green (ICG).

With the support of expert clinicians, an image dataset has been collected from 12 video clips of 12 patients who underwent laparoscopic cholecystectomy. The videos collected from patients that presented complications that did not fall within the scope of this study were rejected. The videos were acquired through a high-definition endoscopic camera system with a 25 Hz frame rate during surgical endoscopic procedures. The frames extracted from the videos were sampled one every ten frames, obtaining 399 frames. The frames were then manually annotated by drawing a bounding box on the bile duct.

The video frames were split to have 208 frames for the training set and 191 frames for the test set, as illustrated in Table I. The training set was used to train the neural network, while the test set was used for evaluation purposes only. The frames of three patients out of twelve (Patients 3, 6 and 12) have been used only in the test set.

The Intersection over Union (IoU) was used as evaluation metrics: IoU compares the annotated bounding boxes with the bounding boxes predicted by the network, as follows:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

For the training phase the following parameters were used:

- Batch size: 64,
- Learning rate: 0.001,
- Epochs number: 1748,
- Average loss: 0.18.

TABLE I: Dataset composition

	<i>Total Frames</i>	<i>Training</i>	<i>Test</i>
Patient 1	142	15	15
Patient 2	171	34	14
Patient 3	219		39
Patient 4	152	74	20
Patient 5	48	5	5
Patient 6	144		29
Patient 7	168	14	10
Patient 8	89	18	10
Patient 9	153	14	10
Patient 10	73	20	10
Patient 11	27	14	10
Patient 12	135		19
Total		208	191

I. RESULTS

To confirm the performance of the proposed method, the network has been used on the test set. Table II shows the results of the biliary duct detection on each patient. Based on the experimental results, the overall IoU is 0.67.

TABLE II: Detection Results

	<i>IoU</i>	<i>STD</i>
Patient 1	0.65	0.04
Patient 2	0.70	0.07
Patient 3	0.53	0.02
Patient 4	0.63	0.11
Patient 5	0.73	0.02
Patient 6	0.65	0.04
Patient 7	0.80	0.01
Patient 8	0.63	0.08
Patient 9	0.76	0.02
Patient 10	0.84	0.02
Patient 11	0.65	0.02
Patient 12	0.58	0.04

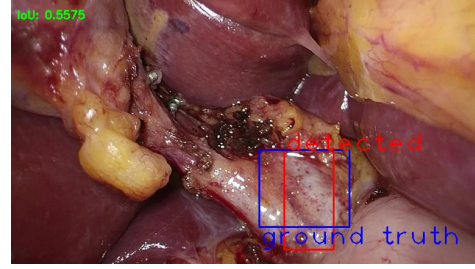
Regarding the videos, which have frames in the training set, the worst-case scenario happened for Patient 4 because in the video, the biliary tract is not easily distinguishable from the background. However, the localization of the surgical site of interest is accurate enough to assist the surgeon in identifying the biliary tract. Since each patient can be very different from the others, a larger dataset is needed to increase the robustness of the algorithm.

The frames of patients 3, 6 and 12 were used only in the test set. The algorithm recognized the area of interest in 26 of 29 images in video 6 and in 14 of 19 images in video 12. The worst-case scenario happened in video 3 where only 6 of 39 images are correctly recognized.

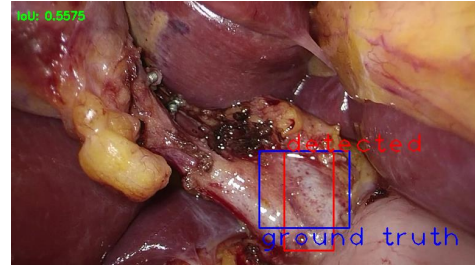
CONCLUSIONS AND DISCUSSION

This work addresses the problem of biliary tract injury during laparoscopic cholecystectomy, using an innovative approach in relation to the work suggested by the literature. The method proposes the application of YOLO for the localization of the biliary tract, creating a dataset of annotated frames of the surgical scene. The average IoU is equal to 67%, despite the small size of the dataset. The future goal is to use the localization of the biliary tract

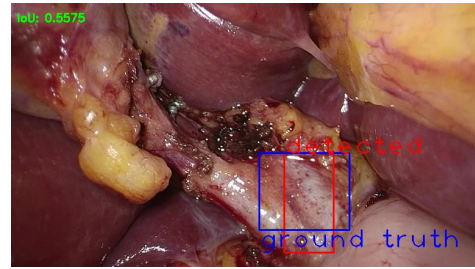
in real time and implement an augmented reality system, in order to help the surgeon to correctly and more easily recognize the area of interest in the crucial phases.



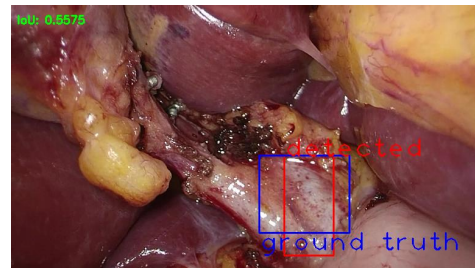
(a) Patient 3



(b) Patient 4



(c) Patient 6



(d) Patient 12

Fig. 2: Results of the localization algorithm on 4 different patients. Patient 3 (a), 6 (c) and 12 (d) belong only on the test set.

REFERENCES

- [1] S. Vlek, D. Van Dam, S. Rubinstein, E. de Lange-de Klerk, L. Schoonmade, J. Tuynman, W. Meijerink, and M. Ankersmit, "Biliary tract visualization using near-infrared imaging with indocyanine green during laparoscopic cholecystectomy: results of a systematic review," *Surgical endoscopy*, vol. 31, no. 7, pp. 2731–2742, 2017.
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

Material Characterization for Magnetic Soft Robots

T. da Veiga, J.H. Chandler, G. Pittiglio, P. Lloyd, M. Holdar, O. Onaizah, A. Alazmani, P. Valdastrì

STORM Lab, School of Electronic and Electrical Engineering, University of Leeds, Leeds, U.K.

eltgdv@leeds.ac.uk

INTRODUCTION

Over the past years, magnetic soft robots have drawn considerable attention due to the wide range of potential applications enabled through the controlled use of highly compliant materials, miniaturization, and untethered control. From flexible soft catheters to micro-robots with a wide range of locomotion capabilities [1], [2], the specific use of magnetic responsive elastomers (MRE) has shown varied possible applications to robotic surgery.

Soft robots rely on accurate material characterizations to achieve precise modeling, actuation and control. Considerable efforts have been made to mechanically characterize frequently employed non-magnetic elastomers [3], [4]. However, characterization of MRE is often restricted to their microstructure, particle behavior or magnetorheological properties [5], failing to provide macro-level properties of interest from a robotics point of view [6]. This often leads to robotic applications using properties based either on the manufacturer's data or models applied to their non-magnetic counterparts [1].

To address this, we provide a material characterization of MRE commonly employed in soft robots [1], [2], from mechanical and magnetic perspectives. For mechanical characterization we consider a tensile analysis of these different MRE and the suitability of linear elastic and commonly employed hyperelastic models to represent their behavior. For magnetic characterization we present an easily implemented, cost-effective method for determining the magnetic moment of MRE samples, using a load cell; studying the influence of particle concentration on the resulting MRE magnetic moment. The results were then applied to a soft magnetic continuum robot to demonstrate the influence of robot-specific mechanical and magnetic characterization on its simulation results.

MATERIALS AND METHODS

Two MRE were tested: Ecoflex™ 00-30 loaded with NdFeB at 0, 50, 100 and 150wt%, and Dragon Skin™ 10 at 0, 50 and 100wt%. Higher concentrations than the ones mentioned prohibited curing and thus were not considered in the study. The MRE were left to fully cure at room temperature before demolding and magnetizing in an impulse magnetizer, under either 2.7 and 5 T fields.

For the mechanical characterization, destructive uniaxial tensile testing, in accordance with ISO37:2017 was performed. A linear elastic model at 100% strain was fitted for all samples to retrieve the corresponding Young's

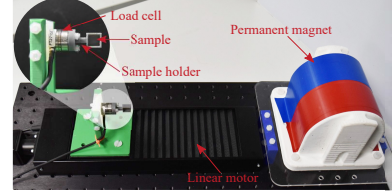


Fig. 1: Setup for the magnetic characterization of MRE.

modulus. To understand the best modeling practices for MRE, a linear elastic model, and hyperelastic models: Mooney-Rivlin, Neo-Hookean, Ogden (three coefficients), Polynomial (5 coefficients), and Yeoh were fitted to the whole strain range of the obtained stress-strain curves.

To magnetically characterize the different MRE, samples were placed under an external magnetic field and field gradient, and the generated forces and torques measured (Figure 1). The presence of the linear motor allows the collection of data at different distances from the permanent magnet (PM). By measuring the forces F (1) and torques τ (2) experienced by the samples under an external magnetic field B_e , their magnetization vector m , and magnetic moment m_a (3) can be determined.

$$F = (m \cdot \nabla) B_e \quad (1)$$

$$\tau = m \times B_e \quad (2)$$

$$m_a = \frac{\|m\|}{V} \quad (3)$$

This method was first validated on small PMs, and then used to examine the influence of particle concentration, impulse magnetizing field intensity, and MRE stiffness on the resulting MRE's magnetization.

RESULTS

Table I lists the values of Young's modulus at 100% strain, as well as the mean absolute percentage errors (MAPE) of the linear and hyperelastic models for the whole strain range. For both MRE, an increase in concentration of NdFeB microparticles led to a stiffening of the composites. However, the rate of increase of the Young's modulus decreases as the concentration goes up. The addition of NdFeB microparticles also translates to a loss of hyperelasticity, which can be seen through the lower fitting errors for the linear model as the concentration increases in Table I. Nonetheless, hyperelastic models still have lower fitting errors when compared linear models, thus showing their applicability to MRE.

TABLE I: Results of model fitting to MRE tensile test data: showing the values of Young's modulus (E) for fitting up to 100% strain, and the mean absolute percentage error (MAPE) for all fitted models.

Elastomer	Con. (wt%)	Linear model			Hyperelastic models					
		100% strain		Full strain	Neo-Hookean	Mooney-Rivlin	Yeoh	Ogden	Polynomial	
		E [kPa]	MAPE [%]	MAPE [%]	MAPE [%]	MAPE [%]	MAPE [%]	MAPE [%]	MAPE [%]	MAPE [%]
Ecoflex™ 00-30	0	42.7 ± 3.9	28.7	98.0	23.0	12.2	13.2	4.8	6.5	6.5
	50	73.2 ± 10.2	22.9	66.8	17.2	19.7	14.9	7.1	5.0	5.0
	100	102.1 ± 7.3	17.8	15.2	18.3	10.0	11.3	4.0	5.3	5.3
	150	128.2 ± 4.6	24.6	21.6	14.5	16.2	13.6	6.5	4.9	4.9
Dragon Skin™ 10	0	201.1 ± 12.0	13.6	33.0	11.3	18.9	16.0	3.6	2.1	2.1
	50	343.2 ± 9.1	10.0	9.7	20.6	13.2	9.4	3.4	1.6	1.6
	100	360.1 ± 10.9	9.9	9.2	19.4	9.2	8.5	3.3	1.3	1.3

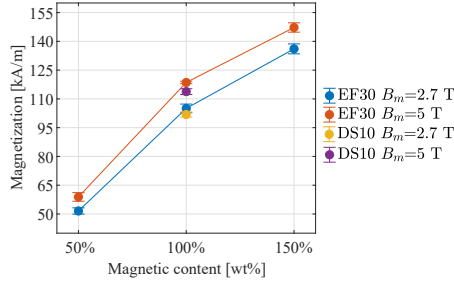


Fig. 2: Effect of the magnetic particle content and magnetizing field on the values of magnetization of MREs.

Figure 2 depicts the values of magnetization m_a obtained for the MRE samples. Higher magnetic content results in a higher value of m_a . However, this increase is non-linear; reducing as the concentration increases. No significant difference was found between Ecoflex™ 00-30 and Dragon Skin™ 10 samples. This indicates that these matrices do not affect the magnetic properties of the final MRE when submitted to a single magnetizing field. Lastly, different values of B_m gave rise to different values of m_a , even though both B_m were over the particles' intrinsic coercivity. This verifies the limitation that impulse magnetizing fields have in which the depth of penetration depends on the sample's shape and size, obeying the skin effect [7].

The results achieved for the fully soft discrete magnetic continuum robot are presented in Figure 3. The theoretical model assumes constant Young's modulus of 69 kPa for the whole length of the robot and remanence of 107 mT, as per the materials' datasheet. The corrected model, on the other hand, takes into account the difference in Young's modulus between magnetic and non-magnetic sections, as well as a more accurate value of magnetic remanence, lowering the MAPE error from 8.5% to 5.4%.

CONCLUSIONS AND DISCUSSION

In this paper, we presented a material characterization for magnetic elastomers commonly employed in soft robots that is able to reduce errors in their modeling and simulation. Mechanical analysis showed that an increase in concentration of NdFeB content translates into a loss of hyperelasticity. Even so, hyperelastic models Ogden and Polynomial are still more accurate than linear models. Magnetic characterization showed that higher magnetic content leads to higher magnetization non-linearly, slowing down with concentration. These findings were subse-

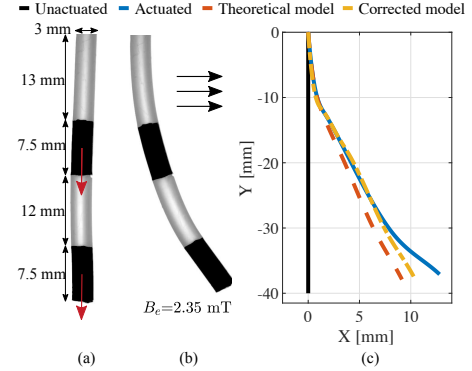


Fig. 3: Fully soft discrete magnetic continuum robot (magnetization shown by red arrows, applied uniform field by black arrows). (a) no field, (b) actuation in an uniform magnetic field, (c) corresponding simulation results considering theoretical and our corrected model.

quently validated on a fully soft magnetic continuum robot and reduced modeling errors on average by 37% when compared to using theoretical parameters. This reduction can be further increased by utilizing application strain specific mechanical properties and models, testing how the mechanical properties of the MRE changed when under actuation, and the development of more repeatable and accurate fabrication methods.

REFERENCES

- [1] P. R. Lloyd, T. Da Veiga, A. Attanasio, N. Marahrens, J. H. Chandler, P. Valdastrì *et al.*, "A learnt approach for the design of magnetically actuated shape forming soft tentacle robots," *IEEE Robotics and Automation Letters*, 2020.
- [2] Y. Kim, G. A. Parada, S. Liu, and X. Zhao, "Ferromagnetic soft continuum robots," *Science Robotics*, vol. 4, no. 33, p. eaax7329, 2019.
- [3] C. Schumacher, E. Knoop, and M. Bacher, "Simulation-ready characterization of soft robotic materials," *IEEE Robotics and Automation Letters*, 2020.
- [4] D. Steck, J. Qu, S. B. Kordmahale, D. Tscharnutter, A. Muliana, and J. Kameoka, "Mechanical responses of ecoflex silicone rubber: Compressible and incompressible behaviors," *Journal of Applied Polymer Science*, vol. 136, no. 5, p. 47025, 2019.
- [5] G. Stepanov, D. Y. Borin, A. Bakhtiiarov, and P. Storozhenko, "Negative coercivity of magnetic elastomers filled with magnetically hard particles," *Journal of Magnetism and Magnetic Materials*, vol. 498, p. 166125, 2020.
- [6] N. Bira, P. Dhagat, and J. R. Davidson, "A review of magnetic elastomers and their role in soft robotics," *Frontiers in Robotics and AI*, vol. 7, p. 146, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2020.588391>
- [7] J. Stöhr and H. C. Siegmann, "Magnetism," *Solid-State Sciences. Springer, Berlin, Heidelberg*, vol. 5, 2006.

Dual-Arm Platform for Control of Magnetically Actuated Soft Robots

Michael Brockdorff, Giovanni Pittiglio, Tomas da Veiga, James Chandler, and Pietro Valdastrì

STORM Lab UK, University of Leeds

INTRODUCTION

The present work discusses a novel approach for remote magnetic actuation. In the following, we present a full characterization of the dual External Permanent Magnet (dEPM) actuation system. Herein, we discuss how this system can be applied to fully control the magnetic field in a predefined workspace. We discuss how it can generate a *homogeneous magnetic field*, in every direction and control every *independent* gradient in the same workspace. We prove how up to 8 Degree of Freedom (DOF), 3 independent field components and 5 gradients directions, can be controlled fully independently.

The rise in popularity of magnetic actuation comes from the fact that it allows for reduction in size of magnet agents, including wireless magnetic micro-robots and magnetic Soft Continuum Robots (SCRs). SCRs have a theoretical infinite number of DOFs and thus can adapt to various nonlinear environments, minimising contact and pressure on surrounding tissue. However, millimeter scale magnetic actuation is characterized by several difficulties, related to generating uniform fields in a larger workspace and requiring higher actuating wrench (forces and torques), thus higher field strength. While successful multi-DOFs magnetic actuation has been demonstrated at small scale [1], by using systems of *coils*, large-scale manipulation is yet to be fully proven. In fact, it might require several independently-controlled coils [2] to be effective along any possible direction of motion. Despite their ability to generate both homogeneous fields [3] and gradients [2], systems of coils are less scalable, compared to permanent magnet-based magnetic field control systems [3]. In fact, due to lower field density, energy-consumption and need for high-performance cooling systems, they are generally characterized by limited workspace [4].

By further developing the idea of remotely actuating 1 Internal Permanent Magnet (IPM) (internal since, generally, inside the human body) with 1 External Permanent Magnet (EPM) [5], we discuss how 2 robotically actuated EPMs are able to magnetically manipulate 2 IPMs, independently. This is achieved by independently controlling the torque (magnetic field) and the force (field gradients) applied to each IPM.

MATERIALS AND METHODS

Magnetic manipulability is the measure of the number of magnetic DOFs that can be magnetically manipulated by a magnetic actuation system. This means that, given a set of inputs, we aim to measure the number of variables that can be independently actuated. In the following, we

prove that with 2 EPMs ($M = 2$) we can control 8 DOFs of 2 orthogonal IPMs ($N = 2$) in the same point of the workspace. Assuming that 2 IPMs are in the same point within the workspace, they will experience the same magnetic field (B) and magnetic field jacobian ($dB = \frac{\partial B}{\partial p}$). This will induce a magnetic wrench on the IPMs consequent to its magnetization m_i and location p_i as shown in (1).

$$w_i = \begin{pmatrix} 0_{3,3} & m_{i+} \\ m_{i \times} & 0_{3,5} \end{pmatrix} \begin{pmatrix} B(p) \\ dB(p) \end{pmatrix}. \quad (1)$$

$$w_i = S_i U$$

Where $w_i = \begin{pmatrix} \tau_i \\ f_i \end{pmatrix}$ and τ_i and f_i refer to the respective torque and force on the agent i . Here we introduce the operator $\cdot_+ : \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 5}$ which rearranges any vector $v \in \mathbb{R}^3$, as well as the operator $\cdot_\times : \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$ as $v_\times = (v \times e_1 | v \times e_2 | v \times e_3)$. Where e represents the canonical basis of \mathbb{R}^3 .

Finding the rank of S allows us to determine the number of controllable DOFs. It is known that for any agent i , $rank(S_i) = 5$ [6]. Moreover, one can notice that maximum manipulability can be obtained when the 2 agents are orthogonal. This comes from the fact that with this configuration $S = (S_1^T S_2^T)^T$ and $m_1 \times m_2 \neq 0$. Resulting in $rank(S) = 8$, thus proving that with $M = N = 2$ we can control 8 independent magnetic DOFs. Since 2 EPMs are used, we refer to this actuating system as the dEPM system.

Finding 8 independent DOFs is equivalent to finding 8 poses of the EPMs that lead to 8 orthogonal directions of

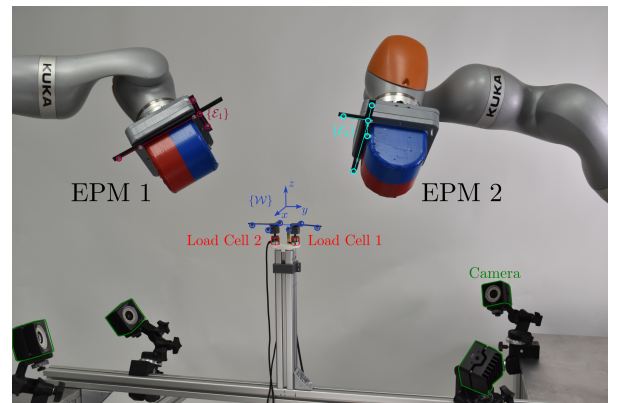


Fig. 1: Setup for magneto-mechanical actuation experiments using the dEPM setup

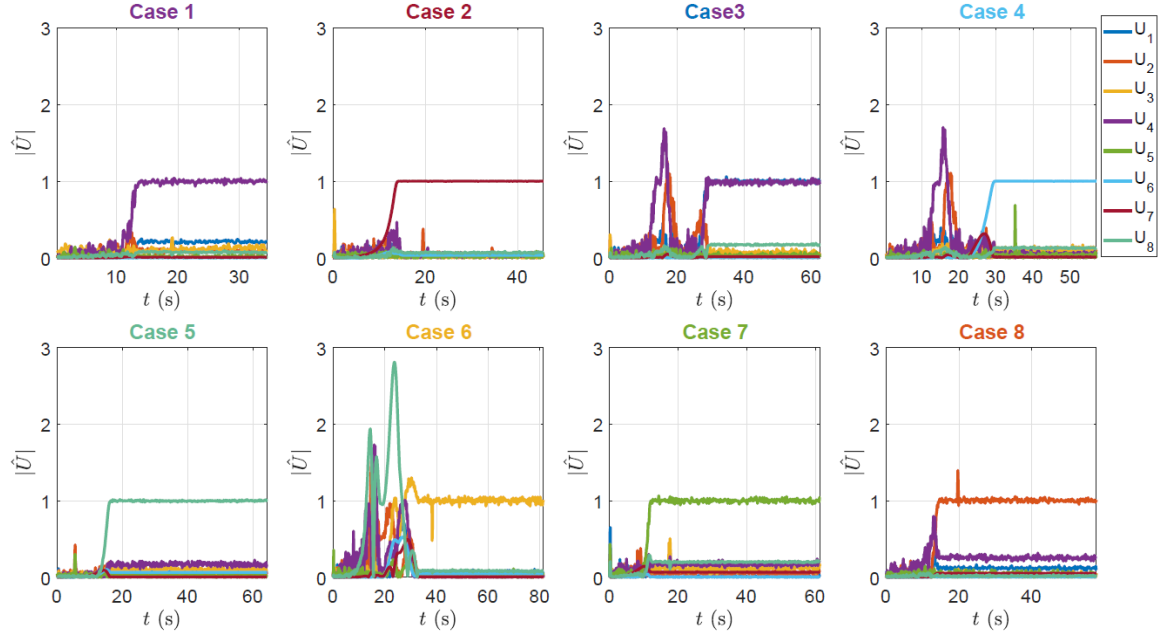


Fig. 2: Normalized response for magnetic field and differentials. Title colors are referred to the component activated for each case.

the wrench onto the IPMs. Thus, obtaining 8 independent $U(T)$, where $T = 1, 2, \dots, 8$. Due to the nonlinearities associated with solving (1), we opted for a direct analysis of primitive poses which show independent activation of field and differential component. To achieve *independent field control* we look for configurations where the magnets are aligned. In contrast, to obtain independent components of U related to the *differentials* of the field, we consider solutions with no field components. This is achieved by positioning the EPMs in opposite directions.

RESULTS

Validation of the proposed inferences, was performed through a series of experiments, aimed at proving the 8 DOFs manipulation capabilities. This was done by using 8 configurations of the EPMs for which we can control, independently, the 8 components of the field U . Each experiment was performed by placing 2 IPM coupled 6-axis load cells (Nano17 Titanium, ATI, USA) between 2 robotic arms (LBR iiwa 14, KUKA, Germany); each manipulating one of the actuating EPMs (Cylindrical permanent magnet with a diameter and length of 101.6 mm and an axial magnetization of 970.1 Am^2 (N52)), as shown in Fig. 1. Two load cells were used, each with a IPM attached to it, with IPMs orthogonal to each other. By measuring the wrench induced by each unique pose of the EPMs and inverting (1) by use of the Moore-Penrose pseudoinverse, U for each pose was measured. The normalized results for each case (independent actuation of each component of U) can be seen in Fig. 2.

CONCLUSIONS AND DISCUSSION

The present work discussed the manipulation capabilities of robotically manipulated magnetic sources. In

particular, we showed that 2 actuated EPMs are able to independently manipulate 8 DOFs. Both theoretical dissertation and experiments prove that the proposed approach can achieve the same capabilities as coil based actuation [6]. The findings can be used to potentially improve a vast range of diagnostic and interventional medical procedures through the employment of smaller and softer instruments. For example, applying the dEPM system to actuate multi-DOFs magnetic SCRs.

REFERENCES

- [1] S. Salmanipour, O. Youssefi, and E. D. Diller, "Design of Multi-Degrees-of-Freedom Microrobots Driven by Homogeneous Quasi-Static Magnetic Fields," *IEEE Transactions on Robotics*, vol. 37, no. 1, pp. 246–256, 2021.
- [2] M. Richter, V. K. Venkiteswaran, and S. Misra, "Multi-Point Orientation Control of Discretely-Magnetized Continuum Manipulators," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3607–3614, 2021.
- [3] J. J. Abbott, E. Diller, and A. J. Petruska, "Magnetic Methods in Robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 57–90, may 2020. [Online]. Available: <https://doi.org/10.1146/annurev-control-081219-082713>
- [4] M. Yousefi and H. Nejat Pishkenari, "Independent position control of two identical magnetic microrobots in a plane using rotating permanent magnets," *Journal of Micro-Bio Robotics*, vol. 17, no. 1, pp. 59–67, jun 2021. [Online]. Available: <https://doi.org/10.1007/s12213-021-00143-w> <https://link.springer.com/10.1007/s12213-021-00143-w>
- [5] G. Pittiglio, J. H. Chandler, M. Richter, V. K. Venkiteswaran, S. Misra, and P. Valdastrì, "Dual-Arm Control for Enhanced Magnetic Manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 7211–7218.
- [6] S. Salmanipour and E. Diller, "Eight-Degrees-of-Freedom Remote Actuation of Small Magnetic Mechanisms," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3608–3613.

Fiber Reinforcement for Miniaturized Magnetically Actuated Soft Continuum Robots

P Lloyd, Z Koszowska, M Di Lecce, O Onaizah, J.H. Chandler, P Valdastri

STORM Lab, School of Electronic and Electrical Engineering, University of Leeds, Leeds, U.K.

men9prl@leeds.ac.uk

INTRODUCTION

Soft continuum robots (SCR) are becoming increasingly popular in the field of medical and surgical robotics because of their ability to safely navigate tortuous pathways and access previously infeasible locations. These robots are generally designed and fabricated using low elasticity materials such as silicone polymers and can be actuated in many different ways including pneumatic, electrostatic, magnetic and chemical based methods [1]. Magnetic actuation is becoming common in the field of surgical robotics due to its ability to penetrate most non-metallic environments and because, at low strengths and frequencies, it is considered safe for biomedical applications. Furthermore, magnetic actuation is remote, since it relies on ferromagnetic bodies buried inside the actuating robot, enabling huge miniaturization potential [2] [3] [4].

Magnetic actuation relies on externally generated magnetic fields. Applying a homogeneous magnetic field will produce torque whereas applying a magnetic field exhibiting a spatial gradient will also produce magnetic force. When a magnet with magnetic moment (m) is placed in a homogeneous external magnetic field (B) it will attempt to rotate to align with the external field, producing a magnetic torque ($\tau = m \times B$). When the angle between the magnetization direction and the external magnetic field is obtuse ($>90^\circ$) the cross product *increases* as the magnet rotates under this torque. This increasing torque property can be beneficial in magnetic SCRs, producing large manipulator deformation. It is however, also prone to twisting about the torsional axis [5] as seen in Fig. 1.

Previous work has demonstrated the capabilities of fiber reinforcement in non magnetic manipulators [6] [7]. Here we have fabricated both unreinforced and fiber reinforced SCRs with permanent magnet actuation to demonstrate a significant reduction in unwanted twisting when helical fibers are present. Whilst we demonstrate significant twist mitigation, this gain was limited by largely manual manufacturing methods.

In this work we expand upon how we are developing our manufacturing process to facilitate miniaturization and improve design flexibility. This will enable us to further reduce unwanted twist and apply reinforcement in far smaller SCRs. With this process improvement we can design SCRs capable of accessing tiny and sensitive areas of the human anatomy. In particular, we are targeting an

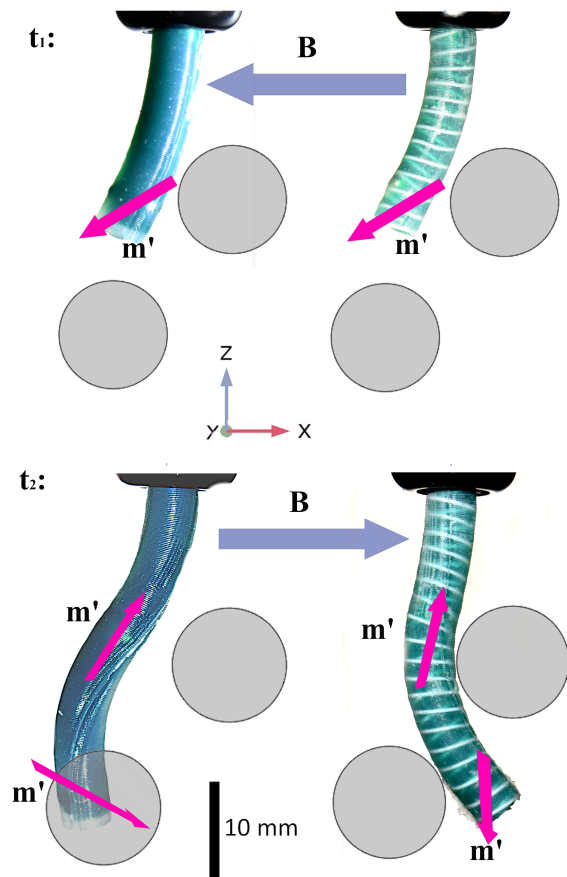


Fig. 1: Impact of Fiber Reinforcement: Unreinforced and fiber reinforced manipulator under stable actuation (top) with no twist and pure bending. Un-reinforced and fiber reinforced manipulator under unstable actuation (bottom) where twisting occurs in both but is greatly reduced in the fiber reinforced prototype.

external diameter of 1 mm at an Elastic modulus of 100 kPa with the ambition to gain access to the pancreatic and bile ducts.

MATERIALS AND METHODS

The fiber reinforcement shown in Fig 1 uses 400 μm diameter polylactic Acid (PLA) filament wound in two separate helices, a clockwise helix inside a counter-clockwise helix at an angle of 85° . Two 3.2 mm x 3.2 mm cylindrical NdFeB permanent magnets provide torques at the locations illustrated by the cyan arrows. This arrangement is embedded in a SCR of diameter 6 mm and

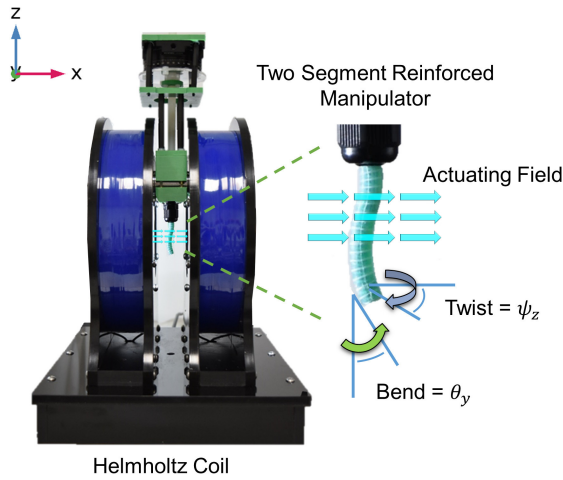


Fig. 2: A sample tentacle under actuation in the Helmholtz coil and the two motion primitives being measured: Twist about the local z axis (ψ) and Bend about the y axis (θ).

actuated in a one dimensional Helmholtz coil (DXHC10-200, Dexing Magnet Tech. Co.) up to a maximum of 20mT as shown in Fig. 2.

RESULTS

The tip of the unreinforced SCR, at worst, twisted by an angle of 145° . Under identical actuation conditions the tip of the reinforced SCR twisted by 49° . An improvement of 67% was observed in the unwanted primitive. As can be observed in Fig 1, this reduced twist has a material effect on the navigational capability of the SCR.

CONCLUSIONS AND DISCUSSION

The underlying purpose of this feasibility study is to produce a magnetic manipulator which remains mechanically stable under unstable magnetic actuation. This will allow for large deformations in a millimetre scale soft robot.

To achieve this, we are exploring a range of micro-manufacturing options. We are looking at the production of sub-millimeter scale fibrous braids to embed within magnetic silicone. We are also exploring the helical printing of material directly onto the doped silicone manipulator body using a helix printing stage which can extrude PLA filament down to a diameter of $50\mu\text{m}$ and wind this into a left or right handed helix of any chosen diameter, pitch and length.

Our ambition is to apply this technology to enter the notoriously inaccessible pancreatic and bile ducts in a safe and efficient manner. A phantom of this region of the digestive tract is shown in Fig. 3 with centrelines extracted in red. These are extremely soft ($E \approx 3\text{kPa}$ [?]) and sensitive regions of the body with tight corners requiring large manipulator deformation. At present less than 10% of pancreatic cancer patients are able to receive therapeutic or long-term treatment by surgery [8]. We believe that this study will enable a significant next step on the route towards safe, minimally invasive access for one of the most

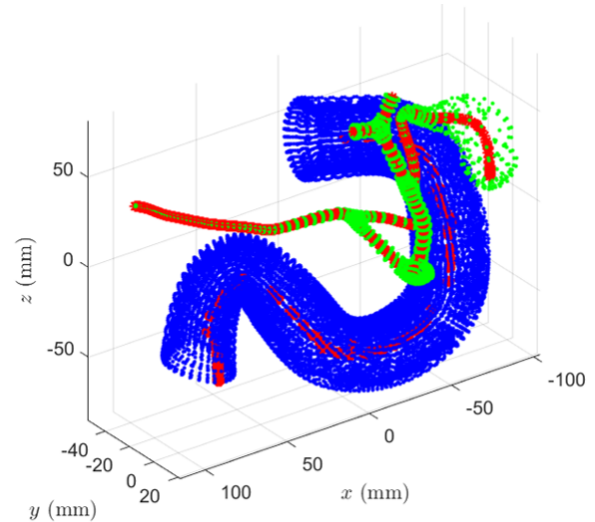


Fig. 3: Proposed path: Centerline extraction of navigation path from MRI scans of the duodenum, pancreatic and biliary ducts.

fragile and inaccessible regions of the anatomy, a region currently with terrible oncological outcomes.

REFERENCES

- [1] J. Burgner-Kahrs, D. C. Rucker, and H. Choset, "Continuum robots for medical applications: A survey," *IEEE Transactions on Robotics*, vol. 31, no. 6, pp. 1261–1280, 2015.
- [2] J. Edelmann, A. J. Petruska, and B. J. Nelson, "Magnetic control of continuum devices," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 68–85, 2017.
- [3] Y. Kim, G. A. Parada, S. Liu, and X. Zhao, "Ferromagnetic soft continuum robots," *Science Robotics*, vol. 4, no. 33, p. eaax7329, 2019.
- [4] W. Hu, G. Z. Lum, M. Mastrangeli, and M. Sitti, "Small-scale soft-bodied robot with multimodal locomotion," *Nature*, vol. 554, no. 7690, pp. 81–85, 2018.
- [5] P. Lloyd, Z. Koszowska, M. Di Lecce, O. Onaizah, J. H. Chandler, and P. Valdastrì, "Feasibility of fiber reinforcement within magnetically actuated soft continuum robots," *Frontiers in Robotics and AI*, vol. 8, p. 214, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2021.715662>
- [6] F. Connolly, P. Polygerinos, C. J. Walsh, and K. Bertoldi, "Mechanical programming of soft actuators by varying fiber angle," *Soft Robotics*, vol. 2, no. 1, pp. 26–32, 2015.
- [7] P. Polygerinos, Z. Wang, J. T. Overvelde, K. C. Galloway, R. J. Wood, K. Bertoldi, and C. J. Walsh, "Modeling of soft fiber-reinforced bending actuators," *IEEE Transactions on Robotics*, vol. 31, no. 3, pp. 778–789, 2015.
- [8] O. Strobel, J. Neoptolemos, and D. Jäger et al., "Optimizing the outcomes of pancreatic cancer surgery," *Nat Rev Clin Oncol*, vol. 16, p. 11–26, 2019.

Geometric Reinforcement towards Fully Soft Magnetic Endoscopic Endonasal Surgical Manipulators

Zaneta Koszowska, Giovanni Pittiglio, James H Chandler, Michael Brockdorff, and Pietro Valdastrì

School of Electronic and Electrical Engineering, STORM lab, University of Leeds, United Kingdom

INTRODUCTION

Endoscopic Endonasal Surgery (EES) targets the sinuses or base of the skull for treatment of lesions, tumors or polyps. The endonasal approach for these procedures is much safer than a craniotomy approach, involving removing part of the skull to access the operating site. Even though current EES is highly beneficial, technological limitations are still present [?]. EES is performed by inserting a rigid endoscope and accompanying tools through the nostrils. The coupled constraints of narrow, curved anatomy with straight and rigid tools present a significant challenge in EES, limiting visibility and maneuverability within the workspace. Enhancing the flexibility and controllable Degrees of Freedom (DoF) of the tools used in EES could make procedures safer and easier to perform, and significantly improve surgical outcomes. Most of the proposed robotic solutions for EES are mechanically driven with sizes close to standard and do not address the issue of miniaturisation or delicate tissue interaction [?], [?]. The field of soft robotics may offer solutions to the issues faced in EES, allowing small compliant manipulators with increased DoF for superior maneuverability and safer tissue interactions. Of these robotic designs, the relatively new field of Soft Magnetic Manipulators (SMMs) presents some notable advantages when designing for small scales. SMMs can be entirely soft and can be magnetically pre-programmed to produce desired deformations under exposure to specific and controlled external magnetic fields [?]. External, remote actuation of this type thus enables extreme miniaturization without loss of controllable DoFs, which is often impossible to achieve with other approaches due to the accommodation of onboard actuation. Due to these advantages magnetically actuated devices can be very beneficial in medical environment – specifically for improving navigation through torturous anatomical pathways and difficult to access surgical sites, such as in EESs. In this paper we investigate the application of soft continuum magnetic manipulators to EES procedures. In combination, the presented approach aims to enable delicate interaction with tissue, and higher maneuverability with respect to current EES tools; overcoming issues associated with torsion.

MATERIALS AND METHODS

A. PRINCIPLE OF MAGNETIC ACTUATION

A magnetic agent with magnetization μ is subject to magnetic force (f) and torque (τ), under an applied field

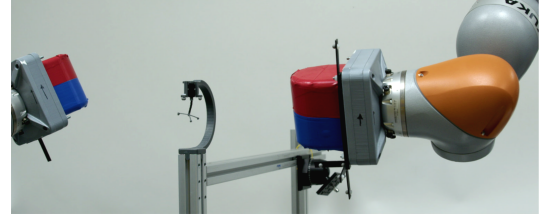


Fig. 1: Experimental set up: a manipulator with optical markers, robotic arms with External Permanent Magnets (EPMs) and frame with optical markers.

B , respectively as

$$f = \nabla(\mu \cdot B) \quad (1)$$

$$\tau = \mu \times B. \quad (2)$$

The magnetisation direction of a controlled body is a crucial element when it comes to magnetic actuation of soft structures. According to (2), when the magnetisation direction of the magnetic body aligns with the applied external magnetic field direction, it is subject to no resultant torque. If not aligned, magnetic torque will tend to rotate the body; it can be expressed as the cross product of magnetisation direction and applied external magnetic field (2). Using this phenomenon, the magnetic manipulator can be pre-programmed in a manner to achieve desired deflection, when exposed to specific external magnetic fields.

In cases where the angle between the magnetisation vector and applied magnetic field is higher than 90° , SCRs will commonly twist around the z axis of the manipulator, rather than producing expected deflection. The concept of constraining torsion in magnetic soft robots was explored by [?], through the addition of helical fibres; succeeding with torsion reduction of 67 %.

B. MANIPULATOR DESIGN AND FABRICATION

To address the issue of twisting, we consider a manipulator design with a monolithic elastic double helix reinforcement structure. Pursuing a geometrical solution to the torsional effect maintains the benefits of fully soft structure without a need for inclusion of stiff, hard structures such as springs or fibres. Figure (2) shows the design parameters in Y-Z and X-Z planes, related to the double helix SCR design. This results in different stiffness when bending about X and Y axes. Specifically, it results

TYPE	OD (mm)	ID (mm)	R	h (mm)	w (mm)
CYL	3.5	N/A	N/A	N/A	N/A
SMM 1	3.5	1	30	0.5	1.25
SMM 2	3.5	1	12.5	1.5	1.25

TABLE I: Design parameters of SMMs including geometrical variation of helical reinforcement.

in relatively low stiffness along X and high stiffness along Y directions. This variability allows for minimal bending in the Y direction and torsion around Z direction, while keeping high deflection in the X direction.

Two designs of mechanical reinforcement with varied geometrical parameters and a cylinder were considered as defined in Table I. Basing on clinical needs for miniaturisation, the outer diameter of the manipulators is set to 3.5 mm. The core cylinder diameter and width of the helices was held constant across designs (Table I). All samples were fabricated by casting, with 3D printed moulds (Grey V4 resin, Form III, Formlabs, USA). Equal parts of Dragon Skin™ 30 (Smooth-On, Inc., U.S.A.) A and B were mixed with 100 wt% of hard magnetic micro-particles (Nd-FeB with an average 5 μ m diameter and intrinsic coercivity of $H_{ci} = 9.65$ kOe, MQFP-B+, Magnequench GmnH, Germany). The magnetic slurry was then treated in a high vacuum-mixer (ARV- 310, THINKYMIXER, Japan) for 90 seconds at a speed of 1400 rpm and pressure of 20.0 kPa. The degassed slurry was injected in to the closed molds and cured at 45° for 30 minutes. To track samples during characterization in the dual arm system, a 1 mm channel of 2 mm length was added to the design to accommodate a frame with optical markers. Each design was fabricated twice to be magnetised in two directions for testing: along their X and Y axes to allow comparison of stiffness variability between the axes. (Fig. 2).

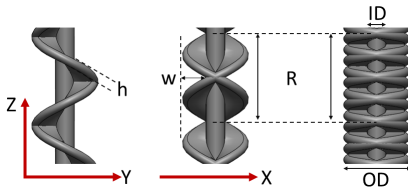


Fig. 2: Examples of monolithic reinforcement design, where h is the thickness of the helical structure, c is a core diameter, R is number of revolutions per unit length and D is an overall diameter.

C. EXPERIMENTAL SET UP

Three candidate designs (Table I) magnetised in X and Y directions were evaluated by recording data on manipulator deformation under varied conditions of external magnetic fields applied by dual arm robotic system with permanent magnets. The tip poses of manipulators are recorded via an Optitrack system (<https://optitrack.com/>) with optical markers attached to the manipulator during testing. The bending information around X, Y and Z was obtained from rotation of the rigid body.

RESULTS

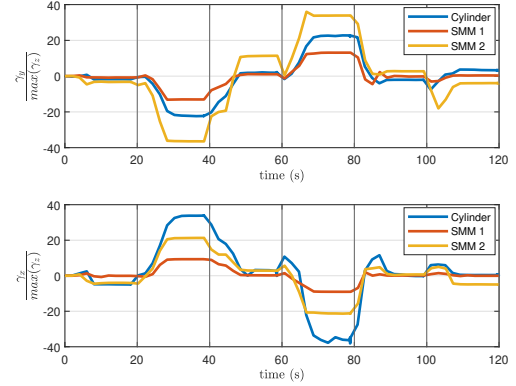


Fig. 3: Results of testing 3 candidate designs under different magnetic field conditions as a ratio of deflection to torsion. Top: Samples magnetized in X direction; Bottom: Samples magnetized in Y direction.

To find the optimal mechanical reinforcement geometry, we consider a ratio of bending in the magnetized direction to the maximum recorded torsion for all six samples Fig. 3. In the Fig. 3(top), SMM2 shows a greater ratio than a cylinder of the same diameter. In addition, the ratio value for SMM1 is lower than the value for both, cylinder and SMM2. In the Fig. 3(bottom), SMM1 shows the lowest ratio of all three samples. However, analysing both plots, it can be seen that only the design SMM2 shows desired behaviour in both magnetization cases. In the case of magnetization along X, the optimal design is expected to have the highest ratio and the lowest in case of magnetization along Y.

CONCLUSIONS AND DISCUSSION

Results collected for three candidate designs showed that the SMM2 design reduces torsion while keeping high deflection in X direction and relatively low deflection in Y direction. Therefore from samples tested in this paper, SMM2 is an optimal design to be used as a soft magnetic manipulator for EES. Future work will include investigation on a greater range of designs, varying more parameters of reinforcing geometry.

Simulated Force Control of the da Vinci PSM for Autonomous Intraoperative Ultrasound Scanning

D.P. Jones, L. Minotti, N. Marahrens, B Scaglioni, P Valdastri

*School of Electronic and Electrical Engineering, University of Leeds, UK,
D.P.Jones@leeds.ac.uk*

INTRODUCTION

Intraoperative ultrasound (US) in minimally invasive surgery allows surgeons to visualize subsurface structures and localize them as the surgery progresses. Typically, the US probe is gripped by one of the two active manipulators, reducing surgeon dexterity by limiting the available manipulators. To reduce cognitive load and allow the total usage of both manipulators, introducing autonomy to the scanning task may be implemented.

Robotic-driven US allows scans to be used in the 3D reconstruction of underlying structures, such as blood vessels and tumors. The contact force between the probe and specimen is one key factor in generating good quality scans. Contact force should be sufficient to maintain contact with the surface and to ensure quality acoustic coupling. As contact forces increase, the induced tissue deformation also increases. This is useful in certain cases, such as elastography, but limits the quality of structure localization, thereby decreasing scan accuracy [1].

In robotic surgery, ultrasound probes may be inserted as a separate entity and be held by the jaws of the manipulator. There are two main methods available to measure interaction forces: 1) integrate some form of force sensing to the tooltip, or 2) use the applied motor torques and the kinematic chain to infer the applied forces. The integration of force sensing to the tooltip is a heavily researched field [2], with researchers typically modifying existing instruments to allow for sensing. Other researchers have attempted to use motor demands and the kinematic chain to estimate tool interaction forces [3]. This method uses neural networks to identify the externally applied forces at the tooltip. While the approach is valid, mechanical resistances can cause variation in joint efforts, leading to errors in the tooltip forces.

This work details the force control method for ensuring controlled contact forces for autonomous ultrasound scanning. We deploy a simulated da Vinci Patient Side Manipulator (PSM) on a selection of different bodies. The simulation framework is modified to allow for contact forces to be estimated, and an example control system is evaluated.

MATERIALS AND METHODS

To simulate the PSM and testing bodies, we opted to use the Asynchronous Multi-Body Framework (AMBF) developed by Munawar et al. [4]. This allowed the usage of soft bodies in the testing phase. One limitation of the framework is the lack of monitoring of the contact forces

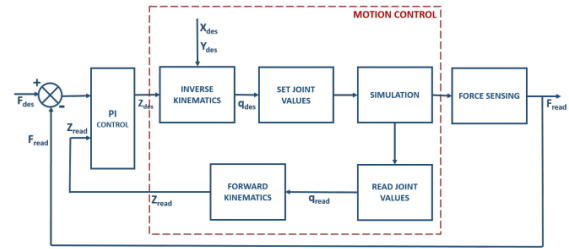


Figure. 1. Control scheme for the simulated PSM

between objects. An estimation of the manipulator contact force was inferred from the joint interaction forces in the physics solver. The engine calculates these forces and torques whenever a multi-body (in this case the robot) comes into contact with any other body. We utilized this by replacing the wrist section of the tool with a spherical tip modelled on the tip radius of an intraoperative US probe, and connecting it to the body with a rigid joint in the physics engine. The forces in this joint equated to the surface contact forces, as long as the tooltip mass was negligible.

These recorded forces were validated within the simulation by changing the mass of the tooltip indenter, and correlating the resultant change in joint forces output from the physics engine. The PSM was oriented so that the tool was perfectly vertical, and the mass of the modified tooltip was set to between 1 and 10g. As mass

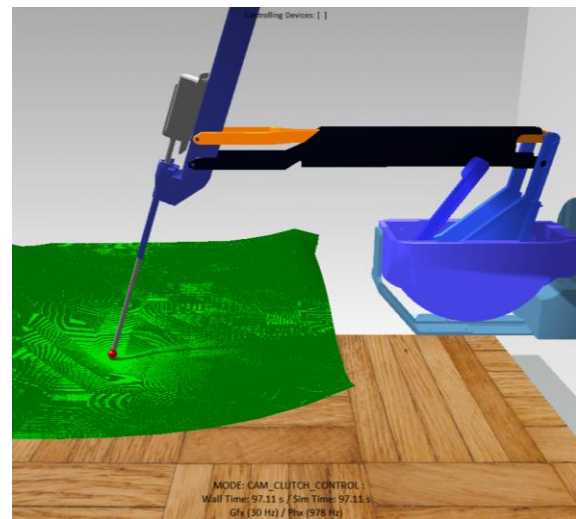


Figure. 2. PSM in the environment indenting a cloth sheet

increased, the magnitude force on the joint due to simulated gravity proportionally increased from 0.0098 to 0.098N.

The control diagram for the hybrid force/motion controller is presented in Fig 1. The forward and inverse kinematics of the modified PSM were determined to perform the Cartesian control and feedback. The contact force perpendicular to the surface was then modulated using a manually tuned PI controller. Three materials were tested in the model, each utilizing a different material method in the physics engine: a rigid cylinder, soft sponge, and soft cloth. To evaluate the control system, we devised two testing regimes: 1) a sinusoidal loading to ensure validity under varying force, and 2) a sliding motion to ensure validity in motion. For the sinusoidal loading the adapted PSM indented each material between 1.5 and 2.5N, at four target frequencies: 0.1, 0.5, 1.0 and 5.0 Hz. To ensure that the deformation was utilizing multiple joints of the PSM, the tool shaft was oriented at 70 degrees to the plane of the object (Fig 2). The object was then indented vertically along the Z-axis, maintaining the current position in X and Y. In the sliding tests, the sample was first indented to 2.5N before sliding back and forth 10mm in each direction. The indenter was set to maintain a 2.5N load at all times.

RESULTS

In the sinusoidal loading tests comparable accuracy was observed between the three bodies at each indentation frequency. The system performs best at 0.1Hz, yielding a root mean squared (RMS) error of 0.06N. The controller performed similarly in the 0.5 and 1.0Hz tests, with 0.13 and 0.18N RMS error, respectively. The 5 Hz test was notably worse than the others, with 0.38N RMS error. In sliding motion, the controller performed best on the rigid body, with 0.06N RMS error. The performance on the softer bodies was slightly worse, with 0.10 and 0.11 N RMS error in the sponge and cloth, respectively. Examples of the recorded forces are shown in Fig 3.

CONCLUSION AND DISCUSSION

The control system was evaluated, yielding less than 10% force error in the force applications under 1Hz, and less than 3% error in the constantly loaded sliding tests. In the sliding tests, spikes were observed both when the movement began and at the change in direction. This spike was deemed to be an artifact of the simulation environment. As the indenter begins to move, the mesh tended to buckle slightly, causing a steep angle between the indenter and the mesh. As the motion is continued, the controller corrects for this unusually high force. This is inherent to the physics simulation in the way that the soft body interactions are calculated, and therefore must be solved in the underlying code.

While this work shows the initial steps toward the simulation of force control in the da Vinci, in the future we wish to extend the force measurement and control to the full kinematic chain of the PSM. The adapted simulation framework will allow for us to apply the

Proceedings of the 11th Joint Workshop on
New Technologies for Computer/Robot Assisted Surgery (CRAS 2022)

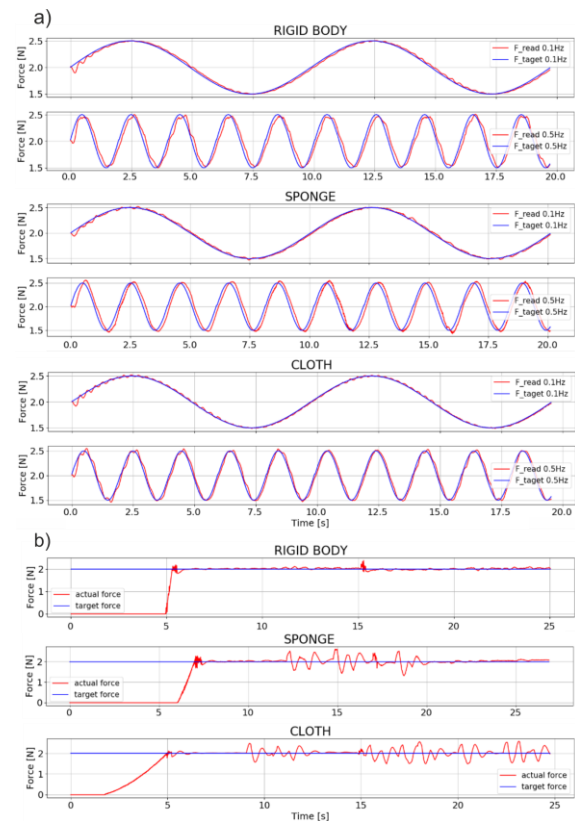


Figure 3. a) Results of the sinusoidal loading: from the 0.1 and 0.5Hz tests on the Rigid Body, Sponge, and Cloth; b) Results from the sliding test on the Rigid Body, Sponge, and Cloth.

‘dummy’ joints to the surface of each grasper face, allowing for grasping forces to be recorded. This will offer a full interaction with the grasped ultrasound probes used in surgery.

The work carried out represents a valid starting point from which the more complex environments and conditions observed in surgery may be created. The technique may also be used to generate large datasets from different surgical tasks, essential in the use of techniques such as reinforcement learning, a fundamental step in the provision of autonomy to robotic surgery.

REFERENCES

- [1] Ting-Yun Fang et al. Force-assisted ultrasound imaging system through dual force sensing and admittance robot control.. *Int. J. Comput. Assist. Radiol. Surg.* 12, 983–991. doi: 10.1007/s11548-017-1566-9
- [2] Chao Huang et al. Tactile Perception Technologies and Their Applications in Minimally Invasive Surgery: A Review. *Frontiers in Physiology*, doi: 10.3389/fphys.2020.611596
- [3] N. Yilmaz et al. “Neural Network based Inverse Dynamics Identification and External Force Estimation on the da Vinci Research Kit”. *IEEE ICRA 2020* doi: 10.1109/ICRA40945.2020.9197445
- [4] Adnan Munawar et al. A Real-Time Dynamic Simulator and an Associated Front-End Representation Format for Simulating Complex Robots and Environments. *IEEE IROS 2019*, doi:10.1109/IROS40897.2019.8968568

Vision-based Assessment and Adaption of Ultrasound Probe Tissue Coupling during Robotic Surgery

N. Marahrens¹, B. Scaglioni¹, P. Valdastrì¹

¹*STORM Lab UK, University of Leeds, Leeds, UK
elnma@leeds.ac.uk*

INTRODUCTION

While autonomous robotic ultrasound systems have been a subject of research for more than two decades, the majority of work has been performed with extracorporeal ultrasound probes, employing precisely maneuverable serial actuators, with elaborate force sensing and the possibility of mapping out the scanning surface via 3D cameras [1]. Minimally invasive surgical (MIS) systems, on the contrary, are built compliantly, commonly with tendon driven instruments as their end-effectors, making both precise kinematics and force measurements inherently difficult [2]. Previous research has commonly omitted this issue by using ultrasound phantoms with planar surfaces [3], not resembling the curved surface geometries found inside the patient. While there has been research on force sensing [4] and 3D scene reconstruction, these may not be deployable under realistic surgical constraints due to their size or their disruption to the workflow.

We, therefore, propose a more general method for purely vision-based assessment of the coupling quality over the width of the image, without the requirement of additional hardware, making it easily deployable in a minimally invasive setting and applicable to non-planar surfaces. To validate the robustness of the approach, we use the extracted coupling information to automatically improve the coupling of a robotically guided ultrasound probe and a curved surface (see Fig. 1).

MATERIALS AND METHODS

Due to the wave-nature of ultrasound, a poor coupling between the ultrasound probe and tissue in a location on the ultrasound sensor array will affect the whole image slice along the propagation direction of the ultrasound wave (depth of the image). Coupling estimation can thus be defined as a regression task with the goal of finding a curve that best describes the quality of coupling over the width of the image. Starting from this assumption we propose to further reduce this task by performing several binary classification tasks on overlapping image slices and combine these into a coarsely sampled representation of the original regression task. We believe this to be advantageous for two reasons: Firstly, this approach vastly reduces the input size to be processed by the algorithm. Secondly, it increases the number of available data samples for training and evaluation, since each labeled ultrasound image divided up into a multitude of data samples. Making the slices overlap may further increase robustness of the system due to the inclusion of

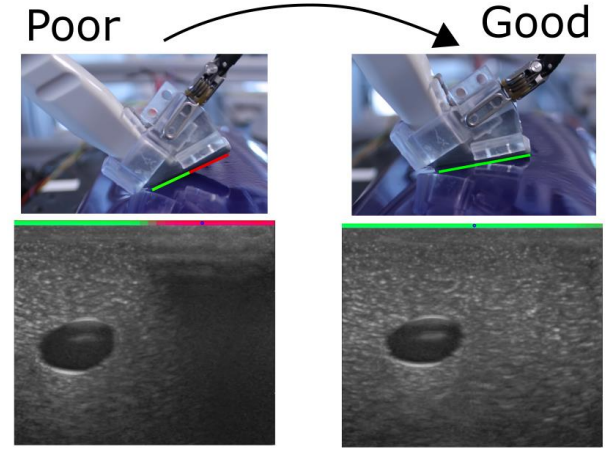


Figure 1: Coupling quality estimation between ultrasound probe and tissue surface with subsequent adaption of the ultrasound probe from poor coupling (left) to good coupling (right)

partly redundant visual information. For the proposed approach, we split the original images (979 pixels wide and between 500 and 1126 pixels deep) into 32 slices, which we found to be a good compromise between resolution and performance. For the classification task, we designed a convolutional neural network (CNN) with four convolutional layers of depth 32, 32, 64 and 64 respectively, followed by two dense layers (see Fig. 2). Between each of the convolutional layers we apply a leaky ReLU activation ($\alpha = 0.05$) function followed by anisotropic max pooling (4 in depth and 2 in width dimension). The latter accounts for the large pixel ratio between depth and width of the extracted slices. Following the convolutional layers, we process the flattened output through two dense layers with an in-between ReLU activation function and dropout of 0.5. The final classification result is generated via applying a softmax activation to the output of the last dense layer. While for training we employ these binary values as ground truth, we transition to using the resulting floating point numbers (between 0.0 and 1.0) during our prediction. This allows for simplified data labelling, while taking advantage of image slices where the classifier is less certain about its predictions.

For the case of automated probe adaption we further propose the center-of-mass (CoM) equivalent of the classified coupling quality as $c_{CoM} = \frac{\sum_{i=0}^n (c_i d_i)}{\sum_{i=0}^n d_i}$, with n being the number of slices in the image, c_i the coupling quality of slice i and d_i the distance of the center of the

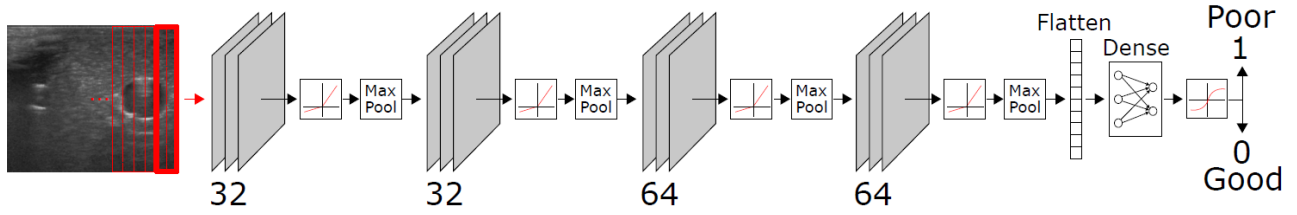


Figure 2: Structure of the proposed Deep Neural Network

of the image, to reduce the information down to a single parameter. The parameter c_{COM} therefore represents the relative imbalance of the coupling across the image. To test our proposed approach, we apply it to the adaption of an ultrasound probe (Philips L15-7io probe driven by iU22 ultrasound machine by Philips, Amsterdam, NL) that is held by a Da Vinci Surgical Robot (see Fig. 1) and automatically controlled via the Da Vinci Research Kit (DVRK) [5]. We start out by placing the probe in a partially decoupled state, equivalent to the image and robot configuration shown in Fig. 1 (left). Subsequently, we apply our algorithm to calculate c_{COM} and use the parameter value to adapt the probe angle towards the tissue surface by rotating around the image normal until the coupling balance approaches the center of the image (equal coupling on both sides – see Fig. 1 right). Rotations are performed in an incremental manner, adding the same small angle increment until full coupling is established.

RESULTS

We trained the neural network on a total of 6634 image slices with 40% of samples labeled poorly coupled and 60% well coupled. All samples were acquired on a BLUEPHANTOM Branched 2 Vessel Ultrasound Training Block Model (CAE Healthcare, Saint-Laurent, CA). We applied a 85%/15% split into training and testing data, respectively. The final mean accuracy on the test data after 250 epochs was 0.98.

Fig. 3 shows the evolution of probe rotation around the image normal axis, for 20 different runs of different positions and initial rotations on the curved phantom surface (see Fig. 1). For the plot we assumed the final rotations (fully coupled), to be an offset angle of zero, hence all lines approach the x axis. Initial angle offsets ranged between 5 and 15 degrees and were adapted between 2s and 4.5s. All twenty runs resulted in successful adaptations of the probe angle with fully coupling established. The resulting angle evolution over time, appears linear, as expected from the successive addition of angular increments.

CONCLUSION AND DISCUSSION

The presented approach for ultrasound coupling estimation showed to be capable of accurately detecting and quantifying the degree of decoupling in ultrasound images. By condensing the resulting information into a single parameter, we further showed how the approach can be directly translated into the probe angle adaption

of a robotic ultrasound probe on the da Vinci Research Kit. Current shortcoming of this approach is the possibility of fully decoupling of the probe, for which further parameters (e.g. mean over all classified slices) might need to be included for adapting the probe angle. This however, poses the question of preventing the application of excessive normal forces onto the tissue.

Future work will focus on employing the algorithm as part of task of automated navigation. To increase reliability and prevent the probe from applying excessive pressures on the tissue, we are further looking into the possibility of using algorithmically estimated normal forces from the da Vinci's motor torques.

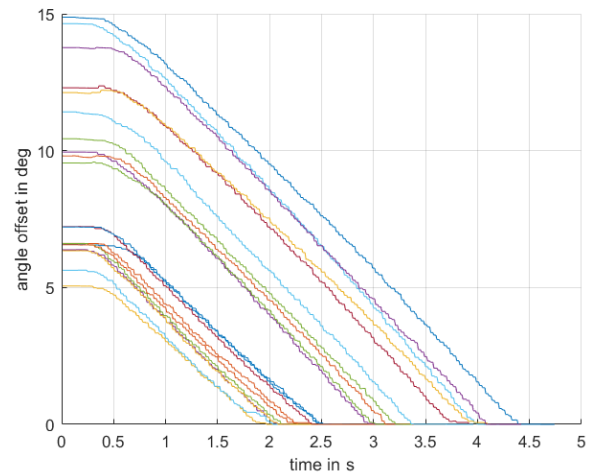


Figure 3: Evolution of ultrasound probe rotation around the image normal axis over time (zero degrees corresponding to good probe tissue coupling).

REFERENCES

- [1] F. Pierrot et al, Hippocrate: A Safe Robot Arm for Medical Applications with Force Feedback, Medical Image Analysis 3(3), 1999, pp.285–300.
- [2] S. Nia et al, Control and Tension Estimation of a Cable Driven Mechanism Under Different Tensions, Proceedings of the ASME Design Engineering Technical Conference 6A, 2013, pp.1-9.
- [3] P. Pratt et al, Autonomous Ultrasound-Guided Tissue Dissection. Medical Image Computing And Computer-Assisted Intervention (MICCAI), 2015
- [4] Lu Li et al, Development of an inexpensive tri-axial force sensor, IEEE International Conference on Intelligent Robots and Systems (IROS), pp. 906-913, 2017
- [5] Zihan Chen et al. Software Architecture of the Da Vinci Research Kit, IEEE International Conference on Robotic Computing, 2017

Autonomous Tissue Detection in Robotic Assisted Minimally Invasive Surgery

J. Hu¹, A. Attanasio¹, B. Scaglioni¹, P. Valdastrì¹

¹ Storm Lab UK, School of Electronic and Electrical Engineering, University of Leeds,

elj@leeds.ac.uk

INTRODUCTION

Robotic assisted minimally invasive surgery (RAMIS) presents several benefits for patients compared to open surgery, such as reduced trauma to the anatomical structures, shorter recovery time, and reduced blood loss. However, it could become a burden to the surgeon's cognitive load. Thus, some frequently recurring surgical tasks such as tissue retraction were studied to improve the autonomy of surgical robot. [1] Perception is regarded as one of the biggest challenges in the high-level autonomy [2]. In endoscopy-guided RAMIS, extracting useful information from the endoscopic video for robotic motion planning is the common way to increase the autonomy. In the task of tissue retraction, tissue flaps obstructing the surgical field should be detected from the video. Many detection algorithms of automating tasks have been studied. Concerning the use of deep learning algorithms in the context of surgical data, the U-Net neural network has been developed for segmentation of biomedical images [3]. Spatio-temporal feature analysis has been introduced in U-Net [4]. Generative models have been used to augment the data for tissue detection [5]. Semi-supervised learning has been studied in surgical image segmentation [6]. Detection of tissue flaps has not been well-studied, and in the paper, U-net will be applied for the automation of retraction.

MATERIALS AND METHODS

The U-Net network is characterised by 4 encoder and decoder blocks. Each block of the contracting path is structured as follow. The input, in the first place, passes through a 5×5 convolutional Long Short-Term Memory (LSTM) layer, then two 5×5 Convolutional 2D are applied, each one followed by a batch normalization and a leaky rectified linear unit (ReLU) layers. The down-sampling is implemented through a MaxPooling 2D layer, after the two convolutions. At each down-sampling step, the number of feature channels increases, thus raising from 64 to 256. Every step in the expansive path consists of an up-sampling of the feature map by means of a bilinear interpolation followed by a series of two 5×5 convolution that halves the number of feature channels. As in the encoding part, each convolution is followed by a batch normalization and a leaky ReLU layers. The two branches are connected at the bottom by a convolutional 2D layer, with kernel filters dimension $f_k = 512$. The skip connections consist in a concatenation of the upsampled feature map with the correspondingly one from the contracting path, after passing through the attention gate. At the final layer a $1 \times$

1 convolution is used to map each 64-component feature vector to the desired number of classes and the sigmoid layer assigns to each pixel the probability of belonging to the foreground or background. In total the network has 17 convolutional layers.

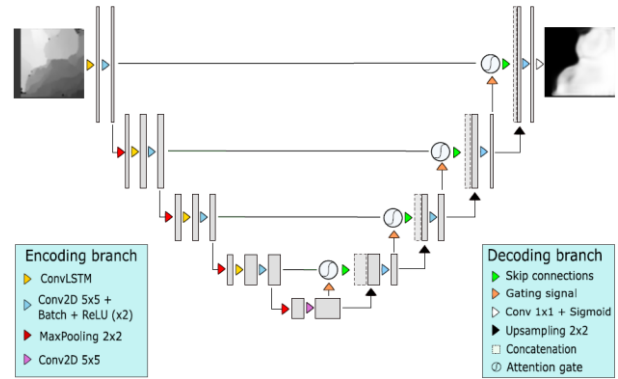


Figure. 1. Adopted neural network architecture. The input is the DM, and the output is the segmented DM.

The resolution required to identify flaps is lower than the original RGB images produced by the endoscope. Moreover, high resolution images would unnecessarily increase the time required to train the U-Net. Consequently, size of input and target images are reduced from 506×466 to 64×64 , thus allowing for faster training. The network is trained for 200 epochs with a learning rate of 0.001 and a batch size of 30 images. The Dice loss function [7] is adopted to compute accuracy and the Adam optimizer [8] is used to update the neurons' weights at every epoch.

In order to create the training dataset, surgical procedures were performed on cadavers with a clinical da Vinci Xi. Each pair of stereo images, collected with the binocular endoscopic camera, was processed in order to obtain the correspondent Depth map (DM), an image that maps the distance of objects from a viewpoint. The augmented dataset is split into a training set (90%) and a test set (10%) randomly. In order to assess the robustness of the U-Net against data variability, a training approach based on K-fold [9] cross-validation is adopted. The training process is repeated $K = 10$ times using different subsets of the dataset as training and validation sets.

RESULTS

In order to determine the best performing model, comparison between three different neural network configurations were carried out:

- UNet containing the Convolutional LSTM layer in the encoding part (ConvULSTMEnc)
- UNet having a Convolutional LSTM layer both in encoding and decoding branches (ConvULSTM)
- UNet containing Convolutional LSTM layer in the contracting path and the attention gate block (ConvULSTMAtn).

The results are shown in the Table 1:

Table.1. Mean and standard deviation comparison between the three different configurations.

	Accuracy		Precision		Recall		Jaccard	
	mean	std	mean	std	mean	std	mean	std
ConvULSTMEnc	82,95	4,00	77.09	5,37	73.41	8,40	74.64	5,61
ConvULSTM	82,80	3,38	74.14	7,47	76.38	5,62	74.90	5,22
ConvULSTMAtn	83,77	2,18	78.42	7,38	74.32	3,83	75.83	3,38

To evaluate the real-time performance of the model, a silicone phantom representing a colon is extracted from a training platform for colonoscopy (Kyoto Kagaku M40) of da Vinci Research Kit (dVRK). A section of the phantom is superposed to a background image, representing the surgical scene, aiming to simulate the presence of a tissue flap obstructing the surgical view.

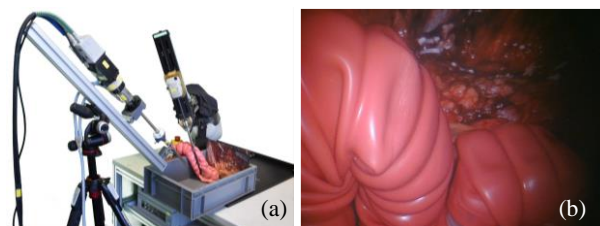


Figure. 2. (a) Phantom experiment setup. (b) Stereo image

The DMs, obtained from the endoscopic stereo images, are fed to the neural networks that outputs the predictions with 2 Hz. For the tissue retraction application, it is an acceptable inference time, since the surgical movements are slow and do not require a fast update.

CONCLUSION AND DISCUSSION

ConvULSTMAtn is considered the model with highest performances. An important step to increase the performances of the model has been the dropout tuning. Several attempts have been performed, modifying the values in the range (0, 0.5). The observed results, as shown in Figure 3, displays better performances for low dropout values. In particular the black line (dropout equals 0, recurrent dropout equals 0) shows better results. Thus, the dropout values in the ConvULSTMAtn have been set to 0 for both the recurrent and convolutional layer. On the contrary, the same tuning approach, applied to the ConvULSTMEnc, led to different results, setting both values to 0.3. Since the main difference between the two architectures are the attention gate blocks, it is very likely that, by setting random pixels to 0, the attention gate performances degrade as well.

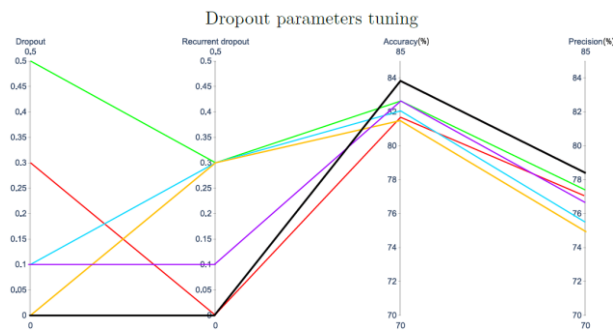


Figure 3: The values of dropout parameters have been tuned in the (0, 0.5) range. Accuracy and Precision means have been considered to evaluate the effects.

The modified U-Net model's performances were significantly improved if compared to the standard U-Net model, trained on the same dataset artificially augmented by means of tools superposition. Hence, the inclusion of temporal layers to exploit temporal properties and of attention mechanisms to refine the predictions, prove to be fundamental to improve the results.

This study is part of the research area focusing on partially automatizing surgical procedures. In the future, we will provide an autonomous approach for the tissue retraction task using the dVRK.

REFERENCES

- [1] Attanasio A. et al. "Autonomy in surgical robotics." *Annu Rev Control Robot Auton Syst*; 2021; 4:651–679.
- [2] Troccaz J, et al. "Frontiers of medical robotics: from concept to systems to clinical translation." *Annu Rev Bio Eng*; 2019; 21:193–218.
- [3] Ronneberger O, et al. "U-net: convolutional networks for biomedical image segmentation." in *Proceedings of Med Image Comput Assist Interv (MICCAI) 2015*; 234–241.
- [4] Attanasio A. et al. "A Comparative Study of Spatio-Temporal U-Nets for Tissue Segmentation in Surgical Robotics," *IEEE Trans Med Robot Bionic*; 2021; 3(1): 53–63
- [5] Lin S, et al. "LC-GAN: Image-to-image Translation Based on Generative Adversarial Network for Endoscopic Images" in *Proceedings of IEEE/RSJ Int Conf Intell Robot Syst (IROS)*, 2020, 2914-2920,
- [6] Zohar M, Bar O, et al. "Accurate detection of out of body segments in surgical video using semi-supervised learning" in *Proceeding of Med Image Deep Learn (MIDL)*. 2020: 923-936.
- [7] Milletari F, et al. "V-net: Fully convolutional neural networks for volumetric medical image segmentation. " in *Proceedings of Fourth Int Conf on 3D Vision (3DV) 2016*; 565–571.
- [8] Kingma DP, Ba J. "Adam: A method for stochastic optimization," in *Proceedings of the 3rd Int Conf on Learn Representations (ICLR) 2015*; 1–15.
- [9] Stone M. "Cross-validators choice and assessment of statistical predictions." *J R Stat Soc Series B*; 1974; 36(2): 111–133.

Human-Robot Collaboration System Using Virtual Fixture Method and Image-Guided for Dental Implant Surgery

Mohammad Hossein Hamedani, Fan Shao, Emanuele Vaia, Alessandro Cuozzo, Luca Ramaglia, Bruno Siciliano, and Fanny Ficuciello

University of Naples Federico II, 80125 Naples, Italy

INTRODUCTION

The main purpose of this work is to develop a methodology and propose an approach to improve the control of human-robot collaboration for robotic dental implant placement. In this study, a human-robotic implant system (HRIS) is designed according to the hand-guiding control in a human-robot collaboration that can increase efficiently the accurate and stable osteotomy drilling based on the surgeon's decision and robotic manipulators motion during the implant placement. The proposed method is able to modify the surgeon's position and orientation hand according to the posture of the desired placement. The implant placement is always tracked by the camera and the data of the desired position are sent to the controller of the robot. The proposed HRIS system can also be enabled or disabled during the drilling according to the decision of the surgeon for increasing the safety performance of the system.

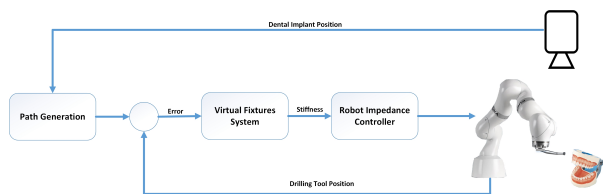


Fig. 1: The proposed structure for the hand-robot collaboration system

MATERIALS AND METHODS

To control and modify the surgeon's hand, the virtual fixture method is used as the main control approach. In fact, the virtual fixture is a software constraint method and it provides a balance between direct human control and autonomy. In this method, the stiffness of the both position and orientation of the robot is adapted according to the dental implant placement. The main idea of the presented approach is that whenever the surgeon deviates from the desired path, the robot will try to revise the motion of the surgeon's hand by increasing stiffness and making a constraint in order to return the surgeon's hand to its desired path. Additionally, using a camera and different defined tags the position and orientation of the implant placed always be available and sent to the controller of the system. The calibration of the position and orientation

of the desired implant placement is another challenge that should be considered. The calibration should be performed in the best mode in such a way that if the patient is moving during the surgical operation, the new and correct posture of the patient should be updated. In order to interact properly between robot and surgeon, the software of the robot controller is developed in a way that the surgeon can observe the position and orientation of the drilling tool during the surgical operation. In addition, the surgeon is able to choose different views from the drilling tool. To improve the safety performance of the system, there are some different gauges and alarm systems that are considered to warn the user as soon as the surgeon's hand has deviated from the desired path.

The structure of the find desired placement for the dental implant can be observed in the Fig.2. Accordingly, using a prob the points on the tooth's surface are collected, and then these points are registered with CT image shown in Fig. 3. To have real-time navigation, in this study a Micron Tracker camera is used.

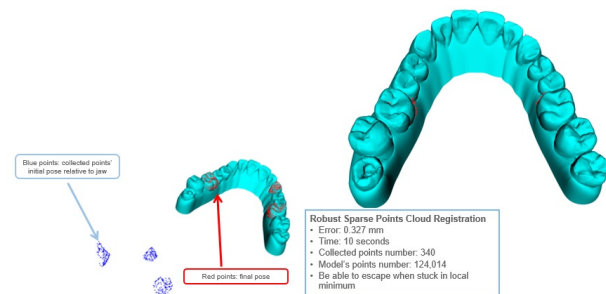


Fig. 2: Structure of identifying the best placement for dental implant

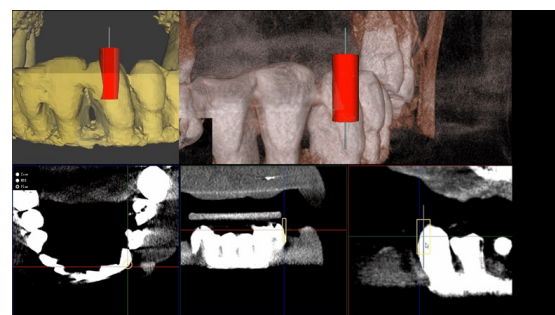


Fig. 3: CT Image of the Dental Implant Placement

CONCLUSIONS AND DISCUSSION

To verify the performance of the introduced method, the KUKA MED robot is used to perform the dental implant placement using the presented approach on a phantom head with a 3D jaw bone model. Additionally, the results between free-hand drilling and HRIS controlled drilling according to the apical center and head center of the implant placement are compared to evaluate the performance of the introduced method. Additionally, all of the required data can be logged during the surgical operation, and using this way the performance of the surgeon can be evaluated by the system and shown with the different graphs. The main structure of the introduced HRIS is presented in Fig.1. According to this structure, the camera tracks the dental placement as the desired point and provided this information for the robot, and then the robot controller is responsible to adapt the stiffness of the robot according to the virtual fixture method. The position of the drilling tool is measured by the robot using the kinematic model of the robot manipulator.

REFERENCES

- [1] T.-s. Kan, K.-j. Cheng, Y.-f. Liu, R. Wang, W.-d. Zhu, F.-d. Zhu, X.-f. Jiang, and X.-t. Dong, "Evaluation of a custom-designed human-robot collaboration control system for dental implant robot," *The International Journal of Medical Robotics and Computer Assisted Surgery*, p. e2346, 2021.
- [2] S. Aghayan and R. Rokhshad, "The art of using computer-assisted navigation systems in guided implant surgery: A review," *Journal of Research in Dental and Maxillofacial Sciences*, vol. 6, no. 2, pp. 51–62, 2021.
- [3] D. Dimri and S. Nautiyal, "Dental robotics-get going," *International Journal of Science and Healthcare Research*, vol. 5, pp. 424–426, 2020.
- [4] S. L. Bolding and U. N. Reebye, "Accuracy of haptic robotic guidance of dental implant surgery for completely edentulous arches," *The Journal of prosthetic dentistry*, 2021.
- [5] F. Yuan, Y. Wang, Y. Zhang, Y. Sun, D. Wang, and P. Lyu, "Study on the appropriate parameters of automatic full crown tooth preparation for dental tooth preparation robot," *Zhonghua kou qiang yi xue za zhi= Zhonghua kouqiang yixue zazhi= Chinese journal of stomatology*, vol. 52, no. 5, pp. 270–273, 2017.
- [6] J. Grischke, L. Johannismeier, L. Eich, L. Griga, and S. Haddadin, "Dentronics: Towards robotics and artificial intelligence in dentistry," *Dental Materials*, vol. 36, no. 6, pp. 765–778, 2020.

Reduced mathematics model of an under-actuated soft hand for Robot-assisted Laparoscopic Surgery: The MUSHA Hand II

Alberto Villani¹, Tommaso Lisini Baldi¹, Rocco Moccia²,
Fanny Ficuciello², and Domenico Prattichizzo^{1,3}

¹Department of Information Engineering and Mathematics, University of Siena, Siena, Italy.

²Department of Electrical Engineering and Information Technology, Università degli Studi di Napoli Federico II, Napoli, Italy.

³Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genova, Italy.

INTRODUCTION

The limited number of Degrees of freedom (DoFs) of robotic surgical end-effectors is an open issue for robot-assisted laparoscopic surgery. The da Vinci surgical system (Intuitive Surgical Inc., US) is a widely used robot for surgical procedures, and its Patient Side Manipulator (PSM) includes end-effector (EE), the EndoWrist surgical tool, can be controlled by the surgeon only using 8 DoFs, 5 DoF for the PSM and 3 DoF for an EE. Advanced and novel surgical instruments introduce modifications of the mechanical structure, offering new DoFs and/or compliance in the surgical system. The MUSHA (MULTifunctional Smart HAnds) Hand II, presented in [1], is a three-fingers miniature hand with force sensors for manipulation and palpation of organs in laparoscopic surgery (Fig. 1). The hand is composed of an actuation unit, three fingers, and a $\varnothing 12$ mm tube connecting them in between. The actuation unit of the hand was designed based on the original da Vinci actuation, with two additional motors to actuate finger segments movements. The hand is highly-underactuated, with 37 joints are driven by only six Degrees-of-Actuation (DoA), and underactuated design and integrated elastic elements enable adaptive prehensile grasp. Its kinematic allows to realize several type of grasp and complete various surgical task as retraction and palpation. This paper proposes a mathematical modeling policy suitable for under-actuated miniaturized surgical tools. Given the under-actuation mechanical system, the paper evaluates the reduced mathematical model based on DoAs of tendon-driven hand start using as a case of study the MUSHA Hand II.

REDUCED MODEL EVALUATION

The MUSHA Hand II has three identical fingers, the index (I), the middle (M), and the thumb (T), each of them includes 12 segments and is connected with 1-DoF wrist. The first six segments of a finger compose the proximal interphalangeal segment (PIP), while the other six compose the distal interphalangeal segment (DIP). The third joint of each finger divides the PIP segments into two parts (PIP1 and PIP2), as shown in Fig. 2 [2]. The actuation of joints of MUSHA Hand II is powered by four motors of da Vinci robot and two additional motors presented at the base of the tool trocar, the action of motors is transmitted by elastic tendons that are firmly

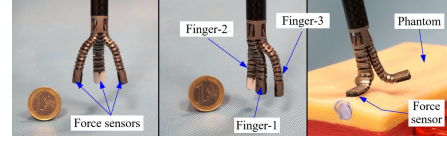


Fig. 1: MUSHA Hand II could be mounted on the PSM of the da Vinci system; in grasper mode; in fan retractor mode; in palpation mode.

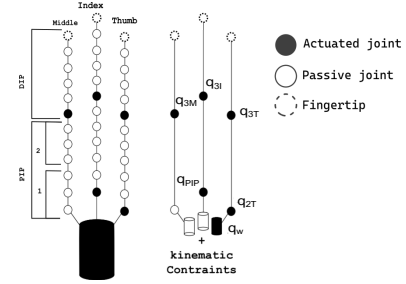


Fig. 2: Details of kinematic structure with completely extended finger and the reduced structure chosen according with the actuation system.

connected with 5 links and at the base of the wrist. Starting from this structure, it is possible to define a reduced kinematic model where the relationship between joints and the tip position $p_{r,\rho} = (x_{r,\rho}, y_{r,\rho}, z_{r,\rho})$ of finger ρ :

$$\begin{aligned} x_{r,\rho} &= \cos(q_{1,\rho})(a_2 \cos(q_{2,\rho}) + a_3 \cos(q_{2,\rho} + q_{3,\rho})) \\ y_{r,\rho} &= \sin(q_{1,\rho})(a_2 \cos(q_{2,\rho}) + a_3 \cos(q_{2,\rho} + q_{3,\rho})) \\ z_{r,\rho} &= a_2 \sin(q_{2,\rho}) + a_3 \sin(q_{2,\rho} + q_{3,\rho}) \\ \forall \rho &\in \{T, I, M\}. \end{aligned} \quad (1)$$

where $q_{i,\rho}$ is the value of i -th joint of the ρ -th finger, and a_i is the length of the i -th link. A peculiarity of the MUSHA Hand II is that the configuration of joints and the antagonist action of tendons in PIP1 and PIP2 allow separating the position and orientation of the end of DIP and PIP segments, so $q_{3,\rho}$ and $q_{2,\rho}$ are decoupled and thus they can be independently controlled. It follows that the model can be rewritten as:

$$\begin{aligned} x_{r,\rho} &= \cos(q_{1,\rho})(a_2 \cos(q_{2,\rho}) + a_3 \cos(q_{3,\rho})) \\ y_{r,\rho} &= \sin(q_{1,\rho})(a_2 \cos(q_{2,\rho}) + a_3 \cos(q_{3,\rho})) \\ z_{r,\rho} &= a_2 \sin(q_{2,\rho}) + a_3 \sin(q_{3,\rho}) \\ \forall \rho &\in \{T, I, M\}. \end{aligned} \quad (2)$$

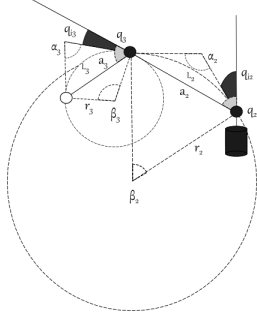


Fig. 3: Details of a bended MUSH Hand II finger.

The complete kinematic model of the MUSH Hand II, $p_r = K(\cdot)$, is obtained exploiting the equation (2) for each finger, giving the following relationship:

$$p_r = K(q_{1,T}, q_{2,T}, q_{3,T}, q_{1,I}, q_{2,I}, q_{3,I}, q_{1,M}, q_{2,M}, q_{3,M}).$$

By adding the remaining sub-actuation constraint of fixed symmetrically y-arranging of a finger around the wrist, the following result is obtained:

$$\begin{cases} q_{1,I} = q_{1,T} + \frac{2\pi}{3} \\ q_{1,M} = q_{1,T} + \frac{4\pi}{3} \end{cases} \quad (3)$$

Moreover, by adding the mechanical constraint, which implies that the proximal segments of the index and middle fingers are actuated by the same tendon: $q_{2,M} = q_{2,I}$. Therefore, the kinematic relationship between tip Cartesian positions of fingers and joint angles is equal to:

$$p_r = K(q_W, q_{2,T}, q_{3,T}, q_{PIP}, q_{3,I}, q_{3,M}). \quad (4)$$

where p_r is a vector containing fingertips position expressed in Cartesian coordinate $q_W = q_{1,T}$, and $q_{PIP} = q_{2,I} = q_{2,M}$. During coupling motion of phalanx segments, the tendons cause links shrinkage modifying the length of equivalent links. The strict requirement on precision in surgery tasks makes not negligible the variation of link length in reduced mode. Then, a dependency of links parameters is introduced, $a_{i,\rho} \forall i \in \{2, 3\}$ and $\forall \rho \in \{T, I, M\}$ with joint angles. This paper exploits this dependency using the chords theorem (Fig. 3), evaluating the length of the links equal to:

$$a_{i,\rho} = 2r_{i,\rho} \sin\left(\frac{\beta_i}{2}\right). \quad (5)$$

Let $q_{i,j,\rho}$ be the joint j in phalanx i of finger ρ , $\forall i \in \{2, 3\}$ and $\forall \rho \in \{T, I, M\}$. Consequently it is possible to write the length of the chord in kinematic model and the length of the link in actuated model as follows:

$$a_{i,\rho} = \begin{cases} \left(\frac{2L_{i,\rho}}{\sum_{j=2}^6 q_{i,j,\rho} + \sum_{j=1}^5 q_{i,j,\rho}} \right) \sin\left(\frac{\sum_{j=2}^6 q_{i,j,\rho} + \sum_{j=1}^5 q_{i,j,\rho}}{2} \right) \\ L_{i,\rho} \text{ if } q_{i,\rho} = 0, \end{cases} \quad (6)$$

where $L_{i,\rho}$ is the length of the arc of the i -th phalanx of the finger ρ . Assuming that the action of tendons is equal on all the segments, the angular displacements are equal

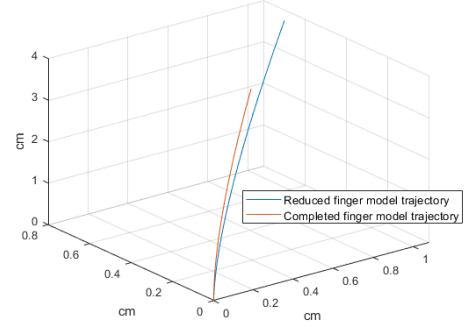


Fig. 4: Path comparison of full and reduced phalanx models .

to a fraction of the equivalent joint in the reduced model:

$$q_{i,j,\rho} = \frac{q_{i,\rho}}{6} \quad \forall j \in [1, 6], \forall i \in \{2, 3\} \text{ and } \forall \rho \in \{T, I, M\}.$$

Therefore, the explicit full hand model has 14 equations, 9 of which linearly independent, in 6 variables, as a result of repetition of the single finger model:

$$\begin{aligned} x_{r,\rho} &= \cos(q_{1,\rho})(a_{2,\rho} \cos(q_{2,\rho}) + a_{3,\rho} \cos(q_{3,\rho})) \\ y_{r,\rho} &= \sin(q_{1,\rho})(a_{2,\rho} \cos(q_{2,\rho}) + a_{3,\rho} \cos(q_{3,\rho})) \\ z_{r,\rho} &= a_{2,\rho} \sin(q_{2,\rho}) + a_{3,\rho} \sin(q_{3,\rho}) \\ a_{2,\rho} &= \begin{cases} \left(\frac{6L_{2,\rho}}{5q_{2,\rho}} \right) \sin\left(\frac{5q_{2,\rho}}{6} \right) \\ L_{2,\rho} \text{ if } q_{2,\rho} = 0 \end{cases} \\ a_{3,\rho} &= \begin{cases} \left(\frac{6L_{3,\rho}}{5q_{3,\rho}} \right) \sin\left(\frac{5q_{3,\rho}}{6} \right) \\ L_{3,\rho} \text{ if } q_{3,\rho} = 0 \end{cases} \end{aligned} \quad (7)$$

The reduced model of a finger is tested, comparing the trajectory of the tip during a complete flexion with the trajectory completed using the full kinematic model of the finger. The results are shown in Fig. 4 and the error introduced by the reduction of the model is equal to $16.04 \pm 0.09\%$ on x- and y-axes and $14.02 \pm 0.04\%$ on the z-axis.

CONCLUSIONS

This paper proposes a mathematical modeling policy for under-actuated miniaturized grippers with the aim to reduce the mapping mathematical complexity of planned trajectory from kinematic model in actuation system. The method was applied to a novel surgical tool called MUSH Hand II and evaluated in preliminary tests, showing its effectiveness. It is worth mentioning that the positioning error introduced by the reduction of the model to the actuation system alone can be neglected for small movements and, being constant in percentage points, it can be compensated for in closed-loop control.

REFERENCES

- [1] H. Liu, M. Selvaggio, P. Ferrentino, R. Moccia, S. Pirozzi, U. Bracale, and F. Ficuciello, "The MUSH Hand II: A Multi-Functional Hand for Robot-Assisted Laparoscopic Surgery," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 1, pp. 393–404, 2020.
- [2] H. Liu, P. Ferrentino, M. Selvaggio, S. Pirozzi, and F. Ficuciello, "Design of the MUSH Hand II for Robotic-Assisted Laparoscopic Surgery," *The Hamlyn Symposium on Medical Robotics*, pp. 21–22, 2019.

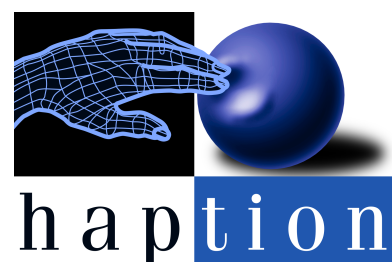
CRAS 2022 Support



UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II



FBGS



CRAS 2022

Proceeding of the 11th Joint
Workshop on New Technologies
for Computer/Robot Assisted
Surgery